

# Get real or keep tryin'

Citation for published version (APA):

Ruof, T. (2019). *Get real or keep tryin'*. [Doctoral Thesis, Maastricht University]. ProefschriftMaken Maastricht. <https://doi.org/10.26481/dis.20190426tr>

## Document status and date:

Published: 01/01/2019

## DOI:

[10.26481/dis.20190426tr](https://doi.org/10.26481/dis.20190426tr)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

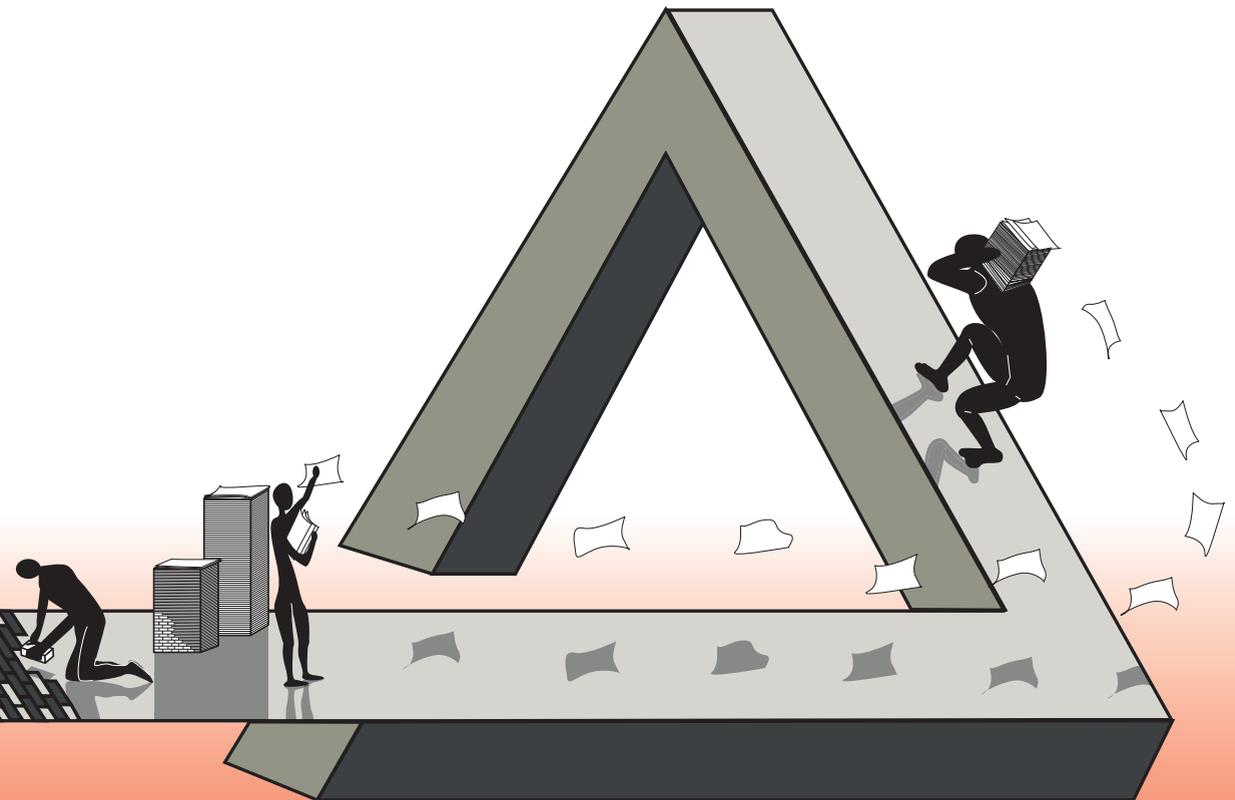
If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# Get Real or Keep Tryin'

Tobias Ruof



# Get Real or Keep Tryin'

Tobias Ruof

© 2019 Tobias Ruof

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

Cover design by Frederik Jöbges.  
This book was typeset using  $\text{\LaTeX}$ .

ISBN 978 94 6380 289 5

Printed by ProefschriftMaken || [www.proefschriftmaken.nl](http://www.proefschriftmaken.nl)

# Get Real or Keep Tryin'

## DISSERTATION

to obtain the degree of Doctor at Maastricht University,  
on the authority of Rector Magnificus Prof. Dr. Rianne M. Letschert,  
in accordance with the decision of the Board of Deans,  
to be defended in public  
on Friday, April 26, 2019, 12:00 p.m.

BY

Tobias Ruof

**Promotor:**

Prof. Dr. Rob Bauer

**Co-Promotor:**

Dr. Paul Smeets

**Assessment Committee:**

Prof. Dr. Peter Schotman, Chairman

Prof. Dr. Caroline Goukens

Prof. Dr. Sébastien Pouget (University of Toulouse)

Dr. Martin Strobel

Dr. Jan Stoop (Erasmus University Rotterdam)

*Für meine Eltern.*

*Denn das Leben der Eltern ist das Buch, in dem die Kinder lesen.  
Und dagegen steht auf diesen Seiten nichts von Relevanz – rein gar nichts.*

*(frei nach Augustinus Aurelius)*



# Acknowledgements

MARCH 22, 2019

Have you ever read an entire doctoral thesis? And I am not talking to those of you who now start thinking "Sure, daily business. I have to sit on a committee every other month." It is rather those of you that I am talking to who hold this book thinking, "Nice, let's quickly check the propositions and the acknowledgements and then I'll put this booklet on the number one spot of my dust collection." I have a challenge for you: Read the first paragraph of each of my chapters. If you do not like it, I invite you for a cup of coffee—or a beer for that matter. If you do like it, you owe me one. Deal?

Johan van Amersfoort once told me that one should never reject the offer to grab a coffee since you never know which opportunities might arise. I loved this piece of wisdom simply because it is so true. I could now write a lengthy essay about my experience as a PhD, bla bla bla. Let me just say this: Life is never too hard as long as you do what you like. Challenges are never too big as long as you enjoy trying to solve them. And there never is just one way of doing it. So do it your way since this is the only way of shaping yourself to become who you really want to be.

Let me now turn to the many thanks I owe to a whole range of people. First and foremost, I owe the deepest gratitude to my supervisors, Rob Bauer and Paul Smeets. Rob, I have always admired your business sense and your people skills. Your time to shine can best be observed when people have to be convinced. What might sound trivial comes with the challenge that these people (whoever they are) have to finance the project while storing the entire executive power in other (i.e. our) hands. Each and every time, I could witness the "Rob-treatment", after which doubts were gone and hopes were high. Your strength of bringing people together and creating an atmosphere which is both, engaging and enforcing, is a rare skill. And it this skill of yours which allowed me to eventually do one of the greatest projects. We both know about the irony behind this project (to the reader: have a look at Chapter 2, a great chapter)<sup>1</sup> and still have to laugh about how it came along. It is probably save to say that the two of us had to work a bit to break the ice. We always got along but we also had quite unique ideas about what we wanted to do. And as time passed by, I was deeply involved in two projects but none of them was a joint project with the two of us. It was not so much that we had not tried but rather that all tries ended somewhere between 'lucky losers' and 'almost, but not quite'. So

---

<sup>1</sup>I know, this sentence includes self-referencing promotion. But it's true.

---

in the summer of 2017, we sat together drinking some coffee... and we talked. You had a great idea, but I was skeptical. I told you that I needed one more project to graduate and that my willingness to take risk was, let's say medium. On top, I had several less risky ideas in mind. So we parted that day with the deal that I would work on *our* project but if it turned out to be too risky, I would jump off and work on my own project. Well, what came out was a high-risk, high-return project of which I am, of which we are still very proud. And this would not have been possible without the "Rob-treatment". So let me just say a deep thank you!

Paul, the great part about starting a PhD with you was that I already knew what I was getting in to. But while working for you had already been a blast, working with you was even better. I could not have imagined a better supervisor than you. You once said that it takes a combination of 70% agreement and 30% disagreement for a fruitful relationship—well let me tell you, we are spot-on. The best examples are the countless Sundays we spent at 'De Brandweer'. We would work hard on our scientific challenges for hours and hours, but we would never miss out on having the most lengthy and interesting discussions about anything that crossed our minds together with good food and tons of coffee. And it was not just once that the waiters asked us to leave when they wanted to close.<sup>2</sup> I truly admire your way of thinking, always critical, always logical, always clear. This made our brainstorming sessions feel like meetings which had been planned and prepared months ahead. You always had an idea, never accepted that something would not work even if truly nothing seemed to work. This attitude makes you the great researcher, the great person you are, and a great friend—and it is those sides of you that will open many doors for you in the future.

I thank all members of my reading committee who took the time to read and evaluate this book: Prof. Dr. Peter Schotman, Prof. Dr. Sébastien Pouget, Prof. Dr. Caroline Goukens, Dr. Jan Stoop, and Dr. Martin Strobel. At the same time, I thank all members of the large committee which further includes Prof. Dr. Alan Fiske, Dr. Nils Kok, Dr. Boris van Leeuwen, and Dr. Leonard Wolk. I feel honored to have such impressive researchers on my corona.

Having been part of the Finance Department has been quite a ride. I thank Jaap Bos for the many talks about research and life which were spiked with ironic comments and challenging remarks—I particularly liked that. I thank Peter Schotman for the great help, especially on technical matters of my work. A big thanks goes out to Piet Eichholtz who tells you right in your face whether he likes your project or not but in a way that is refreshing and of great guidance. Thank you, Rachel Pownall, for the interesting talks we had not only about Brexit; Peiran Jiao, for challenging my ideas; and Thomas Post, for always having an open door. And, last but not least, I thank Stefan Straetmans, Stefanie Kleimeier, Roger Otten, Jeroen Derwall, Dirk Broeders, and Dennis Bams for the good talks whenever we ran into each other.

Even though, the next two are no longer part of the Finance Department, it is as simple as 'Once you go Finance, you stay Finance': So let me particularly thank Michael Viels who has

---

<sup>2</sup> Yes, they close quite late, but don't ask me when. I only know it was dark outside.

---

been a great teaching partner for Institutional Investors as well as Leonard Wolk who helped me multiple times on my research designs and was always great taking to.

I further deeply thank Martin Strobel. I cannot count the times I was sitting in your office talking about research, ztree, or politics. We were always short on time and had too many topics we wanted to discuss. A special thanks goes out to Matthias Wibral and Antoon Pelsser who both came up with great and challenging questions whenever I presented my work.

Now let me turn to what probably made the start of my PhD the funniest and most remarkable to be imagined: my office mates Clarissa Hauptmann and Ehsan Ramezani. Both of you are just such great characters and truly awesome people. I never forget the situation where Clarissa had to write an important email. Once she had made a first draft, she asked both, Ehsan and myself, whether we thought the email to be appropriate. I told her to be more direct, Ehsan immediately jumped in claiming she already was way too direct. Thanks anyway for the good food and great talks with both of you. And, Clarissa, sorry for the many times your belongings had to serve as targets for Ehsan's and my shooting practices. Once Clarissa decided to flee our office, Nora Pankratz replaced her. She was probably the most hard-working office mate we could have imagined but was equally great and always up for great stories—thanks! And finally, this young master student showed up and wanted to join our office. We set the bar high and lowered his working table to make sure he understood the hierarchy. But he would not tolerate this and had the university staff higher his table to the highest position in the room. Matthijs, great working with you, thanks for the many good talks, and thanks for the awesome coffee which made life at TS51 finally acceptable. A great shoutout goes to Inka. While we had our little misunderstandings in the beginning, I could not help but be impressed by your strong will and determination. Keep this up and you will reach all the goals you set yourself. Juan Palacios, thanks for the many talks about research and life, the constant verbal battles which I highly enjoyed, and the great atmosphere you disseminate. Thanks, Adarajew Tamirat, for the many coffees we had in the backyard joint with inspiring conversations. Michael Kurz, thanks for sharing the passion to talk politics. Mike Langen, it was probably the best twist of fate that we never got to move in together. But other than that it was great having you around and listen to you giving advice on how to save money or get a free drink. Thanks, Lidwien Sol, for the many talks and the constant smile on your face; Matteo Bonetti, for spreading your good mood; and Luuk Perik, for always caring about the person behind the researcher. Last but not least, let me thank Joyce Mertens, Martijn Stroom, Marina Gertsberg, Shusen Qi, Hang Sun, Nagihan Mimiroglu, Sanne Jongen, Sjoke Merk, Pomme Theunissen, and Carl Vandenboorn for sharing this journey with me as well as Marten Laudi and Janek Kretschmer who get to replace me—good luck with that! Just kidding, you are doing a great job as far as I can see.

The heart of the department would not beat as lively and joyfully without the secretaries office. I loved coming in, sitting down, and talking about skiing with Francien Schijlen, about dogs and family with Carina van der Velde, about the Swiss Alps with Cecile Luijten, and about

---

the hardships of us commuters with Els van Aernsbergen. Each of you saved my ass not only once and each of you made my life a blast. Thanks!

To my paranymphs: Paulo Rodrigues, it has been fantastic with you. The love for skiing united us and the love for politics kept us together. I never forget the moment when we, a couple of new PhD students, asked you whether you wanted to join our skiing trip. You joined and it was the best trip ever. Since then, we have pretty much solved all the big problems of the world while drinking about two tons of coffee (there you go again, coffee is the solution). Above and beyond, you are one of the very few and special people who combine intelligence, empathy, and a broad interest for pretty much anything. That makes you the great personality you are and a unique friend.

Alessandro Pollastri, the Italian who is more German than me (okay, that might not be too difficult to achieve... especially giving my relaxed relationship with punctuality), but who is a great cook and pretty much provided me with food, a sleeping place, and great coffee for the better part of my PhD. I have come to miss our conversations a great deal since you have left beautiful Maastricht. Your unique humor, your big heart and hospitality, and your genius make you a great, unmatched friend.

Julia, du hast die letzten zwei Jahre meines Lebens zu einer einzigartigen Reise gemacht. Und das liegt nicht nur daran, dass du immer wissen wolltest, woran ich genau forsche und mich mit fantastischen Ideen versorgt hast. Und dennoch zeigt genau diese deine Seite, welch umwerfender Mensch du bist, der sich für alles interessiert, was ich mache. Umso mehr freue ich mich auf die Zeit, in der ich nicht mehr an dieser Arbeit sitze und endlich die vielen Versprechen einlösen kann, bei denen ich immer meinte "Nach meiner Doktorarbeit..." ♡

Finally, let me thank those people to whom I owe everything, my family. Mom und Dad, ihr seid die besten Eltern, die ich mir je hätte erträumen können. Eure bedingungslose Liebe, eure fortwährende Unterstützung, eure Geduld mit meinem teilweise etwas ungeduldigen Gemüt, eure auf einem solch tiefen und standhaften Wertefundament basierende Hingabe: ihr seid wahrlich das Buch, in dem ich am liebsten und öftesten gelesen habe und aus dem ich so Vieles gelernt habe. Wäre es nicht darum, ich stünde heute definitiv nicht hier und schriebe an den letzten Zeilen meiner Doktorarbeit. Genauso danke ich meinen unglaublichen Geschwistern. Es gibt keine besseren und ich freue mich jedes Mal wieder aufs Neue, euch zu sehen und euch in die Arme zu schließen. Eure Ratschläge, euer Lachen, eure Erlebnisse—um nichts in der Welt möchte ich das alles missen.

Now, go ahead and accept the challenge. These chapters await to be read. - Tobias

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 On the challenges of getting real . . . . .	2
1.2 A call for making decisions in (female) groups . . . . .	4
1.3 On amenities and opportunities . . . . .	7
<b>2 Get Real!</b>	<b>9</b>
2.1 Introduction . . . . .	9
2.2 The experimental design . . . . .	12
2.3 The data . . . . .	15
2.3.1 The participants . . . . .	15
2.3.2 Representativeness of our sample . . . . .	15
2.4 Results . . . . .	17
2.4.1 Financial return expectations . . . . .	18
2.4.2 Strong social preferences . . . . .	20
2.4.3 Default effects . . . . .	21
2.4.4 External validation: voting behavior and sustainable investments . . . . .	23
2.4.5 Confusion and misunderstanding do not explain our results . . . . .	24
2.5 Additional findings . . . . .	27
2.5.1 Social signaling . . . . .	27
2.5.2 Pivotal voting . . . . .	27
2.5.3 Exclusion of non-sustainable companies . . . . .	29
2.6 Conclusion . . . . .	29
A Appendix . . . . .	31
A.1 Tables . . . . .	31

A.2	Figures . . . . .	38
A.3	Survey . . . . .	40
A.4	Treatment '3 SDG Default' . . . . .	46
A.5	Treatment '4 SDG Default' . . . . .	50
<b>3</b>	<b>Dare to Pair!</b>	<b>55</b>
3.1	Introduction . . . . .	55
3.2	The experimental design . . . . .	59
3.2.1	The elicitation of beliefs . . . . .	60
3.2.2	The formation of joint beliefs . . . . .	62
3.2.3	The updating of beliefs . . . . .	63
3.2.4	The calculation task . . . . .	64
3.2.5	The payoff . . . . .	64
3.2.6	Methodology . . . . .	65
3.2.7	Restriction of the sample . . . . .	66
3.3	The data . . . . .	67
3.3.1	The participants . . . . .	67
3.3.2	Updating behavior and calculation scores . . . . .	68
3.3.3	Subjects in non-communicating pairs barely react to their teammate's belief	68
3.4	Results . . . . .	69
3.4.1	Paring causes subjects to form underconfident prior beliefs . . . . .	69
3.4.2	Pairs update symmetrically . . . . .	70
3.4.3	Dominance and updating . . . . .	72
3.4.4	Pairs update conservatively . . . . .	74
3.5	Robustness check of the main findings: including everyone . . . . .	76
3.6	Discussion and Conclusion . . . . .	77
A	Appendix . . . . .	79
A.1	Tables . . . . .	79
A.2	Figures . . . . .	83
A.3	Experimental instructions . . . . .	86
A.4	Calculations . . . . .	101
<b>4</b>	<b>Let's Meet in the Middle!</b>	<b>103</b>
4.1	Introduction . . . . .	103
4.2	Hypothesis development . . . . .	107
4.3	The experimental design . . . . .	109
4.3.1	The elicitation of beliefs . . . . .	109
4.3.2	The formation of joint beliefs . . . . .	111

---

4.3.3	The updating of beliefs . . . . .	112
4.3.4	The calculation task . . . . .	112
4.3.5	The matching . . . . .	113
4.3.6	The payoff . . . . .	113
4.3.7	Methodology . . . . .	113
4.3.8	Restriction of the sample . . . . .	114
4.4	The data . . . . .	116
4.4.1	The participants . . . . .	116
4.4.2	Updating behavior and calculation scores . . . . .	116
4.5	Results . . . . .	117
4.5.1	Men and women wrongly believe in the gender-performance gap . . . . .	117
4.5.2	Gender-diverse pairs update asymmetrically . . . . .	118
4.5.3	Disagreement leads to asymmetric updating among gender-diverse pairs . . . . .	121
4.5.4	Dominant behavior drives asymmetric updating in gender-diverse pairs . . . . .	124
4.5.5	Men and women are responsible for dominant behavior within gender-diverse pairs . . . . .	126
4.6	Additional findings and robustness checks . . . . .	129
4.6.1	Perceived math skills and asymmetric belief updating . . . . .	129
4.6.2	Robustness check of the main findings . . . . .	131
4.6.3	Using a different proxy for dominance . . . . .	133
4.7	Conclusion . . . . .	134
A	Appendix . . . . .	137
A.1	Tables . . . . .	137
A.2	Figures . . . . .	146
A.3	Experimental instructions . . . . .	149
A.4	Calculations . . . . .	163
<b>5</b>	<b>Conclusion</b>	<b>165</b>
<b>6</b>	<b>Research Impact</b>	<b>169</b>
	<b>References</b>	<b>173</b>
	<b>Biography</b>	<b>187</b>



# List of Tables

2.1	The effect of social preferences and financial beliefs on sustainable investing . . .	22
2.2	Default effects and sustainable investing . . . . .	24
2.3	The relation between voting behavior and sustainable investing . . . . .	25
A2.1	Variable definition . . . . .	31
A2.2	Summary statistics . . . . .	33
A2.3	Comparison Dutch election results to reported voting behavior . . . . .	34
A2.4	The preference for sustainable policies . . . . .	35
A2.5	Voting behavior, preferences, and sustainable investing . . . . .	36
A2.6	Approval ratings for various exclusion policies . . . . .	37
3.1	Belief updates in pairs with and without communication . . . . .	74
3.2	Belief updates and dominance . . . . .	75
A3.1	Simulated belief updates . . . . .	79
A3.2	Summary statistics . . . . .	80
A3.3	Belief updates in pairs with and without communication (all participants included)	81
A3.4	Belief updates and dominance (all participants included) . . . . .	82
4.1	Belief updates per gender subgroup . . . . .	121
4.2	Belief updates and disagreement . . . . .	123
4.3	Belief updates and dominance . . . . .	127
4.4	Belief updates per gender subgroup . . . . .	130
A4.1	Variable definition . . . . .	137
A4.2	Simulated belief updates . . . . .	138
A4.3	Summary statistics . . . . .	139
A4.4	Belief updates and reported math skills . . . . .	140
A4.5	Belief updates per gender subgroup (all participants included) . . . . .	141
A4.6	Belief updates and disagreement (all participants included) . . . . .	142
A4.7	Belief updates and dominance (all participants included) . . . . .	143
A4.8	Belief updates and dominance with a different proxy for dominance . . . . .	144
A4.9	Belief updates and dominance in gender-diverse pairs (different proxy) . . . . .	145



# List of Figures

1.1	Keeping or abolishing daylight savings time . . . . .	2
1.2	Switching to permanent summer- or wintertime . . . . .	3
1.3	Share of female ministers in European countries . . . . .	6
1.4	Percentage of management jobs held by women (1980 and 2010) . . . . .	7
2.1	The distribution of choices . . . . .	17
2.2	Financial beliefs . . . . .	19
2.3	Controlling for possible confusion . . . . .	26
A.1	Percentage choosing 4 SDGs if uncertain about return . . . . .	38
A.2	The effect of social signaling on sustainable investing . . . . .	39
3.1	Order of events in the experiment . . . . .	61
3.2	Asymmetry and communication . . . . .	72
3.3	Evolution of beliefs by treatment . . . . .	73
A3.1	Screenshot of the protocol . . . . .	83
A3.2	Screenshot of the chat . . . . .	83
A3.3	Difference in beliefs within non-communicating pairs . . . . .	84
A3.4	Empirical cumulative distribution functions . . . . .	85
4.1	Order of events in the experiment . . . . .	110
4.2	Belief adaption after being paired . . . . .	118
4.3	Asymmetry and gender composition . . . . .	119
4.4	Asymmetry and disagreement . . . . .	124
4.5	Asymmetry and dominance . . . . .	126
4.6	Asymmetry and dominance in gender-diverse pairs . . . . .	128
A4.1	Screenshot of the chat . . . . .	146
A4.2	Asymmetry and reported math skills . . . . .	147
A4.3	Development of beliefs by round . . . . .	148



# Chapter 1

## Introduction

On April 30, 1916, Germany became the first country to introduce the daylight saving time. Being caught in the middle of WWI, Germany tried to preserve resources through a more economic use of natural daylight. Several countries directly followed. But the measure had been unpopular, and as soon as the war had ended most of these countries abandoned the daylight saving time. Despite several further attempts, it was not before the energy crisis in the 1970s, that most European and North American countries permanently re-introduced the daylight saving time.

Over the past years, voices in Europe have grown louder calling for an end of the biannual time shift. But the fight over what to do resulted in political gridlock. The European Commission therefore took action and held public online consultations which allowed all European citizens to state their preferences about whether to keep the daylight saving time. The results were published on August 31, 2018 and showed a clear result. A record-breaking participation of 4.6 million people from all 28 member states had filled out the survey and 84% of the participants stated that they want the European Union to abandon the daylight saving time (see Figure 1.1).<sup>1</sup>

Consequently, Commission President Jean-Claude Juncker pledged to abandon the daylight saving time as early as 2019. But several European countries have already expressed their resistance to the plan. While the majority of respondents wants to get rid of the daylight saving time, the results about whether to stay with the standard (winter) time or the summertime are significantly less clear. Figure 1.2 shows the entire extent to which opinions differ. In Finland, approximately 30% of respondents are in favor of introducing permanent summertime while this number is close to 80% in Portugal.<sup>2</sup> And to make things yet more complicated, the response rate varied massively between countries. Of the 4.5 million participants, 3.1 million were German while only 27,000 Dutch and 20,000 Italians responded.

---

<sup>1</sup> See <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0406&from=EN>. For the complete survey, see <https://ec.europa.eu/transport/sites/transport/files/2018-summertime-arrangements.pdf>.

<sup>2</sup> See <https://www.politico.eu/article/countries-resist-jean-claude-juncker-daylight-saving-clock-change-plan/>.

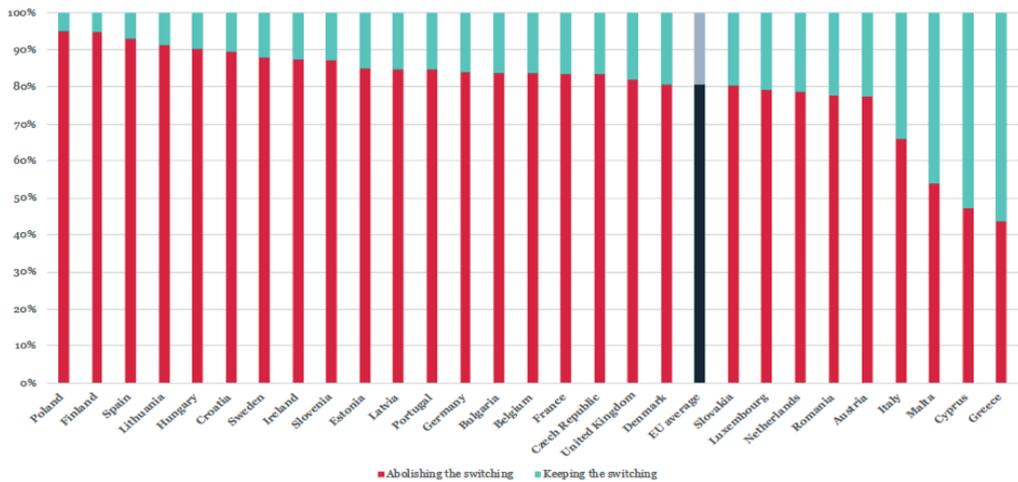


FIGURE 1.1 – Keeping or abolishing daylight savings time.

This figure shows the distribution of answers from European citizens who had to answer the question: *“Evidence suggests that common EU rules in this area are very important to ensure the proper functioning of the internal market. In order to ensure such common rules also for the future, which of the following alternatives would you favour? a. Keeping the current EU arrangements switching between summer and wintertime for all EU Member States; b. Abolishing the switching for all EU Member States”*

Source: European Commission

While the consultation was meant to put an end to the many years of fruitless discussions, it yet opened another round. This book provides three insightful chapters which could help to improve the decision-making process in similar situations.

## 1.1 On the challenges of getting real

One caveat of the approach taken by the European Commission is that the questions could have also been considered to be of hypothetical nature—neither were consequences promised nor was it clear what would have happened in case of a 50-50 decision. Hypothetical answers have been shown to be affected by various confounding factors such as socially desirable answering (Lusk and Norwood, 2009; Grimm, 2010) or a general mistrust that the survey designer might not actually care about the outcome (Terlau and Hirsch, 2015; Kollmuss and Agyeman, 2002; Fowler III and Close, 2012). As a consequence, the so-called hypothetical gap arises which marks the difference between what people say under a hypothetical regime compared to what they say if their answers have direct consequences. This behavior is what makes it difficult for the Commission to address the concerns brought forward by the opponents of the Juncker plan as they cannot say whether people really prefer summertime over wintertime. The jury is out

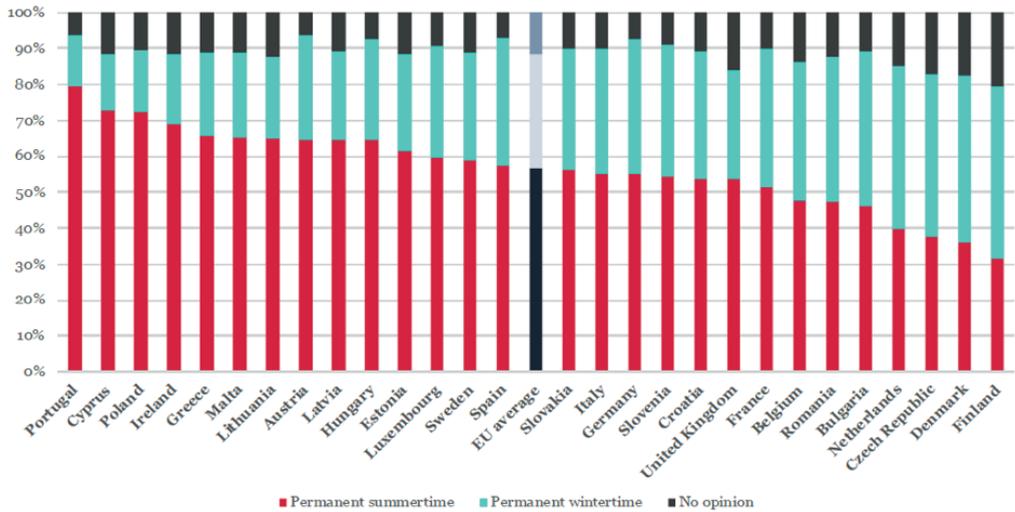


FIGURE 1.2 – Switching to permanent summer- or wintertime.

This figure shows the distribution of answers from European citizens who had to answer the question: "If the switching were to be abolished, what option would you prefer? To abolish the switching and stick with: a. permanent summertime (clocks are advanced one hour compared to wintertime); b. permanent wintertime; c. no opinion / I don't know?"

Source: European Commission

as either option has significant support and both these options could have benefitted from the fact that the question was just asked hypothetically.

One could even go as far as claiming that the results cannot reliably show that people want to abandon the current system at all. While completing the survey, respondents might have thought about what their friends will think of their choice. Equally, if they merely *perceived* the public opinion to be strongly in favor of abandoning the daylight saving time, they might have chosen to abolish it just to go with the flow. Or they might not have thought carefully enough about the consequences of their decision because they did not expect anything to happen after all. And if nothing happens, no costs will occur. These confounding factors make it hard for the European Commission as well as for the European governments to argue in a specific direction. They now will have to *guess* what their citizens really want.

In *Chapter 2*, I tackle a very similar situation. In a large field experiment, I test whether people are willing to invest their pension savings more sustainably. The unique feature about this study is that the pension fund gave its members a real, consequential vote for more or less sustainable investments which allows me to analyze preferences in an incentive-compatible environment. These answers do not suffer from the above-mentioned confounding factors like socially desirable answering. Participants know that they will experience the consequences of their choice as their pension savings will be invested more sustainably if the majority of

respondents chooses this option. With 67.9% in favor of more sustainable investments, I find strong support for more sustainable investments which is driven by strong social preferences.

Yet, importantly, these claims can only be made because respondents were given a real vote. The same findings would have been significantly less useful had I only asked hypothetical questions, as it would simply remain unknown what respondents would have chosen in a real setting. Thus, a commitment of the European commission prior to running the survey promising that they would implement the outcome could have helped them to justify their actions. It would have enabled them to argue that respondents answered based on their true preferences. Reality looks different, though. Imagine the European Union eventually decides to abolish daylight saving time and introduces permanent summertime. Such a decision could spark further dissatisfaction nurturing the disenchantment with politics. To avoid this, I claim that policy designers have to “*Get Real!*”.

### 1.2 A call for making decisions in (female) groups

In situations where a decision has to be taken on whether to abandon daylight saving time, policy makers have to balance countless interests. For example, they have to address the question whether Europe should jointly agree on introducing a permanent summer- or wintertime or whether each country can decide for itself. The Juncker plan suggests the latter but tries to nudge countries into finding a solution that does not result in a hotchpotch of different time zones with unforeseeable consequences for flight and train schedules as well as for daily commuters. To avoid this outcome, information has to be gathered, correctly evaluated, and eventually aggregated into a final decision. Upon the reception of new information, policy makers thereby have to constantly adjust their beliefs about which option best suits all interests.

Research tells us that people tend to update their beliefs more eagerly when receiving positive information while they tend to ignore negative information. Such behavior is called asymmetric belief updating and leads to unjustifiably high beliefs. Asymmetric belief updating thereby is most prominent in situations where the outcome is important to the decision makers. Imagine, the European Commission is in favor of abolishing the daylight saving time. Given the media attention toward this topic, politicians have to fear not being re-elected if they fail to deliver an acceptable solution. Thus, they clearly care about the final outcome. If the Commission now starts to gather information where they overweigh information which is in support of abandoning the daylight saving time while neglecting information against it, their beliefs about abandoning the daylight saving time being the best option would be unjustifiably high which could lead to a suboptimal decision.

However, the studies on asymmetric belief updating cannot be simply applied to this setting as they have only covered decisions of individuals (e.g. Möbius *et al.* 2014; Eil and Rao 2011). Abandoning the daylight saving time, however, is not the decision of a single individual. People

from different parties meet in committees in which they discuss their views. And at the end of that process, they decide as a group. Yet, we know very little about how asymmetric belief updating of individuals affects group decisions. Are groups able to detect such a bias and correct it or do they even amplify it ending up with even less justifiable beliefs? In Chapters 3 and 4, I investigate how beliefs are updated in groups. More specifically, I look at pairs.

*Chapter 3* focuses on the effect of communication on joint beliefs. Communication is an integral and vital part of political decision-making. But there are many examples where groups make decisions without being able to communicate. Consider elections, sport juries (e.g. boxing, dancing, ski jumping), or even the “Ask the audience” lifeline in the known game show “Who Wants to Be a Millionaire?”. Does communication thus help in forming symmetric belief updates in pairs or do they fare better if communication is not allowed? On one side, communication could help to detect wrongly held beliefs; on the other side, it could lead to a dynamic where individuals hold their own opinion back in order to comply with the group. More than 60 years ago, Asch (1956) has shown that subjects submit to group decisions even if those are obviously wrong. In an experiment, he formed groups which had to pick the longest of three drawn lines. 75% of the participants accepted the group decision if all but themselves chose a line which was clearly not the longest.

I show that pairs update symmetrically regardless of whether communication is allowed. Pairs even start out with underconfident beliefs but subsequently update these beliefs to become more confident—yet not overconfident. These findings are good news as they show that decision-making in groups leads to confident judgments where incoming information is considered regardless of its valence. My findings even suggest that while individuals suffer from overconfidence, pairs overcome this bias. Given the vast number of papers consistently finding individuals to form overconfident beliefs, people should “*Dare to Pair!*” before making important decisions. However, I also find that pairs—similarly to individuals—update their beliefs too conservatively. Thereby, available information is not fully captured causing decisions to be postponed while a correct assessment could have justified a decision already.

For decisions among policy makers like the European Commission, these findings suggest good news as its decision-making process is spiked with discussions and assessments of new situations. Yet, the conservative approach to updating beliefs might also be one indicator why decisions like abandoning the daylight saving time take so long.

In *Chapter 4*, I investigate the effect of gender on the belief updating process in pairs. This question is of high relevance for policy makers for two reasons. First, politics is a predominantly male environment. Figure 1.3 shows that in all but three European countries, the share of female ministers is less than 50%, while in more than half of the countries it is even less than 30% with prominent examples like Germany and the UK. Politics, however, is no exception. Executive positions in companies worldwide paint a similar picture. Figure 1.4 shows that in 2010 only

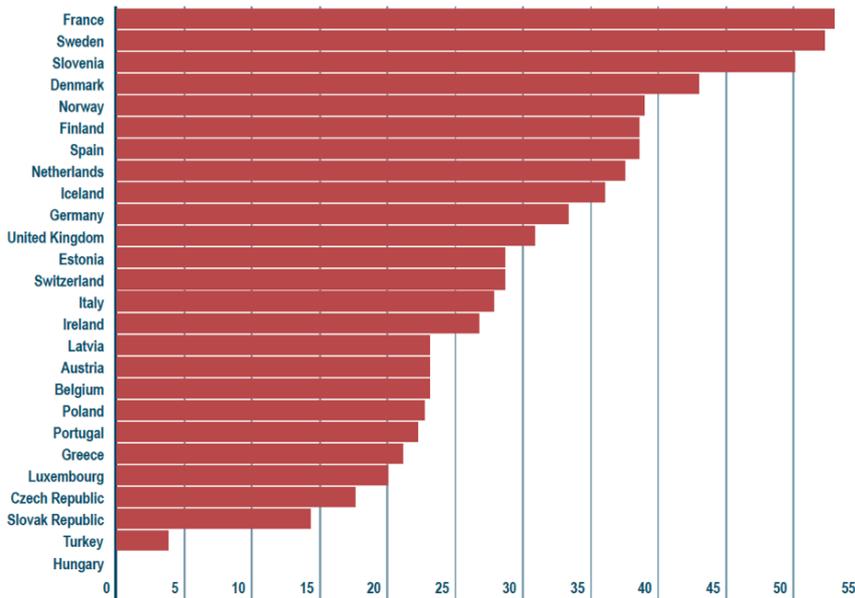


FIGURE 1.3 – Share of female ministers in European countries (March, 2018)

Source: Euronews (<https://www.euronews.com/2018/03/12/women-in-politics-how-does-europe-measure-up->)

23% of all top executive positions were held by women. This number has increased by one percentage point over the course of three decades. Second, men have consistently been found to be more overconfident than women—especially in domains which are dominated by men (e.g. Beyer 1990; Barber and Odean 2001; Bordalo *et al.* forthcoming). Do such conditions affect the belief updating and decision-making process?

At the same time, we observe an increase in the share of women in executive positions. Germany, Lithuania, Estonia, Croatia, Norway and the United Kingdom are currently led by women. And while the number of female top executives has barely increased, Figure 1.4 still shows that female management positions have seen a significant increase even in industries like finance (from 29% in 1980 to 54% in 2010). What is the effect of less overconfident women in management? Do they help to decrease the male tendency of overconfident decision-making or do women play along?

I find male pairs to update their beliefs highly asymmetrically by reacting to positive feedback 1.6 times stronger than to negative feedback. This behavior leads to quickly increasing beliefs which cannot be rationally justified. At the same time, female pairs update perfectly symmetrically causing them to hold confident, yet not overconfident beliefs. But what about gender-diverse pairs? My results show that gender-diverse pairs update as asymmetrically as male pairs. This behavior is mainly caused by dominant behavior of both, men and women.

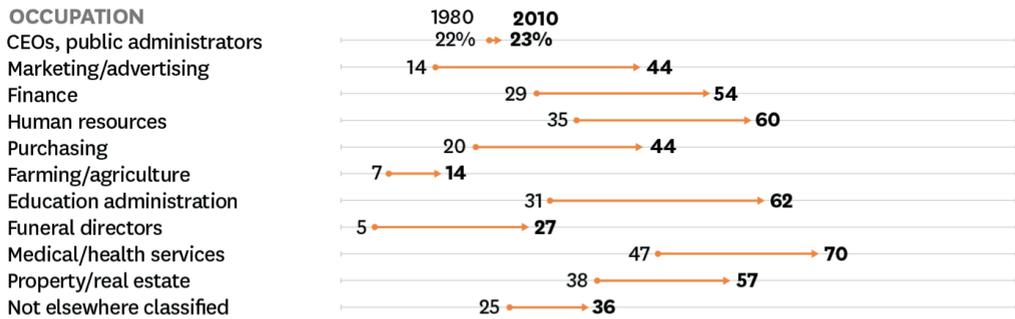


FIGURE 1.4 – Percentage of management jobs held by women (1980 and 2010)

Source: Harvard Business Review (<https://hbr.org/2018/02/what-the-data-says-about-women-in-management-between-1980-and-2010>)

Only if the two abstain from dominant behavior do they overcome asymmetric belief updating patterns and manage to update symmetrically. Therefore, they have to “*Meet in the Middle!*” of their individual beliefs when forming a team belief.

This finding shows that gender-diversity itself does not break the male domain of overconfident decision-making. But it also shows that the status quo, where positions in politics (and executive positions in general) are predominantly held by men is no solution either. As it is unlikely that purely female teams will be in charge in the near future, it will be important to form gender-diverse pairs under the condition that they refrain from dominating behavior. Another solution could be that whenever a task involves the forming and updating of beliefs, it should be done by a purely female team to avoid that decisions are driven by asymmetric belief updates. Eventually, it will be a process of try and error where beliefs and decisions are ex post evaluated to see how gender affects the formation of beliefs and how overconfident male decisions can be overcome even in gender-diverse teams.

### 1.3 On amenities and opportunities

In times where the size of a crowd is bigger even if it is smaller and where fake is the new normal, the potential of science becomes ever clearer. The self-proclaimed goal of trying to falsify what is believed to be known is thereby a scientific key element. However, science also has to fulfil its role to speak up and provide policy implications. This trade-off between composing and decomposing is what runs through my dissertation from the very beginning to the end. All three main chapters provide their policy implication already within the chapter title. In each chapter, I try to provide a better understanding of how decisions should be made to achieve better outcomes.

For it to bring fruit, we have to *Get real*, as scientists, but also as policy makers, lawyers, bankers... the list continues. We have to submit ourselves to scientifically valid assessments of the world around us and thereby take into account what people really want through giving them a real choice. And, while we as scientists try to live by these standards, there are many examples outside academia, where decisions are made without a clear idea of the driving forces. People are not given a real vote in some situations (see the example of the daylight saving time survey), while in other instances, the vote is openly placed into the hands of the people (e.g. the Brexit referendum). Both of these approaches have their advantages and both their disadvantages. It is our duty to find out what these are and how to handle them.

If this does not work, *Keep Tryin'*. This either follows from ignoring facts which prompts those to not only keep tryin', but failin'. If companies do not adhere to the fact that men, especially in groups, form overly confident decisions, they can try as much as they want to make better decisions. Improvement will lack behind. Or it follows from the spirit of a true experimental economist: learning from scientific results, adjusting, and repeating the entire process.

**Get Real or Keep Tryin'!**

# Chapter 2

## Get Real!

### Investors prefer more sustainable investments\*

#### 2.1 Introduction

People often claim to behave in a sustainable manner, but do not back their talk by action.<sup>1</sup> They state to care for animal well-being but buy the cheapest meat in the supermarket (Klink and Langen, 2015). The hypothetical gap is the difference between what people say they do and what they actually do (List and Gallet, 2001; Carson *et al.*, 2006; Harrison, 2006a; Beshears *et al.*, 2008).<sup>2</sup> It is therefore crucial to study real behavior and not mere hypothetical choices. We study sustainable investment behavior in a field experiment in which a pension fund granted its members a real vote on its sustainable investment policy.

---

\* This chapter is co-authored by Rob Bauer and Paul Smeets (both Maastricht University). We are grateful to Pensioenfondsen Detailhandel for enabling us to collect the data used in this paper and we particularly thank Henk Groot, Henk van der Kolk, Andre Snellen, and Rene Upperman. We thank seminar participants at Radboud University, at Nijmegen University, at Maastricht University, and of the 2019 European Commission's Conference on Promoting Sustainable Finance. We are also grateful to the valuable comments of Jaap Bos, Inka Eberhardt, Piet Eichholtz, Dejan Glavas, Peiran Jiao, Matthijs Korevaar, Joyce Mertens, Juan Palacios, Luuk Perik, Thomas Post, Drazen Prelec, Paulo Rodrigues, Lidwien Sol, Martin Strobel, Matthias Wibrál, and Leonard Wolk. This paper is financially supported by an Inquire grant.

<sup>1</sup> See for example Cummings and Taylor (1998, 1999); List and Shogren (1998a,b, 2002); List (2001); Ajzen *et al.* (2004); Harrison *et al.* (2004); Harrison (2006b); Vermeir and Verbeke (2006); Beshears *et al.* (2008); Harrison and Rutström (2008); Gracia *et al.* (2011); FeldmanHall *et al.* (2012); De-Magistris *et al.* (2013); Terlau and Hirsch (2015); for an overview see Harrison (2006b); Loomis (2014); for a meta-study see List and Gallet (2001); Murphy *et al.* (2005).

<sup>2</sup> The gap between what people say and what they do has received various names next to the hypothetical gap, namely: the hypothetical bias, the intention-behavior gap, the attitude-behavior gap, normative vs revealed preferences, the value-action gap, the judgment-action gap, or the green gap (without claiming completeness). In this paper, we will refer to it as hypothetical gap.

Sustainable investments grew in importance over the last years (McKinsey, 2017).<sup>3</sup> Consequently, sustainable investment behavior has attracted increased attention from academics (e.g. Hong and Kostovetsky 2012; Krüger 2015; Bialkowski and Starks 2016; Barber *et al.* 2018; Riedl and Smeets 2017; Hartzmark and Sussman 2018; Krüger *et al.* 2018). Understanding why people invest sustainably is not only important to academics, but also to institutional investors who often invest on behalf of individuals. Pension funds using a defined-benefit scheme hold about 22 trillion dollars in assets under management (Willis Tower Watson, 2018).<sup>4</sup> The United Nations estimated that reaching the Sustainable Development Goals (SDGs) requires five to seven trillion dollars. This shows the dimension of sustainable investing in the defined-benefit pension industry alone. But, how should pension funds manage their fiduciary duty regarding sustainable investments? How should they decide on behalf of their clients? The European Commission installed a High-Level Expert Group on Sustainable Finance in 2017. On the basis of the experts' proposal, the Commission intends to introduce a formal requirement that “*Investment firms... [should ask retail investors] about their preferences for sustainable investments.*”<sup>5</sup>

We provide a scientifically valid method to elicit these preferences. We conducted a field experiment ( $n = 1,669$ ) in cooperation with a Dutch pension fund which had 18.7 billion euros of assets under management in 2016.<sup>6</sup> This defined-benefit pension fund invests on behalf of its members. As part of our experiment, the board of the pension fund gave its members a real vote on its future sustainable investment strategy. Participants faced the choice of more or less sustainable investments. The board guaranteed that it will implement the outcome of the voting. Pension savings of the members are at stake, so the choice is relevant. We told participants that sustainable investments not only focus on financial returns but also on societal returns. In addition, we measured participants' beliefs about the expected financial returns of sustainable investments.

Our experimental design has three key advantages. First, the consequential vote with only two choices makes our experiment incentive-compatible (Cummings *et al.*, 1995, 1997; Harrison, 2006a; Carson and Groves, 2007).<sup>7</sup> We are not aware of any pension fund that gave its members a consequential vote on the investment strategy of the whole pension plan. Second,

<sup>3</sup> McKinsey (2017) reports that investments according to environmental, social, and governance principles are growing by an annual rate of 17%. In the United States, sustainable investments account for over a third and in Europe for half of all assets (JP Morgan, 2018).

<sup>4</sup> According to an estimate by Willis Tower Watson (2018), pension funds worldwide hold USD 45 trillion assets under management.

<sup>5</sup> See [https://ec.europa.eu/info/sites/info/files/180524-sustainable-finance-factsheet\\_en.pdf](https://ec.europa.eu/info/sites/info/files/180524-sustainable-finance-factsheet_en.pdf).

<sup>6</sup> See <https://www.ipe.com/Uploads/v/a/k/Top-1000-European-Pension-Funds.pdf>.

<sup>7</sup> A choice set with more than two choices opens the door for strategic voting. Imagine a person has to choose from three options  $A$ ,  $B$ , and  $C$  and the option chosen by the majority will be implemented. Further imagine the person's preferences to be  $A \succ B \succ C$ . If she expects that option  $A$  will be chosen by only very few people, it can be optimal for her to pick option  $B$  to at least avoid option  $C$ . This would however imply that she did not pick the option which maximized her utility. Therefore, the choice is not incentive-compatible. The lacking incentive-compatibility of stated preferences can be of serious concern as it opens the door for interfering

the political preferences of our sample are close to the outcome of the last national election in the Netherlands. This is important, because it allows us to control for possible biases of social preferences within our sample.<sup>8</sup> Third, we published a pre-analysis plan of our study at the Open Science Framework<sup>9</sup>, in which we specified our measures and sampling procedure.

We find that 67.9% of participants favor to invest their pension savings more sustainably. Only 10.8% are against it while 21.2% do not have an opinion. What drives the strong support for more sustainable investments? We explore three possibilities. First, participants might expect sustainable investments to financially outperform conventional investments. Second, participants could have strong social preferences in favor of sustainable investments in which case they even support sustainable investments when these are financially costly. Third, subjects might have been confused or did not take their real choice seriously.

We show that social preferences rather than financial beliefs drive the choice for more sustainability. The majority voted for the more sustainable option. This was the case even among those who expected lower financial returns with more sustainable investments and subjects who are uncertain about the returns. We find that strong social preferences drive the choice for more sustainable investments. First, a validated measure of social preferences (Falk *et al.*, 2016) is positively related to the choice for more sustainable investments. Second, we find that people who voted for a political party with a stronger sustainability agenda are more likely to support more sustainable investments by the pension fund. Third, the choice for sustainable investments is not influenced by different defaults. The absence of default effects is another sign of strong preferences.<sup>10</sup> The results further show that confusion or lack of information do not drive our results.

Our paper makes two main contributions. First, we contribute to the discussion on sustainable investing. Previous findings show that investors value sustainability in their investment decisions (Bollen, 2007; Hong and Kostovetsky, 2012; Barber *et al.*, 2018; Hartzmark and Sussman, 2018). Yet, these aggregate-level studies cannot inform institutional investors how to invest on behalf of their clients. The data do not show what fraction of individuals has strong social preferences. Riedl and Smeets (2017) use individual-level data to show that sustainable investors indeed have strong social preferences. However, in their study individuals invest on their own behalf and their investment decisions cannot simply be generalized to a setting of delegated portfolio management. In particular, if people's pension is at stake, choices could be very different. We show that social preferences play an important role in delegated investment

---

answering motives such as socially desirable answering (e.g. Lusk and Norwood 2009; Grimm 2010; Norwood and Lusk 2011; Klink and Langen 2015).

<sup>8</sup> Several studies show that political preferences are a valid proxy for social preferences: Fisman *et al.* (2017); Kerschbamer and Müller (2017); Almås *et al.* (2016); Cohn *et al.* (2018).

<sup>9</sup> See <https://osf.io/pz7rn/>.

<sup>10</sup> Two of the main drivers responsible for the status quo bias are loss aversion and regret avoidance (Samuelson and Zeckhauser, 1988; Kahneman *et al.*, 1991; Tversky and Kahneman, 1991; Feldman *et al.*, 1999; Nicolle *et al.*, 2011) which should especially matter for choices with real consequences.

decisions. Pension fund members are even willing to forego financial returns to invest in a sustainable manner.

Second, we contribute to a growing stream of literature on social preferences in the field (e.g. Frey and Meier 2004; Gneezy and List 2006; List 2006; Gneezy *et al.* 2010; DellaVigna *et al.* 2012; Andreoni *et al.* 2017; Kessler *et al.* 2018). Specifically, we study how pension funds should deal with the social preferences of their clients. It has become commonplace to acknowledge the influence of time preferences on retirement decisions (O'Donoghue and Rabin, 1998; Carroll *et al.*, 2009; Beshears *et al.*, 2014; Goda *et al.*, 2015). Similarly, several studies show how to elicit risk preferences of pension fund participants (Donkers *et al.*, 2012; Alserda *et al.*, 2016). However, social preferences are not yet considered (EUROSIF, 2016, p. 82). Investment managers often fear that this is too difficult and too costly.<sup>11</sup> Such concerns could lead to a poor execution of the European Commission's directive. We show a simple, scientifically valid way to elicit preferences for sustainable investments. If pension funds take these preferences seriously, this can have a large economic and societal impact.

## 2.2 The experimental design

In June 2018, we invited 49,552 active members of Pensioenfonds Detailhandel to participate in our online survey. The survey consisted of three parts. Part 1 briefly familiarized all respondents with the concept of sustainable investments and introduced the United Nations Sustainable Development Goals (SDGs).<sup>12</sup> Appendix A.2 shows the instructions and questions. In Part 2, we introduced our treatments. Respondents were randomly assigned to one of two groups and were in either case unaware of the existence of the other treatment group. The first group was confronted with three SDGs as default (3 SDG default) the second with four SDGs (4 SDG default). We explain these treatments in more detail below. Part 3 elicited social preferences and financial return expectations regarding sustainable investments. Further, we asked respondents to provide information on their gender, age, and education; on their financial background; and on their voting behavior in the 2017 Dutch national elections. Part 3 was identical across all treatments.

In Part 2, respondents learned that they would decide on the sustainable investment strategy of their pension fund. The exact framing of the decision depended on the respective treatment (see below). Importantly, all respondents learned, irrespective of their treatment, that they would have to state whether their pension fund should focus on three or four Sustainable Development Goals (SDGs). We explained that investing according to an SDG meant that the

---

<sup>11</sup> See, for example, a statement by the BVI, a German investment fund association representing asset managers with over EUR 3 trillion assets under management ([https://www.bvi.de/fileadmin/user\\_upload/Regulierung/Positionen/2018\\_06\\_21\\_BVIs\\_view-MiFID\\_II\\_sustainability\\_requirements.pdf](https://www.bvi.de/fileadmin/user_upload/Regulierung/Positionen/2018_06_21_BVIs_view-MiFID_II_sustainability_requirements.pdf)).

<sup>12</sup> For more information, see <https://www.un.org/sustainabledevelopment/sustainable-development-goals>.

Pension fund would engage with the companies' board to assure that the company's policy and actions are aligned with the respective SDG. We told them that in 2017 Pensioenfonds Detailhandel had spoken with a total of 246 corporate boards to promote sustainability. In addition, we gave them several examples of earlier SDG engagement by the Pension Fund. We explicitly pointed out that focusing on four SDGs would result in more sustainable investment activities than focusing on three SDGs.

### 3 SDG default vs 4 SDG default

Staying with the default option is a widely found phenomenon, also called the status quo bias (e.g. Samuelson and Zeckhauser 1988; Tversky and Kahneman 1991). Two of its main drivers are loss aversion and regret avoidance (e.g. Samuelson and Zeckhauser 1988; Kahneman *et al.* 1991; Tversky and Kahneman 1991; Feldman *et al.* 1999; Nicolle *et al.* 2011). In particular if people get a real choice, they might anticipate regretting their choice for more sustainable investments, if these were to result in financial losses. With the introduction of different defaults, we can account for differences in the status quo effect.

In the **3 SDG default treatment**, we told participants that the pension fund currently focuses on three SDGs, namely: 'Climate action', 'Decent work and economic growth', and 'Peace, justice, and strong institutions'. Participants were then introduced to the fourth SDG 'Responsible consumption and production'. They could choose:

*"Do you want Pensioenfonds Detailhandel to add the fourth sustainable development goal 'Responsible consumption and production'?"*

- a. Yes, add
- b. No, do not add
- c. I have no opinion regarding this matter"

In the **4 SDG default treatment**, we told participants that the pension fund's future policy will include four SDGs 'Climate action', 'Decent work and economic growth', 'Peace, justice, and strong institutions' and 'Responsible consumption and production'. Participants could then decide:

*"Do you want Pensioenfonds Detailhandel to leave out the fourth sustainable development goal 'Responsible consumption and production'?"*

- a. Yes, leave it out
- b. No, do not leave it out
- c. I have no opinion regarding this matter"

Importantly, the question of interest—whether Pensioenfonds Detailhandel should focus on three or on four SDGs—was the same for both default treatments except for the words ‘add’ or ‘leave out’. Respondents could either choose to add (3 SDG default) or to leave out (4 SDG default) the fourth SDG. In addition, respondents always had the answer option ‘No opinion’.

### The consequentiality of the choices

Participants of *either treatment* learned:

*“[...] **Your choice counts. If the majority** of respondents chooses to add (leave out) the fourth sustainable development goal, it will happen. The board of Pensioenfonds Detailhandel guarantees its implementation.”*

We explicitly stress the two key characteristics of their consequential choice: (1) their vote counts and (2) the board guarantees that the majority vote will be implemented. The vast majority of our participants (86.0%) understood the consequentiality of their choice as they correctly answered the following comprehension question (see Question 2 in Appendix A.4 and A.5):

- “If a majority chooses ‘Yes, add’ (‘Yes, leave it out’), Pensioenfonds Detailhandel...*
- a. guarantees to add (leave out) ‘Sustainable consumption and production’ to (of) its socially responsible investment policy as the fourth Sustainable Development Goal. The Board of Pensioenfonds Detailhandel has decided to implement the outcome of this vote.*
  - b. cannot guarantee that it will add (leave out) ‘Sustainable consumption and production’ to (of) its socially responsible investment policy as the fourth Sustainable Development Goal but may include the results of the survey in its choice.”*

The correct answer was ‘a’. In case, the participants incorrectly answered the question, they were made aware of their misperception.

Everything besides the necessary information within each treatment group was exactly the same. The binary nature of the choice for more or less sustainable investments is essential as it assures incentive compatibility (Cummings *et al.*, 1995, 1997; Harrison, 2006a; Carson and Groves, 2007). Providing participants with multiple options, like choosing between adding different SDGs, would have introduced incentives for strategic voting (see Arrow’s impossibility theorem; Arrow 2012). For example, if participants expect that it is unlikely that their first best choice will get a majority vote, they can strategically vote for the second-best option.

## 2.3 The data

### 2.3.1 The participants

Participants in our study are members of Pensioenfonds Detailhandel, a large Dutch pension fund. The Dutch state requires employees to save for their pension. It is mandatory for participants to contribute to the pension fund through their current employer. Of the invited 49,552 participants<sup>13</sup>, 1,669 (3.4%) answered the question on whether to add or leave out the fourth Sustainable Development Goal and 1,220 (2.5%) completed the survey which took on average 19.3 minutes. Table A2 presents summary statistics of the respondents. The sample consists of 60.9% women and 39.1% men, with an average age of 46.6 years. Slightly less than a third have a university degree or a degree from a university of applied sciences (29.9%). The average monthly net household income is 2,837 Euro.

### 2.3.2 Representativeness of our sample

Surveys face the risk of a biased response sample. This is of special concern in a setting like ours, where the topic is sustainable investing and people with stronger social preferences might be more likely to participate in the survey. Compared to the total population of the pension fund, our sample contains slightly fewer women (60.8 or 9% vs 66.5%) and is older (46.6 years compared to 42.6 years). Further, we compare the yearly gross income of each pension fund participant between our sample and the invited population. Participants in our sample earn with 24,934 euro slightly more than the invited sample (24,294 euro). We show below that men, older people, and those with a higher income are less likely to favor more sustainable investing. Hence, these differences go against our findings and thus cannot explain our results. Note that we use the personal yearly gross income from the administrative data provided by the fund. In subsequent analysis, we will use the self-reported monthly net household income as this provides the monetary basis for decision-making within a household. These household income levels were unavailable for the total population, because the pension fund only observes income saved by the individual and not by the household.

Second, we tested whether participants in our study have comparable political preferences to the Dutch population. We therefore asked participants to provide information on whether they voted in the Dutch national elections in 2017 and, in case they did, to name the party they voted for. Voting behavior is a good proxy for social preferences, because the relation between social preferences and voting behavior has been well-established. Fisman *et al.* (2017) show that equality-focused subjects were more likely to vote for Barack Obama in 2012 and are also more likely to be affiliated with the U.S. Democratic Party. Kerschbamer and Müller (2017) find that social preferences predict political attitudes. Almås *et al.* (2016) find conservatives to

<sup>13</sup>With half of the invited population, we ran a hypothetical treatment which is available upon request.

systematically accept a higher level of inequality than non-conservatives. Hong and Kostovetsky (2012) show that U.S. investment managers who donate to the Democratic Party are more likely to invest in companies which are seen to be socially responsible. Hence, by comparing the voting behavior of our sample to the outcome of the 2017 Dutch national elections, we can observe whether we attracted certain voters more than others resulting in a non-representative distribution of preferences.

Table A2.3 presents the voting behavior of our sample. 1,294 respondents answered the question of which 70 (5.4%) did not want to state whether they voted. The reported voter turnout was 84.5% which is close to the official turnout of 81.6%.<sup>14</sup> Further, 260 (24.9%) did not want to state the name of the party they voted for. The distribution of votes for the remaining 784 respondents deviates slightly from the official voting outcome. Respondents were more likely to vote for parties known to strongly advocate sustainability like the green party (GroenLinks, +4.9 percentage points higher than in the official voting outcome) and the animal party (Partij voor de Dieren, +2.8 percentage points). However, the reported voting behavior also clearly shows that our sample does not mainly consist of people with strong preferences in favor of sustainability. Further, the difference between the reported voting behavior in our sample and the official turnout regarding the votes for the PVV, a Dutch populist party<sup>15</sup>, is noticeable (-8.0 percentage points). This finding is in line with previous work showing that people refrain from admitting their support for extreme groups (Gingerich, 2010; Bullock *et al.*, 2011).

We show below that these deviations cannot explain our results as the majority of voters of all but one party (Forum voor Democratie) chose four SDGs. The reported voting behavior therefore is a strong indicator that we do not have a decisive selection bias of certain preferences which favor sustainable investing.

Hence, while our sample has small biases in terms of demographics, financial background, and social preferences, we will show below that these deviations cannot explain our results but sometimes even go against our results.

---

<sup>14</sup> The official 2017 parliamentary election results can be found online at <https://www.verkiezingsuitslagen.nl/verkiezingen/detail/TK20170315>.

<sup>15</sup> Its leader, Geert Wilders, "told supporters he would ensure there were 'fewer Moroccans' in the country" during an election campaign in 2014 (<https://www.theguardian.com/world/2014/mar/20/dutch-politician-geert-wilders-moroccans-outrage-pvv-party-anti-islam>). As a consequence, he was found guilty of inciting racial discrimination in December 2016 (<https://www.telegraph.co.uk/news/2016/12/09/anti-islam-dutch-mp-geert-wilders-found-guilty-discrimination/>). He has further mentioned that he wants to ban all mosques in the Netherlands (<https://www.independent.co.uk/news/world/europe/netherlands-pvv-leader-geert-wilders-koran-islam-mosque-ban-holland-dutch-pm-favourite-a7214356.html>).

## 2.4 Results

Two thirds (67.9%) of participants favored more sustainable investments and chose four SDGs (see Figure 2.1)<sup>16</sup>. Only 10.8% of respondents chose three SDGs, while 21.2% reported to have no opinion. Put differently, respondents in favor of four SDGs are 6.3 times as many as those against it.

**Result 1:** 67.9% of respondents are in favor of expanding sustainable investments.

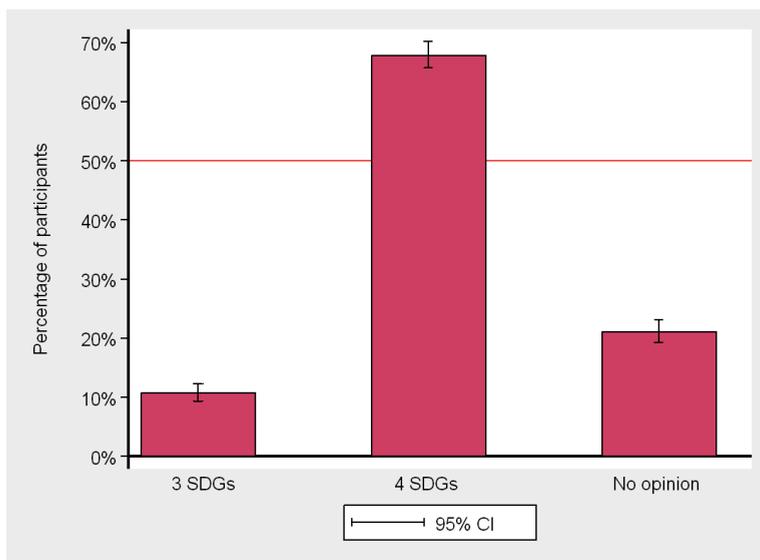


FIGURE 2.1 – The distribution of choices.

The graph presents the distribution of choices regarding the following question: “Do you want Pensioenfonds Detailhandel to add (leave out) the fourth sustainable development goal ‘Responsible consumption and production’?” Answer options were ‘Yes, add (leave it out)’, ‘No, do not add (leave it out)’, and ‘I have no opinion regarding this matter’. Words in parentheses refer to the different status quo treatments which are looked at jointly. Choices were guaranteed to be implemented by the pension fund if more than 50% of respondents would choose in favor of three, respectively four SDGs. Error bars present 95% confidence intervals.

What explains the support for sustainable investing? There are three potential explanations. First, people could decide to invest sustainably as they expect these investments to financially outperform conventional investments. If return expectations are sufficiently high, everybody would opt for more sustainable investments. Second, participants could have strong preferences for sustainable investments. Third, it could be that respondents were confused or unaware of the consequences of their choice. In the following sections, we show that the support for more sustainable investments is mainly caused by strong social preferences.

<sup>16</sup> For now, we do not differentiate between the 3 SDG default and 4 SDG default treatments and look at them jointly. We will get back to this in section 2.4.3.

### 2.4.1 Financial return expectations

In this section, we explore how expectations about financial performance influence sustainable investment choices. An important design feature is how to introduce costs and benefits of sustainable investments. We told participants that investing according to the Sustainable Development Goals means that not only the financial return counts. For those investments it is also important to consider the impact on the environment and society. We deliberately did not mention absolute costs or benefits of sustainable investments, because these could steer participants' choices. Moreover, the expected financial performance of sustainable investments is unclear. Several studies show that sustainable investments outperform conventional investments (Derwall *et al.*, 2005; Edmans, 2011), but other studies rather find an underperformance (Hong and Kacperczyk, 2009; Barber *et al.*, 2018; Riedl and Smeets, 2017; Hartzmark and Sussman, 2018). The return expectations of individual investors reflect these ambiguous findings regarding realized returns on sustainable investments. Some studies find investors to expect lower risk adjusted returns (Renneboog *et al.*, 2008; Riedl and Smeets, 2017). Others find investors to expect higher risk-adjusted returns on sustainable investments (Bauer and Smeets, 2015; Hartzmark and Sussman, 2018).

Rather than assuming a certain financial performance, we asked participants directly about their financial return expectations when the pension fund would focus on 3 or 4 SDGs respectively. Specifically, we asked participants:

*"When do you think the investment return is highest?"*

- a. The investment return is highest with 3 SDGs.*
- b. The investment return is equally high with 3 or 4 SDGs.*
- c. The investment return is highest with 4 SDGs.*
- d. I do not know."*

Figure 2.2(a) plots the distribution of financial beliefs within our sample. The majority of 57.5% of participants provided a return expectation; 14.8% of participants expect a lower return with four SDGs (answer 'a. '), 17.3% expect a higher return, and 25.4% of respondents think that there is no difference in returns between investing in three or four SDGs. Return expectations are clearly not overly favorable toward sustainable investing and are thus unable to explain that two thirds choose to expand sustainable investments at their pension fund. Figure 2.2(b) supports this claim. While we find stronger support for four SDGs if return expectations are more positive, the majority is in favor of four SDGs independent of return expectations. Two findings are especially noteworthy. First, even among those who expect four SDGs to yield lower returns, 57.7% chose four SDGs and only 30.3% chose three SDGs. Thus, even when people expect to be worse off, they favor the more sustainable option. Second, literature shows that people do not like uncertainty and therefore avoid outcomes which increase uncertainty

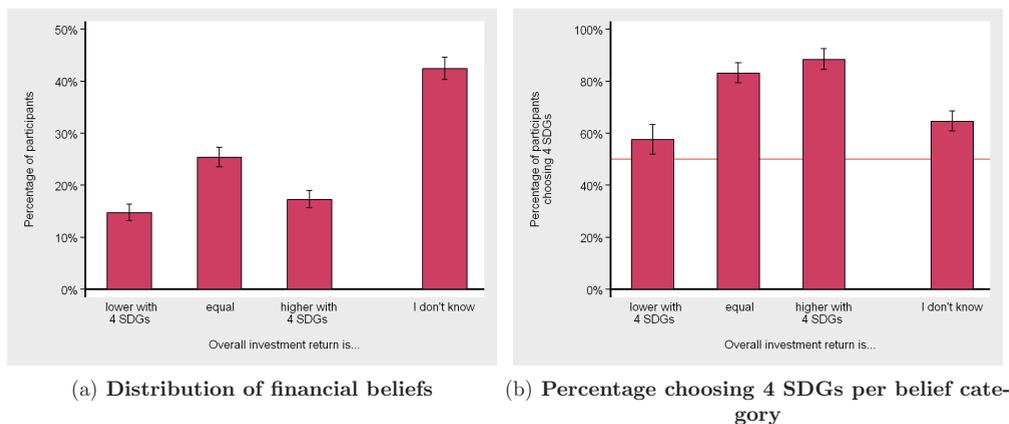


FIGURE 2.2 – Financial beliefs.

Figure 2.2(a) shows the distribution of financial beliefs. Figure 2.2(b) presents the percentage of participants choosing four SDGs for each return expectation separately. Return expectations were elicited by asking participants to answer the following question: “When do you think the investment return is highest?” Answer options were ‘It is highest with 3 SDGs’, ‘The return is equally high with 3 or 4 SDGs’, ‘It is highest with 4 SDGs’, and ‘I do not know’. The categories shown in Figure 2B refer to the following question: “Do you want Pensioenfond Detailhandel to add (leave out) the fourth sustainable development goal ‘Responsible consumption and production’?”. Answer options were ‘Yes, add (leave it out)’, ‘No, do not add (leave it out)’, and ‘I have no opinion regarding this matter’. Words in parentheses refer to the different status quo treatments which are looked at jointly here. Choices were guaranteed to be implemented by the pension fund if more than 50% of respondents would choose in favor of three, respectively four SDGs. Error bars present 95% confidence intervals.

(e.g. Ellsberg 1961; Mukerji and Tallon 2001; Charness and Gneezy 2010). We find that of the participants who do not know what returns to expect, 64.7% favor four SDGs. While this number has to be treated with caution as the uncertain option depends on the default treatment, we find that the majority of these subjects chooses four SDGs irrespective of their default (see Figure A.1).<sup>17</sup> Thus, the majority is in favor of four SDGs independent of return expectations, even if people do not know what to expect.

All in all, there is no evidence that participants only support more sustainable investments because they see it as a free lunch.

**Result 2:** The majority of respondents chooses to expand sustainable investing; this is the case even among those who have negative return expectations or are uncertain about the return.

<sup>17</sup> People in the 3 SDG default should consider the choice of three SDGs to be less uncertain as this is the default option while those in the 4 SDG default should consider the choice of four SDGs to be less uncertain. For those who do not have a return expectation for sustainable investments, we therefore expect that the support for four SDGs is lower in the 3 SDG default than in the 4 SDG default. Figure A.1 shows the default effect for those participants who did not know what return to expect (43.6% of the entire population, see Figure 2.2(a)). As hypothesized, we do find a significant difference where 59.4% of participants in the 3 SDG default and 65.8% of those in the 4 SDG default choose four SDGs (one-sided T-test; p-value = 0.0107). Yet, it is remarkable that even for those who consider choosing four SDGs as the option with higher uncertainty, we find a clear majority to opt for the more sustainable option.

In the following subsection, we show that the effect of social preferences on the choice for four SDGs is indeed unaffected when we control for return expectations in a multinomial logit regression.

## 2.4.2 Strong social preferences

We measure social preferences using the validated social preferences measure of Falk *et al.* (2016) "*How willing are you to give to good causes without expecting anything in return?*". Participants rate their agreement on a 10-point Likert scale from completely unwilling to very willing. Higher ratings correspond to higher levels of social preferences. The average level of social preferences in our sample is 6.1 on a 10-point Likert scale (for summary statistics, see Table A2.2). We investigate to what extent social preferences predict the support for more sustainable investments. Table 2.1 presents relative-risk ratios of a multinomial logit regression that regresses the support for four sustainable development goals on our measure of social preferences. The dependent variable takes on three values for the choices '3 SDGs', '4 SDGs', and 'No opinion'. As baseline, we take respondents who chose three SDGs. To interpret the relative-risk ratios, consider the effect of *Gender (female)* on the choice for 4 SDGs. If the coefficient equals one, men and women are equally likely to choose 4 SDGs. A coefficient larger than one indicates that women are more likely to choose 4 SDGs. Similarly, a coefficient smaller than one indicates that women are less likely to choose 4 SDGs.

In the first column of specification (1), we study the effect of social preferences on the likelihood to choose four SDGs. A one standard deviation increase in social preferences increases the likelihood to choose four SDGs by 53.6%. Given that 67.9% of our sample chose four SDGs, social preferences have a substantial economic impact on sustainable investment decisions. Social preferences are not related to the choice of no opinion.

In specification (2), we control for financial beliefs, demographics, and information on participants' financial background. As anticipated, expecting lower returns with four SDGs decreases the likelihood to choose four SDGs by 79.0% while expecting higher returns more than doubles the likelihood. Importantly, social preferences predict the support for four SDGs when controlling for return expectations and demographics. The effect size of social preferences does not decrease but, on the contrary, even increases slightly ( $\beta = 1.5874$ ).

**Result 3:** Social preferences predict the choice for more sustainable investments; this relation is robust to including financial beliefs. This is consistent with a preference-based explanation for sustainable investments.

Looking more closely at the effects of demographics on sustainable investments in Table 2.1, we see that women are significantly more likely to choose four SDGs compared to men.

Women in our sample also have stronger social preferences than men (F: 6.3 vs M: 5.9; p-value = 0.0124). These results are in line with previous literature findings that women sometimes have stronger social preferences (see e.g. Bolton and Katok 1995; Eckel and Grossman 1998; Güth *et al.* 2007; for a meta-analysis see Croson and Gneezy 2009). Remember that our sample has slightly more men compared to the pension fund's population. The support we find for more sustainable investments therefore is on the conservative side. Noteworthy is that even among men the majority opts for more sustainable investments (69.1%). Second, age is related to the respondents' choice for sustainable investing. Older people are less likely to choose four SDGs. Each 10 years of age decrease the likelihood to choose four SDGs by 15.5%. Given that our respondents are slightly older than the invited sample and older people are less in favor of sustainable investments, the support for sustainable investments is a lower-bound. Third, education has no impact on whether respondents choose three or four SDGs. Not surprisingly, highly educated respondents are 47.7% less likely to have no opinion. And finally, people with a higher income are slightly less likely to choose four SDGs.

### 2.4.3 Default effects

The status quo bias could hide people's actual preferences (e.g. Samuelson and Zeckhauser 1988; Tversky and Kahneman 1991). Consider receiving investment advice, where the default option is to not focus on sustainable investments. Those who prefer to invest sustainably but fear that returns will be too low stay with the default and do not reveal their actual preferences. With our different treatments, we are able to control for status quo effects.

Table 2.2 shows that the status quo does not affect the choice for more sustainable investments. We again use a multinomial logit regression where the dependent variable takes on the three values '3 SDGs', '4 SDGs', and 'No opinion'. If a status quo bias were present, we would expect respondents in the 4 SDG default treatment to be more likely to support four SDGs than respondents in the 3 SDG default. Accordingly, the coefficient of '4 SDG default' in the first column of specification (1) should be significantly greater than one. The non-significance of the coefficient ( $\beta = 0.8013$ ) shows that this is clearly not the case. In specification (2), we include financial beliefs, demographics, and information on the financial background as controls. The coefficient of the '4 SDG default' dummy slightly increases to  $\beta = 1.0133$  and stays non-significant (difference from one: p-value = 0.948). Thus, framing has no effect on the choice between three or four SDGs.

More importantly, the effect size and the significance level of social preferences is not affected by controlling for default effects which indicates that strong preferences drive the decision to invest sustainably. An individual with strong preferences for sustainable investing will choose more sustainable investments, irrespective of what the default option is.

TABLE 2.1 – The effect of social preferences and financial beliefs on sustainable investing.

This table presents relative-risk ratios of a multinomial logit regression where the dependent variable takes on three values for the choices ‘3 SDGs’, ‘4 SDGs’, and ‘I have no opinion regarding this matter’. Baseline are respondents who chose three SDGs. *Social preferences* are elicited by asking respondents to rate the question “How willing are you to give to good causes without expecting anything in return?” on a 10-point Likert scale from completely unwilling to completely willing (see Falk *et al.* 2016). *Financial beliefs* were elicited by asking participants to answer the following question: “When do you think the investment return is highest?” Answer options were ‘It is highest with 3 SDGs’, ‘The return is equally high with 3 or 4 SDGs’, ‘It is highest with 4 SDGs’, and ‘I do not know’. For summary statistics, see Table A2.2, and for definitions of the other variables, see Table A2.1. Significance levels show the difference from one and are indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

	(1)		(2)	
	4 SDGs	No opinion	4 SDGs	No opinion
<b><i>Preferences</i></b>				
Social preferences (std)	1.5361*** (0.1398)	0.8742 (0.0931)	1.5874*** (0.1571)	0.9559 (0.1119)
<b><i>Financial beliefs</i></b>				
Lower returns with 4 SDGs			0.2095*** (0.0571)	0.3897** (0.1476)
Higher returns with 4 SDGs			2.0609* (0.8334)	1.5150 (0.7590)
Do not know			0.8837 (0.2431)	2.8393*** (0.9476)
<b><i>Demographics</i></b>				
Female			2.00355*** (0.4208)	1.6918** (0.4229)
Age			0.9845** (0.0074)	0.9688*** (0.0086)
Highly educated			0.9170 (0.2096)	0.5324** (0.1525)
<b><i>Financial background</i></b>				
Low monthly income			0.7440 (0.1927)	0.8554 (0.2728)
High monthly income			0.7147 (0.2525)	1.1187 (0.4992)
Do not report			0.5544** (0.1578)	1.0735 (0.3603)
Constant	7.3652*** (0.7084)	1.4712*** (0.1736)	20.6432*** (10.2167)	4.3091** (2.5298)
Observations	1,280	1,280	1,280	1,280

**Result 4:** Respondents’ choice of expanding sustainable investments does not depend on how their choice is framed (no status quo dependency). The support for more sustainable investments is equally high, consistent with strong preferences.

### 2.4.4 External validation: voting behavior and sustainable investments

To test for external validity of our results, we explore the relation between the support for more sustainable pension investments and voting behavior in the last national parliamentary elections. We construct respondents' preferences for sustainable policies by evaluating the stance on sustainability of the political party they voted for. We use the official voting guide for the 2017 Dutch national elections.<sup>18</sup> Five of the 30 questions posed in this voting guide deal with issues related to sustainability. All questions are posed such that they can be answered with 'agree', 'disagree', or 'neither of the two'. A party receives one point for each question for which they choose the more sustainable option. If a party chooses 'neither of the two', it receives half a point. For example, consider the statement "All coal-fired power stations may remain open for the time being." If a party agrees to this statement, it receive zero points; if a party disagrees it receives one point; and if it says neither of the two, it receives half a point. Each political party thus receives a score between zero and five. The entire construction of the score is presented in Table A2.4.

Table A2.5 shows that voters of political parties with a stronger preference for sustainable policies show more support for four SDGs. Especially, voters of the animal party (Partij voor de Dieren) and the green party (GroenLinks) favor four SDGs (91.5% and 89.1% respectively). On the other side, voters of parties like the Volkspartij voor Vrijheid en Democratie (VVD) or the Partij voor de Vrijheid (PVV) with a sustainability score of zero show one of the lowest acceptance rates for four SDGs. Nevertheless, across voters for all political parties more than 50% choose four SDGs (except for the party 'Forum voor Democratie'). Also among those who reported not to have voted in the last national elections, 67.9% are in favor of four SDGs. Among those who do not state the party they voted for, the majority of 68.1% chose four SDGs.

The first specification in Table 2.3 shows that a respondent who voted for a party with a one standard deviation stronger preference for sustainable policies is 56.7% more likely to also choose the more sustainable option in our experiment. When we include control variables in specification (2), the effect size slightly decreases to 46.7% and stays highly significant. These results show the external validity of the support for four sustainable development goals. Participants took their choice seriously and chose more or less sustainability in line with how they evaluated political parties on sustainability in the last national election. Remarkably, even when we control for participants' preferences for sustainable policies, the effect size of social preferences barely changes compared to Tables 2.1 and 2.2. This underlines the importance of strong social preferences as explanation for the choice to invest more sustainably.

<sup>18</sup> See <https://tweedekamer2017.stemwijzer.nl>.

TABLE 2.2 – Default effects and sustainable investing.

This table presents relative-risk ratios of a multinomial logit regression where the dependent variable takes on three values for the choices ‘3 SDGs’, ‘4 SDGs’, and ‘I have no opinion regarding this matter’. Baseline are respondents who chose three SDGs. For summary statistics, see Table A2.2. *Social preferences* are elicited by asking respondents to rate the question “How willing are you to give to good causes without expecting anything in return?” on a 10-point Likert scale from completely unwilling to completely willing (see Falk *et al.* 2016). 3 SDG default and 4 SDG default refer to two different status quo treatments to which participants were randomly assigned. In the *3 SDG default*, participants were told that the pension fund currently focuses on three SDGs. Participants were introduced to the fourth SDG ‘Responsible consumption and production’ and could choose: “Do you want Pensioenfond Detailhandel to add the fourth sustainable development goal ‘Responsible consumption and production’?”. Answer options were ‘Yes, add’, ‘No, do not add’, and ‘I have no opinion regarding this matter’. In the *4 SDG default* treatment, it was communicated that the pension fund’s future policy will include a fourth SDG ‘Responsible consumption and production’. Participants could then choose: “Do you want Pensioenfond Detailhandel to leave out the fourth sustainable development goal ‘Responsible consumption and production’?”. Answer options were ‘Yes, leave it out’, ‘No, do not leave it out’, and ‘I have no opinion regarding this matter’. Choices were guaranteed to be implemented by the pension fund if more than 50% of respondents would choose in favor of three, respectively four SDGs. For definitions of the other variables, see Table A2.1. Significance levels show the difference from one and are indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

	(1)		(2)	
	4 SDGs	No opinion	4 SDGs	No opinion
<i>Preferences</i>				
Social preferences (std)			1.5874*** (0.1575)	0.9490 (0.1114)
<i>Treatment effects</i>				
4 SDG default	0.8013 (0.1286)	0.7646 (0.1401)	1.0133 (0.2061)	0.7811 (0.1907)
Constant	7.0119*** (0.8178)	2.2381*** (0.2937)	20.5008*** (10.3080)	4.8554*** (2.8993)
Observations	1,669	1,669	1,212	1,212
Financial beliefs	No	No	Yes	Yes
Demographics	No	No	Yes	Yes
Financial background	No	No	Yes	Yes

**Result 5:** Respondents who voted for a party which cares for sustainability are more likely to expand sustainable investing. This finding provides external validity consistent with a preference-based explanation for sustainable investments.

### 2.4.5 Confusion and misunderstanding do not explain our results

The strong support for sustainable investing could be explained by respondents being confused or lacking information. We address this by re-analyzing our results, excluding: (1) Respondents who started to vote, but dropped out of the survey before finishing; (2) Respondents who did not understand the treatment and failed our comprehension question (see section 2.2); (3) Respondents who reported that they would have liked to receive more information before deciding on whether to choose three or four SDGs.

TABLE 2.3 – The relation between voting behavior and sustainable investing.

This table presents relative-risk ratios of a multinomial logit regression where the dependent variable takes on three values for the choices ‘3 SDGs’, ‘4 SDGs’, and ‘I have no opinion regarding this matter’. Baseline are respondents who chose three SDGs. For summary statistics, see Table A2.2. *Social preferences* are elicited by asking respondents to rate the question “How willing are you to give to good causes without expecting anything in return?” on a 10-point Likert scale from completely unwilling to completely willing (see Falk *et al.* 2016). *Preferences for sustainable policies* is a score which is constructed by evaluating Dutch parties on their stance on sustainability using the official voting guide for the 2017 Dutch national elections (<https://tweedekamer2017.stemwijzer.nl>). Five of the 30 questions shown deal with sustainable issues. All questions are posed such that they can be answered with ‘agree’, ‘disagree’, or ‘neither of the two’. A party receives one point for each question for which they choose the more sustainable option. If a party chooses ‘neither of the two’, it receives half a point (see Table A2.4). 3 SDG default and 4 SDG default refer to two different status quo treatments to which participants were randomly assigned. In the *3 SDG default*, participants were told that the pension fund currently focuses on three SDGs. Participants were introduced to the fourth SDG ‘Responsible consumption and production’ and could choose: “Do you want Pensioenfonds Detailhandel to add the fourth sustainable development goal ‘Responsible consumption and production’?”. Answer options were ‘Yes, add’, ‘No, do not add’, and ‘I have no opinion regarding this matter’. In the *4 SDG default* treatment, it was communicated that the pension fund’s future policy will include a fourth SDG ‘Responsible consumption and production’. Participants could then choose: “Do you want Pensioenfonds Detailhandel to leave out the fourth sustainable development goal ‘Responsible consumption and production’?”. Answer options were ‘Yes, leave it out’, ‘No, do not leave it out’, and ‘I have no opinion regarding this matter’. Choices were guaranteed to be implemented by the pension fund if more than 50% of respondents would choose in favor of three, respectively four SDGs. For definitions of the other variables, see Table A2.1. Significance levels show the difference from one and are indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

	(1)		(2)	
	4 SDGs	No opinion	4 SDGs	No opinion
<b>Preferences</b>				
Social preferences (std)			1.5129*** (0.2030)	0.8932 (0.1461)
Preferences for sustainable policies (std)	1.5668*** (0.1961)	1.0712 (0.1685)	1.4671*** (0.2088)	1.0268 (0.1815)
<b>Treatment effects</b>				
4 SDG default & Real			1.1990 (0.3198)	0.7955 (0.2625)
Constant	8.2138*** (1.0484)	1.2735*** (0.2047)	19.2319*** (12.8936)	5.5426** (4.5371)
Observations	772	772	731	731
Financial beliefs	No	No	Yes	Yes
Demographics	No	No	Yes	Yes
Financial background	No	No	Yes	Yes

We successively exclude each of the three types. The results are presented in Figure 2.3. The first bar represents the choice distribution of all participants, without exclusion. The second bar shows the choice distribution when respondents are excluded who did not complete the survey. The dropout rate was 26.9%. The third bar shows the choice distribution excluding participants who did not understand the treatment they were in. This is the case for 14.0% of participants. The fourth bar shows the distribution of choices excluding participants who wanted more information for their decision (33.5%). Importantly, the support for four SDGs is

unaffected by either of the exclusion criteria ranging between 67.9% (all included) and 73.9% (only those who completed the survey).

**Result 6:** Neither dropouts, confusion, nor a lack of information affect our results. There is always a clear majority supporting more sustainable investments.

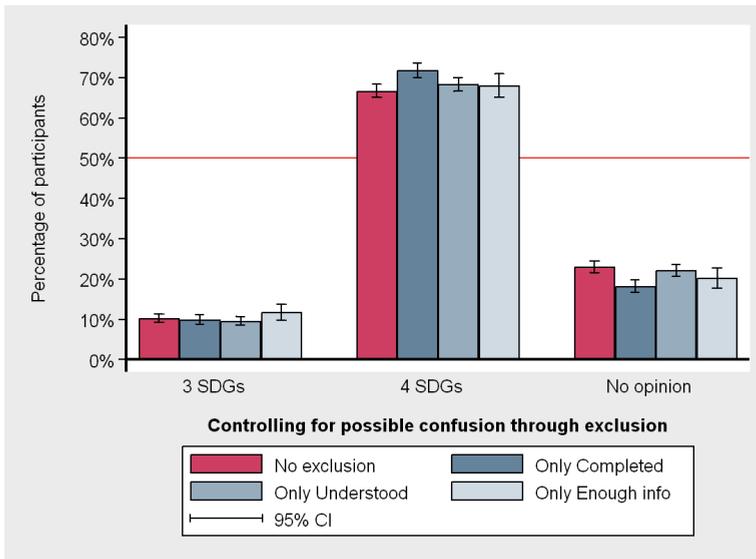


FIGURE 2.3 – Controlling for possible confusion.

The graph presents the distribution of all respondents (leftmost bar) and of three restricted subsamples separately regarding the following question: “Do you want Pensioenfonds Detailhandel to add (leave out) the fourth sustainable development goal ‘Responsible consumption and production’?”. Answer options were ‘Yes, add (leave it out)’, ‘No, do not add (leave it out)’, and ‘I have no opinion regarding this matter’. Words in parentheses refer to the different status quo treatments which are looked at jointly here. Choices were guaranteed to be implemented by the pension fund if more than 50% of respondents would choose in favor of three, respectively four SDGs. The second bar from the left (dark blue) presents the distribution of choices where respondents are excluded who dropped out of the survey after giving a choice on sustainable investing but before the end of the survey. The second bar from the right (medium blue) presents the distribution of choices where respondents are excluded wrongly answered the control question “If a majority chooses ‘yes, add’ (‘no, leave it out’), a. Pensioenfonds Detailhandel guarantees to add (leave out) the fourth SDG; b. Pensioenfonds Detailhandel cannot guarantee that it will add (leave out) the fourth SDG, but may include the results of the survey in its choice.” The rightmost bar (light blue) presents the distribution of choices where respondents are excluded who report that they would have liked to receive more information before deciding on whether to choose three or four SDGs. Error bars present 95% confidence intervals.

## 2.5 Additional findings

### 2.5.1 Social signaling

Social signaling can play a significant role when deciding whether to invest sustainably (Riedl and Smeets, 2017). It is the notion that people prefer to be seen by others as prosocial individuals (Glazer and Konrad, 1996; Bénabou and Tirole, 2006; Ellingsen and Johannesson, 2008; Ariely *et al.*, 2009; Tonin and Vlassopoulos, 2013; Cappelen *et al.*, 2017). To study the effect of social signaling, we use a modified version of the question asked in Riedl and Smeets (2017). Specifically, we asked respondents to rate the question “How likely is it that you will talk to your friends about this survey?” on a 10-point Likert scale from 1 – very unlikely to 10 – very likely. We expect respondents with a higher desire of signaling their pro-sociality to be slightly more likely to choose four SDGs. We therefore split the subjects into a group with an above-median social signaling desire (ratings of 3 and higher, 58.4% of the sample) with a below-median social signaling desire (41.6%).

The results of testing for social signaling are presented in Figure A2. We find a significant difference as 78.3% of those with a high signaling desire and 67.5% of those with a low signaling desire chose four SDGs ( $p$ -value = 0.0000). Yet, within both groups the majority chose to expand sustainable investing which indicates that social signaling matters but cannot explain the large support for more sustainable investments.

### 2.5.2 Pivotal voting

There is a possibility that subjects in our experiment did not perceive their choice to be consequential. They might believe that their vote is unlikely to be pivotal. In that case, the real choice could simply be perceived as a hypothetical choice, which in turn could explain the large support for more sustainable investments. Several aspects render it unlikely that this potential concern is driving our results.

First, voting theory shows that the use of a voting mechanism with a simple majority rule is incentive-compatible if the choice is binary, consequential, and if subjects care about the outcome (Cummings *et al.*, 1997; Harrison, 2006a). The first two conditions are met as part of the experimental setup. Further, we can expect that the third condition is met by the fact that people’s pension money is at stake. Thus, the choices selected by our participants maximized their personal utility, which implies that they reported their preferences honestly.

Second, the paradox of voting predicts that rational people do not vote as their expected benefits of casting the pivotal vote are lower than their expected costs (Downs, 1957). For example, Mulligan and Hunter (2003) find that only one of every 89,000 votes casted in U.S.

Congressional elections, and one of 15,000 in state legislator elections were pivotal<sup>19</sup>. Nevertheless, turnout rates in political elections are usually quite high, which makes it at least difficult to justify that voters do not care about their vote when it is unlikely to be pivotal.

Third, even though theory and field evidence suggest that people take voting seriously, it is possible that in our experiment this was not the case. We therefore distinguish between participants who expected to be more pivotal and those who expected to be less pivotal. If participants only supported more sustainable investments when they perceived their vote not to be pivotal, this would cast doubt on our finding that a majority prefers more sustainable investments. We tackle this concern in two different ways. First, we asked participants to estimate the percentage of participants who would choose three SDGs, four SDGs, or opt for ‘no opinion’. Subjects who predicted that the share of those in favor of four SDGs is close to 50% should perceive their vote to be more pivotal than subjects with a prediction far away from 50%. Following the just outlined expectation, we would expect a gap to arise for subjects with a prediction close to 50% compared to those who have predictions far away from 50%. We define ‘pivotal’ voters as those who predict the share of respondents choosing four SDGs to be between 40% and 60% and ‘non-pivotal’ voters as those with predictions below 40% or above 60%. This leaves us with 458 ‘pivotal’ subjects (31.6% of the entire sample) and 991 ‘non-pivotal’ voters (68.4%). We find no difference between pivotal and non-pivotal voters as 73.6% of ‘pivotal’ and 70.4% of ‘non-pivotal’ subjects chose four SDGs (p-value = 0.2179).<sup>20</sup> Second, we asked subjects to estimate how many of the 25,000 invited people would participate in the experiment.<sup>21</sup> The median estimate was 10,000 participants. We define a ‘pivotal’ voter as someone who estimates that less than 10,000 would participate (meaning that the voter has a relatively higher probability to be pivotal). This leaves us with 586 ‘pivotal’ (52.0%) and 635 ‘non-pivotal’ subjects (48.0%). Again, we do not find a gap as 72.5% respectively 75.1% subjects choose four SDGs (p-value = 0.3033).<sup>22</sup>

As additional check, we look at the two above-mentioned definitions of a ‘pivotal’ voter jointly. There are 192 subjects who predict the share of respondents choosing four SDGs to be between 40% and 60% while at the same time estimating less than 10,000 would participate in the survey (‘double-pivotal’) while there are 437 subjects who predict the share to be less than 40% more than 60% while at the same time estimating more than 10,000 would participate (‘double-non-pivotal’). We again find no gap (percentage choosing four SDGs: ‘double-pivotal’

---

<sup>19</sup> A pivotal vote in this case means that the candidate of that specific election tied for first or won by a single vote.

<sup>20</sup> This finding is robust to altering the definition of a “pivotal” voter. We also do not find a difference when we look at those who predict the share of respondents choosing four SDGs to be between 45% and 55% as well as between 35% and 65%.

<sup>21</sup> We gave 25,000 as reference point as this was the number of participants which was invited to participate in these two treatments. With the other half, we ran a hypothetical treatment which is available upon request.

<sup>22</sup> This finding is robust to altering the definition of a “pivotal” voter. We also do not find a difference when we look at those who estimate that less than 8,000 (bottom 45%) as well as less than 11,000 (bottom 65%) would participate.

71.9% vs ‘double-non-pivotal’ 74.1%; p-value = 0.5540). Taken together, we can rule out that participants support more sustainable investments, because they think that their vote is not pivotal.

### 2.5.3 Exclusion of non-sustainable companies

Investing according to the SDGs offers different approaches. While our question on adding a fourth SDG focused on engagement, we also elicited participants’ preferences for exclusion policies. We told participants that exclusion means that the pension fund will refrain from investing in non-sustainable companies. It was explicitly mentioned that the questions on the exclusion policy in this part of the survey were of hypothetical nature but that the pension fund is interested in learning about their opinion. Thus, respondents’ choices were no longer consequential. Specifically, respondents were asked to choose from a list of topics. Multiple answers were allowed. Each item they could choose indicated that they would want to exclude companies which relate to this item. For example, respondents could state that they want their pension fund to exclude companies which produce tobacco or alcohol (for a full list, see Table A2.1). Respondents could also choose to not exclude any company.

Table A2.6 presents the answer distribution of the exclusion question. Column (1) shows the average answer over all respondents. In columns (2) to (4), respondents are split into the three answer categories ‘4 SDGs’, ‘3 SDGs’, and ‘no opinion’ which correspond to whether they chose three or four SDGs or did not have an opinion. It shows that the willingness to exclude companies is in general high. Across all three answer categories, respondents chose to exclude companies on seven of nine items with more than 50%. We observe the highest rates of approval for exclusion due to corruption (86.9%), child labor (86.2%), and human rights violations (85.3%). Importantly, this table also shows that people do not simply choose more sustainable investments whenever given a choice. Only a minority is in favor of excluding companies that produce alcohol (17.8%) and tobacco (44.0%). Even among those who chose four SDGs, those numbers are with 19.4% and 47.4% clearly below the threshold of 50%. Given the strong pressure that is currently being put on pension funds to divest from tobacco, it is remarkable to see that these pension fund members decide to still keep their tobacco investments.

## 2.6 Conclusion

The United Nations Sustainable Development Goals (SDGs) have created societal and political pressure for companies and investors to promote sustainability. We ran a field experiment with a pension fund which granted its members a real vote on its sustainable investment policy. A majority of two thirds of the plan participants are willing to support increased investments based on the SDGs. Most participants believe that more sustainable investments do not come

at the expense of financial returns. But even among those who do expect a reduction in financial returns, the majority wants to put their pension money on the table to promote sustainability. A key reason is participants' strong social preferences.

Social preferences have a significant influence on many economic decisions. They are key to understanding what incentivizes people (Fehr and Fischbacher, 2002; Bandiera *et al.*, 2005), can decrease free riding problems (Fehr and Gächter, 2000), and explain why people are willing to sacrifice resources to increase social welfare (Andreoni, 1990; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). Yet, it is not clear whether social preferences matter in financial markets (Dufwenberg *et al.*, 2011; Falk and Szech, 2013; Bartling *et al.*, 2014; Sobel, 2015). This could be the reason why European investment managers elicit clients' risk preferences but mostly ignore social preferences (EUROSIF, 2016, p. 82). This raises a simple question: why the reluctance towards social preferences? While previous research shows the importance of social preferences for some investors, it has been hard for institutional investors to use these findings when deciding on behalf of their clients. For instance, Bollen (2007); Hong and Kostovetsky (2012); Barber *et al.* (2018) and Hartzmark and Sussman (2018) show that investors value sustainability. Yet, consider a pension fund manager: with aggregate-level data, he will have trouble to identify the preferences for sustainable investments of his clients. And even with findings on the individual level, the pension fund manager would not know what to do. Take the findings of Riedl and Smeets (2017) who show that 16% of their sample holds a sustainable mutual fund. The pension fund manager could now say that the majority does not want to hold sustainable mutual funds. Case closed. Yet, Beshears *et al.* (2008) show that inertia often keeps people from acting according to their true preferences. Thus, fund managers have to elicit the social preferences of their clients themselves and act accordingly.

We offer a simple method for institutional investors to cater to the social preferences of their clients. Our study took place in the Netherlands, a country known for its large share of assets invested sustainably (EUROSIF, 2016), Cultural differences can have an important influence on economic decisions (e.g. Guiso *et al.* 2006). For example, European institutional investors have a stronger influence on environmental and social performance of countries than do American institutional investors (Dyck *et al.*, 2018). Why is this the case? Are American pension fund participants less willing to support sustainable investments or do fund managers impose their own political preferences onto their clients' investments (see Hong and Kostovetsky 2012)? An important avenue for future research is to investigate the preferences for sustainable investments in other countries and different populations. It will thereby be crucial to study real choices with actual consequences.

Outside of the investment industry, our proposed method can inform decision-makers who decide on behalf of others. Politicians could elicit social preferences to address key sustainable policy questions. If politicians were to rely on real preferences, this could create a wider public support for the outcome—even if it is costly. *But for that to be true, politicians have to get real!*

# A Appendix

## A.1 Tables

TABLE A2.1 – Variable definition

Variable	Description
<i>Preferences</i>	
Social preferences	The participant's response to the question "How willing are you to give to good causes without expecting anything in return?" (1 completely unwilling, . . . , 10 completely willing; Falk <i>et al.</i> 2016).
Preferences for sustainable policies	Preference for sustainable policies is a score which is constructed by evaluating Dutch parties on their stance on sustainability using the official voting guide for the 2017 Dutch national elections ( <a href="https://tweedekamer2017.stemwijzer.nl">https://tweedekamer2017.stemwijzer.nl</a> ). Five of the 30 questions shown deal with sustainable issues. All questions are posed such that they can be answered with 'agree', 'disagree', or 'neither of the two'. A party receives one point for each question for which they choose the more sustainable option. If a party chooses 'neither of the two', it receives half a point.
<i>Social signaling</i>	
Social signaling	The participant's response to the question "How likely is it that you will talk to your friends about this survey?" (1 very unlikely, . . . , 10 very likely; adaption from Riedl and Smeets 2017).
<i>Financial beliefs</i>	
Expected returns with 4 SDGs	The participant's response to the question "When do you think the investment return is highest?" Answer options: 'It is highest with 3 SDGs', 'The return is equally high with 3 or 4 SDGs', 'It is highest with 4 SDGs', and 'I do not know'.
Lower returns with 4 SDGs	Dummy variable equal to one if the participant believes that the returns is highest with three SDGs.
Equal returns with 3 or 4 SDGs	Dummy variable equal to one if the participant believes that the returns is equally high with three or four SDGs.
Higher returns with 4 SDGs	Dummy variable equal to one if the participant believes that the returns is highest with four SDGs.
Do not know	Dummy variable equal to one if the participant reports to not know when the return is the highest.
<i>Treatments</i>	
3 SDG default	Participants in the 3 SDG default treatment were told that the pension fund currently focuses on three SDGs. Participants were introduced to the fourth SDG 'Responsible consumption and production' and could answer the question: "Do you want Pensioenfondsen Detailhandel to add the fourth sustainable development goal 'Responsible consumption and production'?". Answer options: 'Yes, add', 'No, do not add', and 'I have no opinion regarding this matter'.
4 SDG default	Participants in the 4 SDG default treatment were told that the pension fund's future policy will include a fourth SDG 'Responsible consumption and production'. Participants could then choose: "Do you want Pensioenfondsen Detailhandel to leave out the fourth sustainable development goal 'Responsible consumption and production'?". Answer options: 'Yes, leave it out', 'No, do not leave it out', and 'I have no opinion regarding this matter'.

*Continued on next page*

Table A2.1—*Continued*

Variable	Description
<b><i>Demographics</i></b>	
Female	Dummy variable equal to one if the participant reports being a woman.
Age	The participant's self-reported age.
Highly educated	Dummy variable equal to one if the participant reports having a university degree or a degree from a university of applied sciences.
<b><i>Financial background</i></b>	
Individual income (gross, yearly)	Annual gross salary of the pension participant provided by the pension fund (administrative data)
Household income (net, monthly)	Self-reported monthly net household income (survey data)
Low income	Dummy variable equal to one if the participant's reported monthly net household income is below 2,500 euro.
Middle income	Dummy variable equal to one if the participant's reported monthly net household income is between 2,500 euro and 4,000.
High income	Dummy variable equal to one if the participant's reported monthly net household income is above 4,000 euro.
Do not report	Dummy variable equal to one if the investor does not disclose his or her monthly net household income.
<b><i>Comprehension</i></b>	
Understood treatment (Yes)	Dummy variable equal to one if the participant correctly answers the control question (correct answer 'a'): "If a majority chooses 'Yes, add' ('Yes, leave it out'), Pensioenfonds Detailhandel a. guarantees to add (leave out) 'Sustainable consumption and production' to (of) its socially responsible investment policy as the fourth Sustainable Development Goal. The Board of Pensioenfonds Detailhandel has decided to implement the outcome of this vote. b. cannot guarantee that it will add (leave out) 'Sustainable consumption and production' to (of) its socially responsible investment policy as the fourth Sustainable Development Goal but may include the results of the survey in its choice."
More information (Yes)	Dummy variable equal to one if the participant answers the question "Would you rather have received more information to make your choice on adding (leaving out) the fourth SDG?" with 'yes'. Answer options: 'Yes', 'No'
<b><i>Exclusion policy</i></b>	
Exclusion	The participant's response to the question "Pensioenfonds Detailhandel wants to know which companies it should exclude from its investments (multiple answers possible). Companies... a. producing tobacco; b. producing controversial weapons (e.g. cluster bombs and landmines); c. producing alcohol; d. producing weapons (weapons other than controversial weapons); e. violating human rights; f. using forced labor; g. having a bad influence on the environment; h. involving in corruption, extortion or bribery; i. allowing the use of child labor; j. I do not want to exclude companies."

TABLE A2.2 – Summary statistics.

This table presents summary statistics for the survey respondents. Variable definitions are provided in Table A2.1. Differences in the number of observations stem from the fact that not all participants in the experiments answered all survey questions.

	Mean	Median	Std Dev	Obs.
<b><i>Preferences</i></b>				
Social preferences (1–10)	6.1	7	2.5	1,280
Preferences for sustainable policies (0–5)	2.5	4	2.1	772
<b><i>Financial beliefs</i></b>				
Return expectation				
Lower with 4 SDGs	14.8%			208
Equal with 3 or 4 SDGs	25.4%			358
Higher with 4 SDGs	17.3%			244
Do not know	42.5%			598
<b><i>Demographics</i></b>				
Female	60.9%			1,293
Age	46.4	50	14.5	1,295
Highly educated	29.9%			1,229
<b><i>Financial background</i></b>				
Gross individual income (yearly, admin. data)	€ 24,934	€ 20,915	€ 20,189	1,492
Net income (monthly, per household)	€ 2,837	€ 2,750	€ 1,690	969
Low income (€ 0 to € 2,500)	36.6%			473
Middle income (€ 2,500 to € 4,000)	28.3%			366
High income (above € 4,000)	10.6%			130
do not report	25.0%			323
<b><i>Comprehension</i></b>				
Understood treatment (Yes)	86.0%			1,694
More information (Yes)	33.5%			1,542

**TABLE A2.3 – Comparison of the official 2017 Dutch parliamentary election results to the reported voting behavior of respondents in our sample.**

This table presents the distribution of votes in the 2017 Dutch parliamentary elections within our sample and compares it to the official election results. 1,294 respondents answered the question “Did you vote in the last national parliamentary election?”. Respondents were given the answer options ‘Yes’, ‘No’, and ‘I do not want to say’. The last option was chosen by 70 respondents while 1,224 respondents reported to have voted. As a follow-up question, respondents were asked to name the party they voted for. They could also choose to keep this information private (‘I do not want to say’) which 260 (24.9%) respondents did. The voting distribution of the remaining 784 respondents is reported below. The official 2017 parliamentary election results can be found online at <https://www.verkiezingsuitslagen.nl/verkiezingen/detail/TK20170315>.

	Reported voting turnover (n = 1,224)		Official 2017 voter turnover
	absolute	relative	
<b>Voted in 2017</b>	1,034	85.4%	81.6%
	Reported voting behavior (n = 784)		Official 2017 election results
	absolute	relative	
Volkspartij voor Vrijheid en Democratie (VVD)	172	21.9%	21.3%
GroenLinks	110	14.0%	9.1%
Christen-Democratisch Appèl (CDA)	102	13.0%	12.4%
Democraten 66 (D66)	98	12.5%	12.2%
SP (Socialistische Partij)	64	8.2%	9.1%
Partij voor de Dieren	47	6.0%	3.2%
Partij voor de Vrijheid (PVV)	40	5.1%	13.1%
ChristenUnie	40	5.1%	3.4%
Partij van de Arbeid (P.v.d.A.)	39	5.0%	5.7%
50PLUS	21	2.7%	3.1%
Staatkundig Gereformeerde Partij (SGP)	20	2.6%	2.1%
Forum voor Democratie	18	2.3%	1.8%
DENK	1	0.1%	2.1%
Other	12	1.5%	1.5%

TABLE A2.4 – The preference for sustainable policies.

This table presents the preference for sustainable policies which is a score which is constructed by evaluating Dutch parties on their stance on sustainability using the official voting guide for the 2017 Dutch national elections (<https://tweedekamer2017.stemwijzer.nl>). Five of the 30 questions shown deal with sustainable issues. All questions are posed such that they can be answered with 'agree', 'disagree', or 'neither of the two'. A party receives one point for each question for which they choose the more sustainable option. If a party chooses 'neither of the two', it receives half a point. The score is calculated for each of the parties shown in Table A2.3.

	VVD	PVV	CDA	D66	GroenLinks	SP	PvdA
<b>Mileage tax</b>	0	0	0	1	1	1	1
The government must not tax the possession of the car, but the number of kilometers driven. (agree = 1 pt)							
<b>Infrastructure</b>	0	0	0	1	1	1	1
More money needs to be spent on the construction of new roads. (agree = 0 pts)							
<b>Coal-fired power stations</b>	0	0	0.5	1	1	1	1
All coal-fired power stations may remain open for the time being. (agree = 0 pts)							
<b>VAT rate for meat</b>	0	0	0	0	1	0	0
The high VAT rate of 21% must be applied to meat. (agree = 1 pt)							
<b>Development aid</b>	0	0	0	1	1	1	1
The Netherlands should spend more money for the development of poor countries. (agree = 1 pt)							
<b>Total</b>	<b>0</b>	<b>0</b>	<b>0.5</b>	<b>4</b>	<b>5</b>	<b>4</b>	<b>4</b>

	Christen-Unie	Partij voor de Dieren	50PLUS	SGP	DENK	Forum voor Democratie
<b>Mileage tax</b>	1	1	0	1	1	1
The government must not tax the possession of the car, but the number of kilometers driven. (agree = 1 pt)						
<b>Infrastructure</b>	1	1	1	0	1	0
More money needs to be spent on the construction of new roads. (agree = 0 pts)						
<b>Coal-fired power stations</b>	1	1	1	0	1	0
All coal-fired power stations may remain open for the time being. (agree = 0 pts)						
<b>VAT rate for meat</b>	0	1	0	0	0	0
The high VAT rate of 21% must be applied to meat. (agree = 1 pt)						
<b>Development aid</b>	1	1	1	1	1	0
The Netherlands should spend more money for the development of poor countries. (agree = 1 pt)						
<b>Total</b>	<b>4</b>	<b>5</b>	<b>3</b>	<b>2</b>	<b>4</b>	<b>1</b>

TABLE A2.5 – Voting behavior, preferences, and sustainable investing.

This table presents the distribution of preferences and the choice on sustainable investing of respondents who are grouped according to their reported voting behavior in the Dutch national elections 2017. The reported voting behavior refers to the question “Did you vote in the last national parliamentary election?”. Respondents were given the answer options ‘Yes’, ‘No’, and ‘I do not want to say’. As a follow-up question, respondents were asked to name the party they voted for. They could also choose to keep this information private. Preferences for sustainable policies is a score which is constructed by evaluating Dutch parties on their stance on sustainability using the official voting guide for the 2017 Dutch national elections (<https://tweedekamer2017.stemwijzer.nl>). Five of the 30 questions shown deal with sustainable issues. All questions are posed such that they can be answered with ‘agree’, ‘disagree’, or ‘neither of the two’. A party receives one point for each question for which they choose the more sustainable option. If a party chooses ‘neither of the two’, it receives half a point (see Table A2.4). The distribution of choices ‘4 SDG’ and ‘3 SDG’ refers to the following question: “Do you want Pensioenfonds Detailhandel to add (leave out) the fourth sustainable development goal ‘Responsible consumption and production’?” Answer options were ‘Yes, add (leave it out)’, ‘No, do not add (leave it out)’, and ‘I have no opinion regarding this matter’. Words in parentheses refer to the different status quo treatments which are looked at jointly here.

	Preferences for sustainable policies	Percentage of respondents who chose	
		4 SDGs	3 SDGs
Partij voor de Dieren	5	91.5%	4.2%
GroenLinks	5	89.1%	5.5%
Democraten 66 (D66)	4	83.7%	11.2%
DENK	4	<i>too few obs.</i>	
SP (Socialistische Partij)	4	75.0%	4.7%
Partij van de Arbeid (P.v.d.A.)	4	76.9%	10.3%
ChristenUnie	4	82.5%	5.0%
50PLUS	3	76.2%	4.8%
Staatkundig Gereformeerde Partij (SGP)	2	70.0%	5.0%
Forum voor Democratie	1	44.4%	38.9%
Christen-Democratisch Appèl (CDA)	0.5	81.4%	7.8%
Volkspartij voor Vrijheid en Democratie (VVD)	0	70.3%	14.0%
Partij voor de Vrijheid (PVV)	0	55.0%	20.0%
Other	-	75.0%	16.7%
Do not report	-	68.1%	12.3%
Did not vote	-	67.9%	8.4%

TABLE A2.6 – Approval ratings for various exclusion policies.

This table compares the approval ratings of respondents regarding exclusion policies. Exclusion refers to the following question: “Pensioenfondsen Detailhandel wants to know which companies it should exclude from its investments (multiple answers possible). Companies... a. producing tobacco; b. producing controversial weapons (e.g. cluster bombs and landmines); c. producing alcohol; d. producing weapons (weapons other than controversial weapons); e. violating human rights; f. using forced labor; g. having a bad influence on the environment; h. involving in corruption, extortion or bribery; i. allowing the use of child labor; j. I do not want to exclude companies.” Respondents who choose to not exclude any company could not select any of the answer options ‘a’ through ‘j’. Column (1) shows the average of all respondents who answered the exclusion question. The categorization of respondents in columns (2) to (4) is done according to how they answered the question: “Do you want Pensioenfondsen Detailhandel to add (leave out) the fourth sustainable development goal ‘Responsible consumption and production’?”. Answer options were ‘Yes, add (leave it out)’, ‘No, do not add (leave it out)’, and ‘I have no opinion regarding this matter’. Words in parentheses refer to the different status quo treatments which are looked at jointly here. Respondents are categorized into those who chose four SDGs, those who chose three SDGs, and those who reported to have no opinion.

	All	4 SDGs	3 SDGs	No opinion
Corruption	86.9%	90.3%	80.0%	76.1%
Child labor	86.2%	91.1%	73.9%	71.4%
Human rights violation	85.3%	89.9%	69.6%	74.2%
Forced labor	83.2%	86.7%	73.2%	73.7%
Controversial weapons	79.8%	84.0%	71.7%	65.3%
Bad influence on environment	72.3%	78.1%	52.9%	58.2%
Non-controversial weapons	70.7%	74.6%	58.0%	61.0%
Tobacco	44.0%	47.4%	38.4%	31.9%
Alcohol	17.8%	19.4%	14.5%	12.7%
No exclusion	5.7%	3.8%	7.9%	13.1%

## A.2 Figures

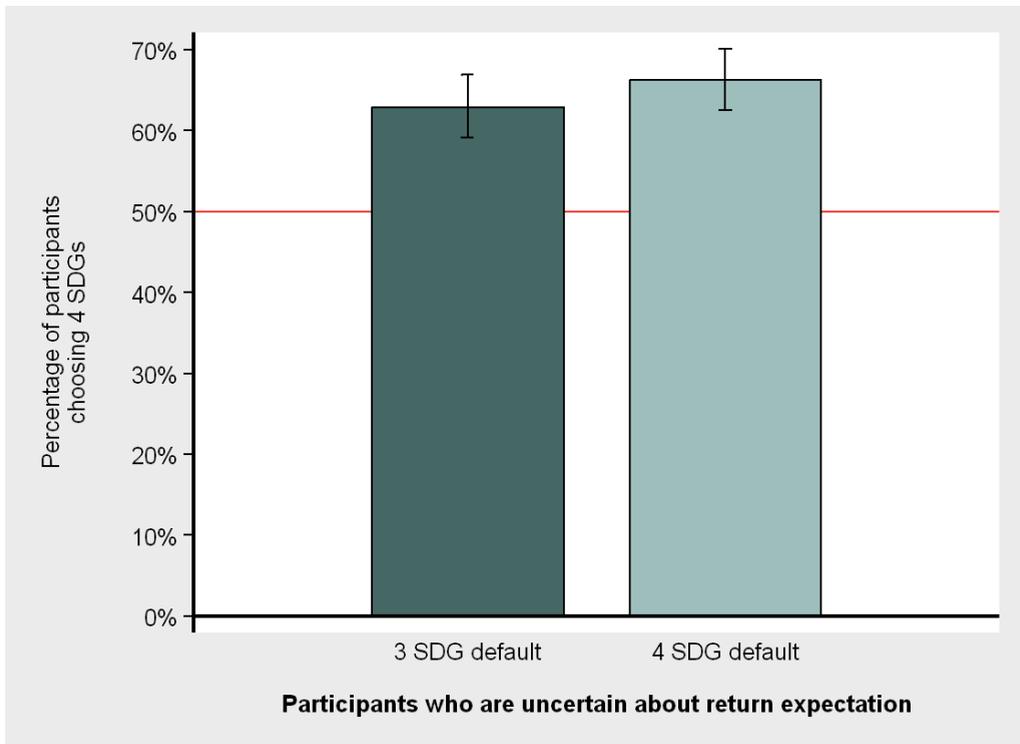


FIGURE A.1 – Percentage choosing 4 SDGs if uncertain about return.

The graph presents the percentage of participants choosing four SDGs if they answered 'I do not know' to the following question: "When do you think the investment return is highest?" Answer options were 'It is highest with 3 SDGs', 'The return is equally high with 3 or 4 SDGs', 'It is highest with 4 SDGs', and 'I do not know'. 3 SDG default and 4 SDG default refer to two different status quo treatments to which participants were randomly assigned. In the *3 SDG default*, participants were told that the pension fund currently focuses on three SDGs. Participants were introduced to the fourth SDG 'Responsible consumption and production' and could choose: "Do you want Pensioenfonds Detailhandel to add the fourth sustainable development goal 'Responsible consumption and production'?". Answer options were 'Yes, add', 'No, do not add', and 'I have no opinion regarding this matter'. In the *4 SDG default*, it was communicated that the pension fund's future policy will include a fourth SDG 'Responsible consumption and production'. Participants could then choose: "Do you want Pensioenfonds Detailhandel to leave out the fourth sustainable development goal 'Responsible consumption and production'?". Answer options were 'Yes, leave it out', 'No, do not leave it out', and 'I have no opinion regarding this matter'. Choices were guaranteed to be implemented by the pension fund if more than 50% of respondents would choose in favor of three, respectively four SDGs. Error bars present 95% confidence intervals.

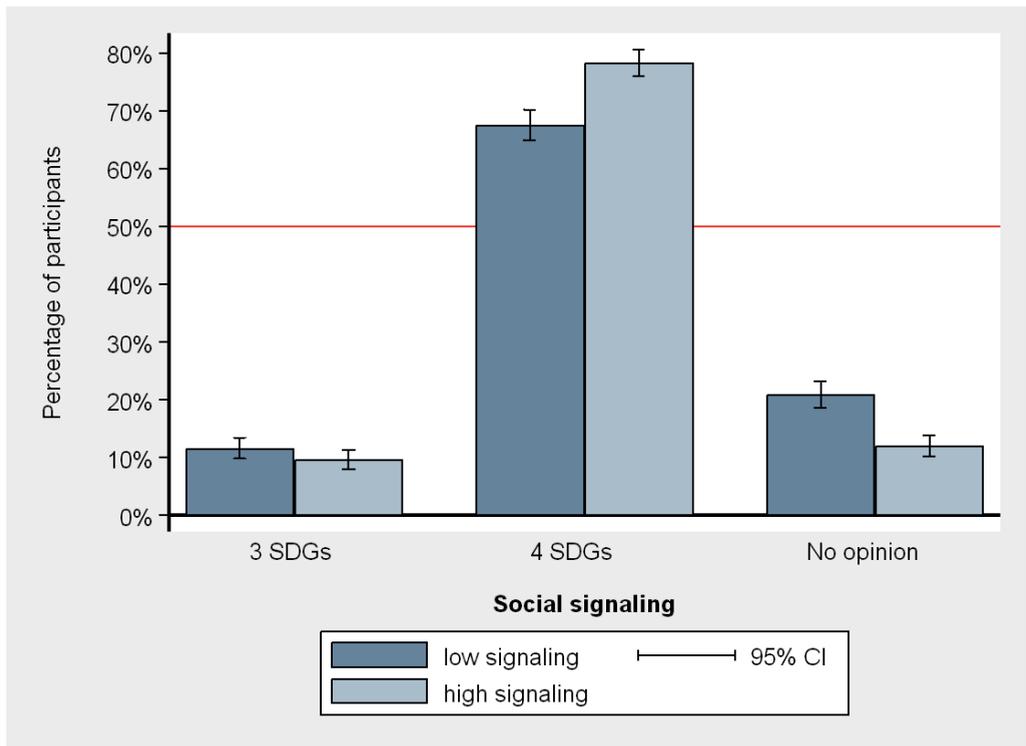


FIGURE A.2 – The effect of social signaling on sustainable investing.

This graph presents the distribution of choices for respondents with a low respectively high desire for social signaling. Social signaling is elicited by asking respondents to rate the question “How likely is it that you will talk to your friends about this survey?” on a 10-point Likert scale from very unlikely to very likely (see Riedl and Smeets, 2017). Respondents are reported to have a low desire for social signaling if they rated this question with 2 or lower (41.6% of the sample) and a high desire for social signaling with a 3 or higher (58.4%). The categories shown in the graph refer to the following question: “Do you want Pensioenfonds Detailhandel to add (leave out) the fourth sustainable development goal ‘Responsible consumption and production’?”. Answer options were ‘Yes, add (leave it out)’, ‘No, do not add (leave it out)’, and ‘I have no opinion regarding this matter’. Choices were guaranteed to be implemented by the pension fund if more than 50% of respondents would choose in favor of three, respectively four SDGs. Words in parentheses refer to the different status quo treatments which are looked at jointly here. Error bars present 95% confidence intervals.

### A.3 Survey

#### Welcome!

Pensioenfonds Detailhandel invests **your pension savings in a sustainable way**. In doing so, Pensioenfonds Detailhandel strives to achieve continuous improvements.

That is why Pensioenfonds Detailhandel has decided to ask you about our investment policy.

Of the people who participate in this survey, 25 randomly selected individuals will receive a **VVV Gift Voucher worth 25 euros**. Additionally, one randomly selected participant will receive a **VVV Gift Voucher worth 250 euros**.

You will be entered in the prize draw if you complete the whole survey. Maastricht University and Pensioenfonds Detailhandel guarantee that all payments will be made.

This survey takes **around fifteen minutes** to complete. Pensioenfonds Detailhandel will make the results available in Autumn.

Thank you in advance for your participation!

Yours sincerely, Maastricht University and Pensioenfonds Detailhandel.

---

#### [Part 1]

#### What is socially responsible investment?

1. Are you familiar with the United Nation's Sustainable Development Goals?
  - (a) Yes
  - (b) No

*[Answer if yes or no:]* The United Nation's Sustainable Development Goals focus on seventeen goals relating to the environment, the climate, poverty, and other themes. The next page explains how Pensioenfonds Detailhandel uses these Sustainable Development Goals.

---

[Insert]

— 3 SDG default —

— 4 SDG default —

---

**[For both treatments]**

6. When do you think that the investment return is the highest?

*[Info button: Investment return is the degree to which an investment earns money when taking the level of risk into account. In other words, more money results in a greater return. However, an investment that takes a greater risk to earn a certain amount results in a lower risk-adjusted return.]*

- (a) The investment return is highest for three sustainable development goals.
  - (b) The investment return is highest for four sustainable development goals.
  - (c) The investment return remains the same for three or four sustainable development goals.
  - (d) I don't know
7. Would you want Pensioenfonds Detailhandel to opt for four Sustainable Development Goals even if this decreased the risk-adjusted return?
- (a) Yes
  - (b) No
  - (c) I don't know
8. Predict which percentage of Pensioenfonds Detailhandel's participants would also opt for four Sustainable Development Goals even if this decreased the risk-adjusted return.
- (a) --- percent

[Part 2]

The following questions aim to clarify your opinion of investing in a socially responsible way.

As opposed to previous questions, your answers in this part may not lead to Pensioenfonds Detailhandel taking direct action but can help us make decisions in the future.

**How does Pensioenfonds Detailhandel currently make decisions regarding socially responsible investments?**

Pensioenfonds Detailhandel now implements its socially responsible investment policy in two ways:

- i. by excluding companies
- ii. by entering into discussions with companies.

The next question concerns excluding companies as a method of investing in a socially responsible way.

9. Pensioenfonds Detailhandel would like you to list which types of companies you think it should exclude from investments (several answers possible). Companies that...
  - (a) produce tobacco
  - (b) produce controversial weapons such as cluster bombs and landmines
  - (c) produce alcohol
  - (d) produce weapons (other than controversial weapons)
  - (e) infringe on human rights
  - (f) use forced labour
  - (g) have a negative impact on the environment
  - (h) are involved in corruption, extortion or bribery
  - (i) allow child labour

**[Part 3]**

We would also like to ask you some background questions.

10. How willing are you to give to good causes without expecting anything in return?  
completely unwilling 1—10 very willing
11. I am...
  - (a) Male
  - (b) Female
  - (c) Other
12. In what year were you born?
  - (a) Year of birth
13. What is the highest level of education you have completed?
  - (a) Preparatory secondary vocational education (VMBO) or lower general secondary education (MAVO or MULO)
  - (b) Higher general secondary education (HAVO)
  - (c) Pre-university education (HBS, HAVO, VWO) or pre-university education with Latin and/or Greek (Gymnasium)
  - (d) Intermediate vocational education level 1 (MBO)
  - (e) Intermediate vocational education level 2 (MBO)
  - (f) Intermediate vocational education level 3 (MBO)
  - (g) Intermediate vocational education level 4 (MBO)
  - (h) Higher professional education (HBO)
  - (i) University (WO)
  - (j) Other: ---
  - (k) I did not follow any of the above types of education
14. Did you vote in the last national parliamentary elections?
  - (a) Yes
  - (b) No
  - (c) I prefer not to answer

15. *If 'Yes' in previous question:* Did you vote in the last national parliamentary elections?
- (a) VVD
  - (b) CDA
  - (c) D66
  - (d) ChristenUnie
  - (e) PVV
  - (f) GroenLinks
  - (g) SP
  - (h) PvdA
  - (i) Partij voor de Dieren
  - (j) 50PLUS
  - (k) SGP
  - (l) DENK
  - (m) Forum voor Democratie
  - (n) Other, specifically: ---
  - (o) I prefer not to answer

We have a few questions regarding your financial situation. We understand that this information is sensitive and you can decide to skip these questions. However, as your answers to these questions are very valuable, we hope you will answer them. We will handle your data with due care.

16. What is your household's net monthly income?

*[Info button: net = your and your partner's income combined (if applicable) minus taxes and contributions.]*

- (a) Less than € 930
- (b) Between € 930 and € 1,500
- (c) Between € 1,500 and € 2,000
- (d) Between € 2,000 and € 2,500
- (e) Between € 2,500 and € 3,000
- (f) Between € 3,000 and € 4,000
- (g) Between € 4,000 and € 7,000
- (h) Between € 7,000 and € 10,000
- (i) Over € 10,000
- (j) I don't know
- (k) I prefer not to answer

17. How likely are you to discuss this survey with friends?  
completely unlikely 1—10 very likely

18. We asked approximately 25,000 participants to complete this survey. How many people do you think will complete the survey? ---

19. If you win a VVV Gift Voucher, we will contact you by e-mail. If you want to win the prize, please tick the box.  
 I want to win the prize.

20. If you would like to receive an e-mail with information about the results of this survey, please tick the box.  
 I would like to receive information.

21. Please fill in your e-mail address below.

\_\_\_\_\_

---

Thank you for participating.

Yours sincerely, Maastricht University and Pensioenfonds Detailhandel.

---

## A.4 Treatment '3 SDG Default'

Pensioenfond's Detailhandel currently focuses on three of the United Nations' 17 Sustainable Development Goals. These three Sustainable Development Goals are:

1. **Climate action:**

e.g. taking urgent action to combat climate change such as by ensuring that businesses emit less CO<sub>2</sub>

2. **Decent work and economic growth:**

e.g. full employment for all women and men

3. **Peace, justice, and strong institutions:**

e.g. eradication of corruption and bribery

For more information on the Sustainable Development Goals, please click here: <http://www.sdgnederland.nl/sdgs/>

### What does it mean to invest according to the United Nations' Sustainable Development Goals?

Pensioenfond's Detailhandel uses its influence in the companies in which it invests. In 2017, the pension fund spoke **with a total of 246 company boards to promote sustainability.**

#### **Royal Dutch Shell case:**

Working in collaboration with other parties, **Pensioenfond's Detailhandel has contacted Shell** on a number of occasions and made it clear that Shell has a lot of work ahead of it to achieve the **objectives of the Paris Climate Agreement** (to become CO<sub>2</sub> neutral by 2015) and that it is urgent to do so.

#### **The example of the diesel scandal in 2015:**

**Pensioenfond's Detailhandel contacted managers at Volkswagen (VW)** in the aftermath of the diesel scandal. **Pensioenfond's Detailhandel's involvement helped VW** focus more on its long-term emissions strategy.

### **You have been selected!**

Pensioenfond's Detailhandel has decided to give you a say in how it determines its socially responsible investment policy.

### **But how does this work?**

1. You can let Pensioenfond's Detailhandel know **whether you agree with the current policy** that focuses on three Sustainable Development Goals, or whether you would like to add a fourth one.
2. **Your choice counts. If the majority** of respondents chooses to add a fourth sustainable development goal, it will happen. The board of Pensioenfond's Detailhandel guarantees its implementation.

Below, you find the fourth Sustainable Development Goal that you can add to the current policy.

**Sustainable consumption and production:** This Sustainable Development Goal ensures that parties, such as Pensioenfond's Detailhandel, are obliged to do the following:

e.g. actively work against child labour, guarantee fair wages.

**Please note:** implementing Sustainable Development Goals means that financial returns are not the only factor that is taken into consideration. Making investments with this in mind means that it is important to take the impact on the environment and wider society into account.

### **The influence of your choice**

If Pensioenfond's Detailhandel focuses on four Sustainable Development Goals, this means that it will contact companies to discuss their sustainable business practices more often. The fund will also enter into discussions with companies about the fourth Sustainable Development Goal, in addition to the discussions it has about the other three Sustainable Development Goals.

As it is important that you understand this explanation, we will ask you a short question to test your understanding on the next page.

On the next page you will be asked the following:

*Do you want Pensioenfonds Detailhandel to add the fourth sustainable development goal 'Responsible consumption and production'?*

- (a) Yes, add*
- (b) No, do not add*
- (c) I have no opinion regarding the matter*

**Which of the following statements is correct?**

- 2. If a majority chooses 'Yes, add', Pensioenfonds Detailhandel
  - (a) **guarantees** to add 'Sustainable consumption and production' to its socially responsible investment policy as the fourth Sustainable Development Goal. The Board of Pensioenfonds Detailhandel has decided to implement the outcome of this vote.
  - (b) **cannot guarantee** that it will add 'Sustainable consumption and production' to its socially responsible investment policy as the fourth Sustainable Development Goal but may include the results of the survey in its choice.

*If a:* That is correct. Pensioenfonds Detailhandel **guarantees** that it will add 'Sustainable consumption and production' to its socially responsible investment policy as the fourth Sustainable Development Goal if the majority of participants chooses 'Yes, add'. The Board of Pensioenfonds Detailhandel has decided to implement the outcome of this vote.

*If b:* Your answer is incorrect. Pensioenfonds Detailhandel **guarantees** that it will add 'Sustainable consumption and production' to its socially responsible investment policy as the fourth Sustainable Development Goal if the majority of participants chooses 'Yes, add'. The Board of Pensioenfonds Detailhandel has decided to implement the outcome of this vote.

Please note that this is your choice. When considering your decision, please do so carefully.

- 3. Do you want Pensioenfonds Detailhandel to add the fourth sustainable development goal 'Responsible consumption and production'?
  - (a) Yes, add
  - (b) No, do not add
  - (c) I have no opinion regarding the matter

4. Would you have preferred receiving more information to make a decision regarding the development goals?
  - (a) Yes
  - (b) No
  
5. Please indicate below which percentage of participants of this survey will choose which answer regarding the question of adding the fourth Sustainable Development Goal.
  - (a) Yes, add: \_\_\_ percent
  - (b) No, do not add: \_\_\_ percent
  - (c) I have no opinion regarding the matter: \_\_\_ percent

## A.5 Treatment '4 SDG Default'

Pensioenfonds Detailhandel currently focuses on three of the United Nations' 17 Sustainable Development Goals. These three Sustainable Development Goals are:

1. **Climate action:**

e.g. taking urgent action to combat climate change such as by ensuring that businesses emit less CO<sub>2</sub>

2. **Decent work and economic growth:**

e.g. full employment for all women and men

3. **Peace, justice, and strong institutions:**

e.g. eradication of corruption and bribery

The future policy also includes this fourth goal

4. **Sustainable consumption and production**

For more information on the Sustainable Development Goals, please click here: <http://www.sdgnederland.nl/sdgs/>

### **What does it mean to invest according to the United Nations' Sustainable Development Goals?**

Pensioenfonds Detailhandel uses its influence in the companies in which it invests. In 2017, the pension fund spoke **with a total of 246 company boards to promote sustainability.**

#### **Royal Dutch Shell case:**

Working in collaboration with other parties, **Pensioenfonds Detailhandel has contacted Shell** on a number of occasions and made it clear that Shell has a lot of work ahead of it to achieve the **objectives of the Paris Climate Agreement** (to become CO<sub>2</sub> neutral by 2015) and that it is urgent to do so.

#### **The example of the diesel scandal in 2015:**

**Pensioenfonds Detailhandel contacted managers at Volkswagen (VW)** in the aftermath of the diesel scandal. **Pensioenfonds Detailhandel's involvement helped VW** focus more on its long-term emissions strategy.

### **You have been selected!**

Pensioenfonds Detailhandel has decided to give you a say in how it determines its socially responsible investment policy.

### **But how does this work?**

1. You can let Pensioenfonds Detailhandel know **whether you agree with the future policy** that focuses on four Sustainable Development Goals, or whether you would like to leave out the fourth one.
2. **Your choice counts. If the majority** of respondents chooses that the fourth sustainable development goal must be left out, it will happen. The board of Pensioenfonds Detailhandel guarantees its implementation.

Below, you find the fourth Sustainable Development Goal that you can leave out from the future policy.

**Sustainable consumption and production:** This Sustainable Development Goal ensures that parties, such as Pensioenfonds Detailhandel, are obliged to do the following:  
e.g. actively work against child labour, guarantee fair wages.

**Please note:** implementing Sustainable Development Goals means that financial returns are not the only factor that is taken into consideration. Making investments with this in mind means that it is important to take the impact on the environment and wider society into account.

### **The influence of your choice**

If Pensioenfonds Detailhandel focuses on four Sustainable Development Goals, this means that it will contact companies to discuss their sustainable business practices more often. The fund will also enter into discussions with companies about the fourth Sustainable Development Goal, in addition to the discussions it has about the other three Sustainable Development Goals.

As it is important that you understand this explanation, we will ask you a short question to test your understanding on the next page.

On the next page you will be asked the following:

*Do you want Pensioenfond's Detailhandel to leave out the fourth sustainable development goal 'Responsible consumption and production'?*

- (a) Yes, leave it out*
- (b) No, do not leave it out*
- (c) I have no opinion regarding the matter*

**Which of the following statements is correct?**

- 2. If a majority chooses 'Yes, leave it out', Pensioenfond's Detailhandel
  - (a) **guarantees** to leave out 'Sustainable consumption and production' of its socially responsible investment policy as the fourth Sustainable Development Goal. The Board of Pensioenfond's Detailhandel has decided to implement the outcome of this vote.
  - (b) **cannot guarantee** that it will leave out 'Sustainable consumption and production' of its socially responsible investment policy as the fourth Sustainable Development Goal but may include the results of the survey in its choice.

*If a:* That is correct. Pensioenfond's Detailhandel **guarantees** that it will leave out 'Sustainable consumption and production' of its socially responsible investment policy as the fourth Sustainable Development Goal if the majority of participants chooses 'Yes, leave it out'. The Board of Pensioenfond's Detailhandel has decided to implement the outcome of this vote.

*If b:* Your answer is incorrect. Pensioenfond's Detailhandel **guarantees** that it will leave out 'Sustainable consumption and production' of its socially responsible investment policy as the fourth Sustainable Development Goal if the majority of participants chooses 'Yes, leave it out'. The Board of Pensioenfond's Detailhandel has decided to implement the outcome of this vote.

Please note that this is your choice. When considering your decision, please do so carefully.

- 3. Do you want Pensioenfond's Detailhandel to leave out the fourth sustainable development goal 'Responsible consumption and production'?
  - (a) Yes, leave it out
  - (b) No, do not leave it out
  - (c) I have no opinion regarding the matter

4. Would you have preferred receiving more information to make a decision regarding the development goals?
  - (a) Yes
  - (b) No
  
5. Please indicate below which percentage of participants of this survey will choose which answer regarding the question of adding the fourth Sustainable Development Goal.
  - (a) Yes, leave it out: \_\_\_ percent
  - (b) No, do not leave it out: \_\_\_ percent
  - (c) I have no opinion regarding the matter: \_\_\_ percent



# Chapter 3

## Dare to Pair!

Random pairs update beliefs symmetrically with and without communication\*

### 3.1 Introduction

When Muhammad Ali stepped into the ring to fight Joe Frazier in 1971, it had already been billed “The Fight of the Century”. Either of them had been undefeated, Ali had won 25 fights by knockout, Frazier 23. After an intense battle over 15 rounds, the jury proclaimed Frazier to be the winner. 24 years later, a jury faced what came to be known as “The Trial of the Century”. In 1995, then former National Football League player O. J. Simpson was tried on two counts of murder for the deaths of his ex-wife and a friend of hers. After an eleven-month trial, the grand jury acquitted O. J. Simpson. Both times, the jury came to a unanimous verdict. Yet, there was a crucial difference. Boxing jurors make their verdict independent of each other, no communication is allowed. The grand jury, on the other hand, is forced to communicate. Two century-shaping events, two different ways to process information and find a verdict. Is one objectively superior? Does a group decision profit or suffer from communication between its members?

In this paper, we study the causal effect of communication on the updating of beliefs. Not only jurors have to constantly update their beliefs. Consider a company on the brink to expand its business. Should the company rely on experts who jointly discuss new information on the project, or should these experts separately assess the situation for the company to then decide on the basis of several individual beliefs? Or consider the hiring process of a new assistant professor. Should professors of the hiring department assess the candidate jointly or separately? While

---

\* This chapter is co-authored by Paul Smeets (Maastricht University) and Pim van Vliet (Robeco). We are especially grateful to the valuable feedback of Rob Bauer, Jaap Bos, Paulo Rodrigues, Martin Strobel, and Antoon Pelsser. We thank seminar participants at Maastricht University.

we work in groups in many areas of our daily life, we know surprisingly little about how group beliefs are formed.

Several papers have looked at how well groups process information. Many find groups to perform better than individuals (Sniezek and Henry, 1989, 1990; Hinsz, 1990; Sniezek, 1992; Bonner *et al.*, 2007; Adams and Ferreira, 2009), while others show that groups do worse (Hill, 1982; Puncochar and Fox, 2004). One of the most prominent examples for poor group judgment is provided by Asch (1956) who shows that conformity can prompt people to subordinate to group decisions even if they are clearly wrong. In a task where a group has to pick the longest of three different lines, Asch finds individuals to go with the group decision even if the group picks an obviously shorter line as the longest just to conform with the group—a behavior that has been found manifold since then (Bond and Smith, 1996). Group conform behavior is not possible if groups are not able to communicate. Another example for bad group decisions is portrayed by a situation where a group of friends is looking for a place to have dinner. Most of the time, the group ends up discussing for the better part of the evening to eventually end up in a restaurant nobody actually likes. Would it be better in such a situation if everybody wrote down his favorite restaurant on a piece of paper and the one that gets picked by the majority is chosen? To answer these questions, we study how groups should make a decision and whether communication should be allowed.

We designed an experiment where pairs form beliefs about their team performance over several rounds. Thereby, one half is allowed to verbally communicate (communication treatment) while the other half is not (no communication control). Assignment is random. Subjects are informed that they have been placed in a pair before the calculation task. Yet, they do not learn about the actual treatment until after the task. Subjects individually complete a timed, ego-relevant calculation task. After completion, pairs receive their team score which is the sum of all correctly solved calculations between the two. All subjects then are asked to indicate how likely it is that their team score is in the top half of all team scores. Then, those in the communication treatment are told that after indicating their individual belief, they will have to chat to their teammate to find a unanimous belief. Subjects in the no communication control are told that after indicating their individual belief, they will be asked to write down the reasoning for their belief in a protocol. At the same time, they will see the belief of their teammate to which they can react after the protocol. They are not required to coordinate their beliefs. Successively, noisy feedback on the team performance is given in the form of a binary signal which is either positive (“Top half”) or negative (“Bottom half”). After each of the four signals, pairs can update their beliefs in the same way as just described. Bayesian updating requires pairs to respond equally strong to positive and negative signals. If pairs show a stronger response

to positive than to negative signals, they update asymmetrically<sup>1</sup>. In the long-run, asymmetric updating leads to unjustifiably optimistic beliefs.

In our experiment, we focus on asymmetric belief updating for two reasons. First, the formation of beliefs resembles the very core of a decision-making process. Take the example of expanding a business. Incoming information will either support an expansion or suggest to not do it. With each piece of information, the decision-makers within the company have to form a belief about how likely a successful expansion is. On the basis of this belief updating process, the decision will eventually be taken. Second, while some papers find overconfidence to have positive consequences (e.g. Bernardo and Welch 2001; Hirshleifer and Luo 2001; Johnson and Fowler 2011; Gervais *et al.* 2011), the majority of papers find costly effects (e.g. Camerer and Lovallo 1999; Barber and Odean 2001; Biais *et al.* 2005; Malmendier and Tate 2005, 2008; Möbius *et al.* 2007, 2014; Kim *et al.* 2016; Bordalo *et al.* forthcoming). These papers underline the necessity to study how to reduce overconfidence when it is harmful. We deliberately chose an experimental design where overconfidence is financially costly and where subjects are aware of this. Further, we only look at pairs as there are multiple examples where decisions are made by two people. In the academic world, many papers are written by two researchers; in the industry, many company decisions are taken by CEO and CFO. And even outside the professional world, couples take daily household decisions. On top, decision making in pairs allows us to better control for group dynamics which get significantly more complex as soon as we introduce more people to the group. While research has only focused on overconfident decision-making among individuals, we take the research on this topic to the next level by looking at randomly matched pairs.

To understand how pairs update beliefs, we have to first look at individuals. Individuals update their beliefs highly asymmetrically in ego-relevant tasks (see e.g. Möbius *et al.* 2014; Eil and Rao 2011; Cacault and Grieder 2018) where they value positive news more than negative. This individual behavior leads to two forces which can potentially shape the belief formation process in pairs where communication is allowed.

(1) The force of disagreement: Members of a group have different individual beliefs which serve as anchor points in the process of forming a team belief (Eriksen and Hake, 1957; Tversky and Kahneman, 1974; Sniezek and Henry, 1989; Kahneman, 1992; Sniezek, 1992; Whyte *et al.*, 1997). These individual beliefs compete with each other and lead to disagreement among teammates which in turn leads to a thorough discussion of available information. Sniezek and Henry (1989) hypothesize that competing individual beliefs thus increase the likelihood to detect asymmetric individual belief updates (the ‘competing anchors hypothesis’). This process, however, can be impeded if individual beliefs lie closely together or suffer from the same bias (Sniezek and Henry, 1989). In this case, there is less disagreement between the individual

---

<sup>1</sup> Unless stated otherwise, when we write asymmetric updating, we refer to pairs responding stronger to positive than to negative feedback.

anchor beliefs and consequently a less thorough discussion. Hence, disagreement should lead to more accurate (less asymmetric) group beliefs and thereby depends on individual beliefs being sufficiently diverse.

(2) The force of dominance: Next to disagreement, interaction in groups is affected by dominant behavior which is defined as the power to impose one's will upon others (see e.g. Halebian and Finkelstein 1993; Brown and Sarma 2007). Dominant individuals are wrongly believed to be more competent (Kenny and LaVoie, 1985) and consistently gain a high level of influence in groups (Anderson and Kilduff, 2009). Thereby, dominance has been found to be positively related to overconfidence. Brown and Sarma (2007) study the impact of overconfidence and dominance on the acquisitiveness of CEOs and find a positive correlation between dominance and overconfidence. Burks *et al.* (2013) study the behavior of 1,000 truck drivers. They find dominant truck drivers to make more confident judgments about their own relative IQ score even when controlling for their actual ability. Hence, the force of dominance suggests that the presence of a dominant person in pairs leads to asymmetric belief updating.

For non-communicating pairs, neither of these two forces apply as verbal interaction is not possible. Given the finding that individual belief updates are asymmetric, belief updates in non-communicating pairs should also be. This hypothesis rests on the argument that taking the mean of two individual biased beliefs, which is the weighting we use for non-communicating pairs in this analysis, cannot lead to a debiased (less asymmetric) team belief (Sniezek and Henry, 1989). Hence, it has to be analyzed empirically whether the positive force of disagreement or the negative force of dominance overweighs in communicating pairs, and consequentially, whether communication leads to better or worse belief updates.

In our sample containing 121 students of which half are in each treatment, we find that non-communicating and communicating pairs update equally well with symmetric reactions to positive and negative feedback. In line with previous findings among individuals, our results show that the presence of dominant members in communicating pairs causes slight asymmetric updating, an effect which is less pronounced in non-communicating pairs. Yet, dominant behavior is not strong enough to affect the aggregate level. Thus, the force of disagreement outweighs the force of dominance in communicating pairs. Last, we find pairs to update conservatively suggesting that they are too slow in adapting to new information—a bias which is independent of communication. In summary, pairs show highly accurate belief updates regardless of whether communication is allowed.

With our work, we contribute to the literature in two important ways. First, we contribute to the literature on overconfidence. Our experiment is a modification of the design of Möbius *et al.* (2014) which we adapted to pairs. The advantage of this method is that the measurement for overconfidence cannot be rationalized (e.g. Möbius *et al.* 2007, 2014; Eil and Rao 2011; Ertac 2011; Buser *et al.* 2016; Cacault and Grieder 2018; Coutts 2018). Beliefs which can be rationalized can be the result of optimal Bayesian updating and thus can be objectively

correct while they seem to be overconfident. Benoît and Dubra (2011) show that “a rational population can appear to be twice as confident as reality would suggest” without necessarily being overconfident. In other words, the mere observation that the entire population believes to be better than the median does itself not yield any information on whether the population is overconfident. Asymmetric updating behavior, however, cannot be rationalized and therefore is a valid indicator for overconfidence under Bayes’ Law. Findings of overconfidence with this new measurement have thereby supported previous work on overconfidence (e.g. Barber and Odean 2001; Hoelzl and Rustichini 2005; Moore and Healy 2008). More importantly, our finding shows that with pairs we do not observe the pattern of asymmetric updating.

Second, we add to the literature on decision-making in groups. Some papers have studied the influence of groups on beliefs through priming individuals with a group identity and have found that this causes individuals to become overconfident (Healy and Pate, 2007; Cacault and Grieder, 2018). Yet, these papers look at individual, not at group decisions. Others have actually studied group beliefs where communication is allowed (Sniezek and Henry, 1989; Hinsz, 1990; Adams and Ferreira, 2009). Our advantage over these papers is, however, that we do not allow for face-to-face interaction which makes a clean identification of the underlying mechanisms easier. Being able to control for the degree of interaction allows us to avoid confounding factors like sympathy, visual stimuli, or potential inactivity of some members. We can further analyze to which degree each individual was involved in forming the group belief as we elicit individual beliefs before we elicit team beliefs. This design allows us to cleanly address the question of how pairs form beliefs. The fact that we do not find any sign of asymmetric belief updating among pairs – especially among communicating pairs – is strong evidence that two people form accurate joint beliefs.

Overall, our study provides new and important insights into the formation of beliefs in pairs. Take the mutual fund industry which has seen significant growth of team-managed funds of which most are led by two investment managers (Bliss *et al.*, 2008). Our findings show that this development could help to decrease the tendency of overconfident investment decisions. Yet, fund managers still have to be aware that they likely under-appreciate new information causing them to react too slowly to new trends.

## 3.2 The experimental design

We randomly<sup>2</sup> assigned subjects to two distinct types of pairs: one where both teammates are allowed to interact through chatting and one, where no verbal interaction is allowed. A graphical

<sup>2</sup> Given the findings of Ruof and Smeets (2018) that the salience of gender has a causal impact on the updating behavior of pairs, we paired respondents conditional on gender. This assured an even amount of male, female, and gender-diverse pairs even though gender was not salient to subjects. Subjects were able to observe that there are male and female students participating in the experiment. However, we deliberately did not mention gender in any form throughout the experiment to avoid gender becoming salient.

illustration of the experiment is provided in Figure 3.1. In the beginning of the experiment, all participants were told that they are about to do a timed calculation task. Subjects then should estimate how likely they are to end up in the top 50 percent of all individual scores in their session (see belief A, Figure 3.1). Only then, we told subjects that they have been randomly paired with another subject of their session. The sum of the two individual scores would form their team score, individual scores would not be reported. Importantly, we stressed that only their team score and not their individual score would be considered for payoff to strengthen their identification with the team. No information was given on the teammate. We then asked subjects to estimate how likely their team score is to end up in the top half of all teams (belief B). The two beliefs were elicited before the calculation task and for each subject individually. Teammates did not get to see the belief of each other. Those in the communication treatment did not know that they would have to interact with their teammate until after the calculation task. Thus, beliefs A and B are not influenced by the teammate's belief. The calculation task itself will be explained in section 3.2.4.

### 3.2.1 The elicitation of beliefs

Participants were asked to state their true beliefs. In the beginning of the experiment, the experimenter instructed all participants that they would have to state beliefs throughout the experiment and that stating true beliefs would maximize their expected payoffs. The experimenter gave participants the paper by Karni (2009) whose payoff mechanism we use and who shows that stating true beliefs is incentive-compatible.<sup>3</sup> As participants were unlikely to be able to read and understand the paper during the experiment, we stressed that the abstract of the paper already stressed the main finding within two sentences, where especially one sentence is of importance: "The game induced by the mechanism has a dominant strategy equilibrium in which the players reveal their subjective probabilities." This procedure allowed us to make it as explicit as possible to participants that stating their true beliefs would yield them the highest expected payoffs. Of all participants, four participants requested to see the paper. The experimenter also gave a quick example which served to stress the fact that subjects should only state their true belief and that anything but their true belief would diminish their expected payoff.

In addition, we made sure that everybody understood the procedure of the experiment by thoroughly explaining the procedure before as well as during the experiment including several example questions. Subjects had to complete a similar non-incentivized updating game before the calculation task to get familiar with the theme.

---

<sup>3</sup>The stated belief is assumed to represent the mean of someone's belief distribution. The elicitation method we employed follows the method developed by Karni (2009) and is strictly incentive-compatible. This method does not allow for hedging behavior, which can happen with the quadratic scoring rule where under certain circumstances expected payoffs for subjects are marginally decreasing in solving more calculations leading to perverse incentives (for proof, see Möbius *et al.* 2014).

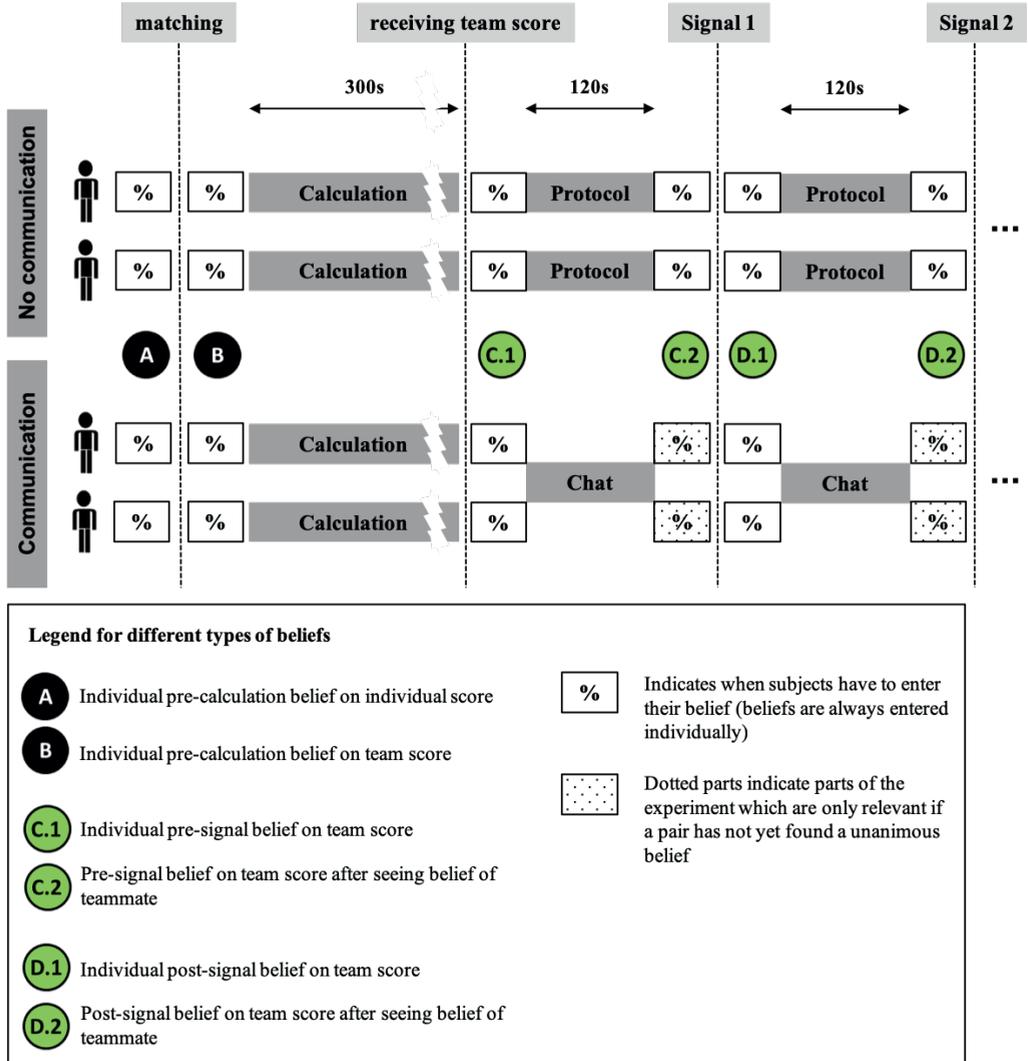


FIGURE 3.1 – Order of events in the experiment (Continued on next page)

FIGURE 3.1—*Continued* —

This figure presents the experimental procedure. The upper part of the figure presents the procedure for pairs with no communication, the lower part for pairs with communication. All subjects were told that they are about to do a timed calculation task. Subjects would receive a score which equals the number of correctly solved calculations. The experiment introduced the calculation task as an individual task and asked subjects to estimate how likely they are to end up in the top 50 percent of all individual scores in their session (belief A). Then, subjects were told that they had been randomly matched to another subject of their session and that the sum of the two individual scores will form their team score. No information was given on the teammate. Subjects were then asked to estimate how likely their team score is to end up in the top half of all teams (belief B). Then subjects individually completed the calculation task (300 seconds) and received their team score. Subjects first had to state their individual belief (belief C.1) that the team score would end up in the top half of all team scores in the session. Subjects in the no communication treatment knew that after entering their individual belief, they would get to see the belief of their teammate, but they would not get to talk to each other. Once each teammate had submitted the individual prior, a protocol window opened and each of the two could see the prior of their teammate (see Figure A3.1). They were told to write down their reasoning for choosing this particular belief and that they could revise their belief after the time for the protocol ended (belief C.2). Yet, they were not forced to adapt as they could just re-enter their previous belief. No unanimous belief was required. Subjects in the communication treatment knew that after entering their individual belief, they would get to chat to their teammate to find a unanimous belief. Once each teammate had submitted the individual prior, a chat window opened and the two could see the prior of their teammate. If the priors did not match, subjects had 120 seconds to find a unanimous belief (see Figure A3.2). After the chat closed, subjects individually entered their revised belief (belief C.2). This belief had to be the same between both teammates otherwise they would not be paid in case this belief was chosen for payment. Then, subjects of either treatment group received a sequence of four informative, but noisy signals on their team performance. After each signal, subjects had to enter their updated beliefs (beliefs D.1 and D.2, etc.) by following the same procedure of their respective treatment group as just described. Updating in this case could also mean to stay with the same belief even after having received a signal. Pairs were not forced to update.

### 3.2.2 The formation of joint beliefs

Consider the following example where we for simplicity refer to all players as being male. Player 1 (P1) and Player 2 (P2) have been matched, as have Player 3 (P3) and Player 4 (P4). Assume, they individually complete the calculation task (300 seconds) and receive the information that they correctly solved a total of 20 calculations within each team. Individual scores are not provided.

In the *no communication treatment*, P1 is asked to state his belief (see Figure 3.1, belief C.1) that their team score is in the top half of all scores in her session. He is also informed that after he has entered his belief, a protocol window opens where he has to state the reasons for choosing his belief (see Figure A3.1). At the same time, he will get to see the belief of his teammate. He is not forced to adapt to his teammate's belief but will have the chance to do so after the time for writing a protocol has ended. If he wants to stay with his belief, he can just re-enter his current belief. He is further told that only his own belief is payoff-relevant for him. P2 follows the same procedure. After receiving all this information, suppose P1 estimates the likelihood to end up in the top half given their team score to be 60%. P2 on the other side believes that their score has a mere 40% chance to make it to the top half. After independently submitting these beliefs (C.1), the protocol window opens. P1 decides to stay with his belief

of 60% and re-enters it on the next screen (C.2). P2 takes the prior of P1 into account and updates his belief to 50% which he re-enters on the next screen. Both beliefs are entitled for payment as no unanimous decision is required. The average of 55% is considered for analysis.

In the *communication treatment*, P3 is asked to state his belief (belief C.1) that their team score is in the top half of all scores in his session. He is also informed that after he has entered his belief, he will get the chance to chat with his teammate to find a unanimous belief. They are informed that only their team belief is payoff-relevant for each teammate. After receiving all this information, he enters his individual belief. P4 follows the same procedure. Once each of them has submitted an individual prior (belief C.1), a chat window opens for 120 seconds. The two can see the prior of their teammate on the screen (see Figure A3.2). In case their beliefs do not match, P3 and P4 only have one more chance after the chatting round to enter a unanimous belief (belief C.2). If the revised priors are not the same, there is no possibility for revision. Both are told in advance that disagreement after the chat results in losing the entitlement to payment in case this round is chosen for payment. Suppose P3 estimates the likelihood to end up in the top half given their team score to be 60%. P4 on the other side believes that their score has a mere 40% chance to make it to the top half. After independently submitting these beliefs (C.1), they get to chat with each other to find a unanimous decision. Both agree on 50% and individually enter the belief on the next screen (C.2) enabling them to be eligible for payment in case this belief is chosen.

Note that subjects in the communication treatment and in the no communication control receive equal information. Specifically, all subjects learn about the team score and about the teammate's belief. The only difference between the treatment groups is whether they can communicate or not.

### 3.2.3 The updating of beliefs

After pairs have entered their C.2 beliefs, they are informed that they will receive a sequence of four conditionally independent, noisy, but informative signals. Teammates get to see the same signal. The signal provides an indication whether the team score is in the top or bottom half and has an accuracy of 75 percent. In other words, if the team score is in the top half, the signal says "Top" with a probability of 75 percent and "Bottom" with a probability of 25 percent. Equivalently, if the score is in the bottom half, the signal says "Bottom" with a probability of 75 percent and "Top" with a probability of 25 percent (symmetric signals). After each signal, teammates in the no communication treatment can update their beliefs following the procedure described in section 3.2.2 (for the procedure after signal 1, see Figure 3.1, beliefs D.1 and D.2). Analog, teammates in the communication treatment have to find a unanimous belief following the previously mentioned procedure.

### 3.2.4 The calculation task

At the start of the experiment, participants were told that they would have to complete a timed calculation task. This task is a self-serving or ego-relevant task which requires ability and effort to be solved. This feature is important as self-serving tasks have been shown to cause asymmetric updating while self-irrelevant tasks do not (see e.g. Eil and Rao 2011; Ertac 2011). This task had to be done by everybody individually to assure that we would not get any synergy effects. The task consisted of 25 calculations, for which participants had a maximum of five minutes (for the exact calculations, see Appendix A.4). Each calculation showed five randomly drawn two-digit numbers which had to be added (e.g.  $93 + 53 + 32 + 69 + 19 = \square$ ). Only one calculation was visible at a time and could not be revisited once an answer had been submitted. Participants were not allowed to skip calculations. The sequence of 25 calculations was the same for everyone.

We deliberately chose a calculation task for two reasons. First, there is no performance difference between men and women (see e.g. Hyde *et al.* 1990; Catsambis 1994; Guiso *et al.* 2008; Hyde *et al.* 2008). Even though, the gender of participants is not salient, we want to single out the effect of communication without confounding performance effects. Second, we chose a test where subjects care about the outcome but where English language skills do not matter. Most papers on belief updating use an IQ test which combines logical and English language skills (e.g. Möbius *et al.* 2007, 2014; Eil and Rao 2011; Buser *et al.* 2016; Cacaault and Grieder 2018) or mathematical tasks and English language skills (Coutts, 2018). As subjects in our sample mostly are not English natives, we decided to go with a purely mathematical task.

### 3.2.5 The payoff

Subjects were explicitly told that throughout the experiment they would have to complete multiple payoff-relevant tasks of which only one would be chosen at random and implemented as their payoff-determining decision. With a likelihood of 50%, the calculation task was chosen to be payoff-relevant. The higher likelihood to be paid for the team score was designed to further strengthen team identification among individuals. For each correctly solved calculation as a team, subjects would earn 0.50 euro in case the calculation task was selected for payment. The remaining 50% were equally spread across all belief elicitation tasks (and two additional small experiments). Participants in the no communication treatment were told that only their own belief was payoff-relevant to them. Participants in the communication treatment were told that only their unanimous team belief was payoff-relevant to them.

### 3.2.6 Methodology

With our methodology to elicit overconfidence through asymmetric updating behavior, we adhere to the critique of Benoît and Dubra (2011). They show that the better-than-average effect does not unambiguously show overconfident behavior even if the majority of people believes to be better than average as these beliefs could be rationalized in a Bayesian way given a certain distribution of priors. To infer unjustifiably optimistic beliefs according to Bayes' Law, subjects would have to value positive feedback more than negative feedback. This behavior is called asymmetric updating as new information is valued asymmetrically and leads to overconfident beliefs in the long-run. Several papers have already looked at asymmetric updating (e.g. Möbius *et al.* 2007, 2014; Eil and Rao 2011; Cacault and Grieder 2018; Coutts 2018).

For the empirical analysis, we follow the approach by Möbius *et al.* (2014). Suppose a subject finishes the calculation task and has a prior belief  $\mu^t = P(A = H)$  that his true ability  $A$  is high  $H$  and assume this subject observes a high signal  $h$ . A subject receives a high signal  $x = h$  with probability  $p$  if his true ability is high and with probability  $q$  if his true ability is low. We employ symmetric signals with  $p = 0.75$  and  $q = 0.25$  which means that the signal accuracy is 75% irrespective of the true ability. People of high ability receive a high signal with a 75% chance and people of low ability receive a low signal with a 75% chance. The subject then derives his posterior belief  $\mu^{t+1}$  following Bayes' Law:

$$\mu_{t+1} = \frac{P(x = h|A = H) * \mu^t}{P(x = h|A = H) * \mu^t + P(x = h|A = L) * (1 - \mu^t)} \quad (3.1)$$

This translates into

$$\text{logit}(\mu_t) = \text{logit}(\mu_0) + X_h^t * \ln\left(\frac{p}{q}\right) + X_l^t * \ln\left(\frac{1-p}{1-q}\right) \quad (3.2)$$

where  $\text{logit}(z) = \ln\left(\frac{z}{1-z}\right)$  and  $X$  the number of high, respectively low signals received at time  $t$ .<sup>4</sup> This allows us to empirically test whether subjects update optimally by using the following specification:

$$\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_t = h) * \lambda_h + \beta_l * I(x_t = l) * \lambda_l + \epsilon_t \quad (3.3)$$

where  $\lambda_h = \ln\left(\frac{p}{q}\right)$  and  $\lambda_l = \ln\left(\frac{1-p}{1-q}\right)$  denote the loglikelihood ratios of positive and negative signals. For an optimal Bayesian updater, the coefficients equal one, i.e.  $\delta = \beta_h = \beta_l = 1$ .

<sup>4</sup> For the derivation, see Cacault and Grieder (2018).

### 3.2.7 Restriction of the sample

To account for confusion, misunderstanding of the experiment, or factors potentially indicating that subjects were otherwise not committed to attentively follow the experiment, we exclude all pairs which never update (one non-communicating individual and two communicating pairs) or update in the wrong direction at least once (14 non-communicating individuals and 13 communicating pairs). This is in line with the procedure of previous papers (Möbius *et al.*, 2007, 2014; Coutts, 2018). In total, we exclude 15 individuals (17.9%) and 15 communicating pairs (i.e. 30 participants, 36.6%). This leaves us with 121 participants (26 communicating pairs and 69 non-communicating individuals).

We further exclude observations from the analysis when the prior belief is either below 10% or above 90%. The rationale behind this exclusion is that minor deviations from optimal Bayesian updates have a huge impact on the updating function if priors lie within the extreme part of the distribution but have almost no effect if priors are close to 50%. Consider Jake and Lindsey, who both deviate from optimal Bayesian updating by just one percentage point. Jake has a prior of  $\mu^0 = 50\%$  and receives a positive signal. According to Bayes' Law, he should update his belief to  $\mu^1 = \frac{0.75*0.5}{0.75*0.5+0.25*(1-0.5)} = 75\%$ , that is by 25 percentage points. Given his one percentage point deviation, he ends up with a posterior belief of  $\hat{\mu}^1 = 76\%$ . Jake therefore overreacts by  $\frac{26\%-25\%}{25\%} = 4\%$  compared to the optimal Bayesian updater. Now consider Lindsey with a prior belief of  $\mu^0 = 95\%$  who also receives a positive signal. Her optimal Bayesian posterior is  $\mu^1 = \frac{0.75*0.95}{0.75*0.95+0.25*(1-0.95)} \approx 98.3\%$ , that is she has to update by 3.3 percentage points. Lindsey also deviates by one percentage point which results in  $\hat{\mu}^1 = 99.3\%$  or a relative deviation of  $\frac{4.3\%-3.3\%}{3.3\%} \approx 30.3\%$ . In other words, while both overreact to a positive signal by one percentage point, we only observe a slight relative overestimation with Jake but a significant relative overestimation with Lindsey.

This algebraical effect has a strong impact on our regression analysis. For visualization, we ran simulations mimicking 1,000 individual belief updating processes over the course of four signals analog to the design of our experiment. Column (1) of Table A3.1 shows the pooled regression results when all participants update optimally Bayesian. All coefficients equal one as follows directly from equation 3.2.<sup>5</sup> In column (2), we simulate that individuals with a prior belief between 50% and 60% respond one percentage point stronger to positive feedback than an optimal Bayesian; the response to negative feedback remains optimally Bayesian. Subjects with priors below 50% and above 60% update optimally Bayesian for all signals. Take Jake from the example above. Given his prior belief of  $\mu^0 = 50\%$ , he falls within the relevant belief decile and updates to  $\mu^1 = 76\%$  upon receiving a positive signal (instead of the optimal 75%). One round later, he further updates  $\mu^1 = 76\%$  after receiving another signal. Since this belief does not lie within the decile from 50% to 60% anymore, he now updates optimally Bayesian.

---

<sup>5</sup> Coefficients are significantly different from one as we include i.i.d. noise.

In other words, we create an environment which allows us to understand the impact that a one percentage point deviation from optimal Bayesian updating within a certain part of the belief distribution has on the *entire* regression. In columns (3) and (4), we change the asymmetric updating decile to the 9th, respectively 10th decile.

The results show that a one percentage point deviation in response to a positive signal has basically no effect if observed only in the sixth decile with a coefficient of  $\beta_H = 1.0041 \approx 1$ , while there is a small effect observable if subjects deviate in the ninth decile ( $\beta_H = 1.0213$ ). If belief updates are asymmetric by one percentage point in the tenth decile, however, we observe a huge effect. Even though, subjects by construction deviate in the tenth decile in absolute terms as much as in the 6th or 9th decile, the coefficient of  $\beta_H = 1.1088$  suggests that subjects now update strongly asymmetrically. This effect can equally be observed in the first decile and for overreacting to negative signals.

It is unrealistic to expect such updating precision of subjects in our experiment, especially since we limit their updates to integers. We therefore exclude observations where the prior belief lies within the first or the tenth decile from our analyses to avoid that extreme beliefs drive our results. Importantly, we only exclude those observations where the prior belief lies within the first or tenth decile. If, after an update, the belief lies within the range of 10% to 90% again, their next update is included. Example: assume, a pair starts with a prior of 95%, receives a negative signal, and updates downwards to 80%. The second signal is positive, and the pair updates their belief back to 95%. Then we exclude their first update from our regression as the pre-signal prior (95%) lies within the tenth decile. Yet, we keep their second update as their pre-signal prior (80%) now does not lie within the tenth decile any longer. This leads to a total exclusion of 2.3% or 11 observations (9 non-communicating individuals, 3.3%; and one communicating pair, 1.0%). We show in section 3.5 that the inclusion of all observations does not change our results.

## 3.3 The data

### 3.3.1 The participants

The experiment was programmed using ztree (Fischbacher, 2007) and was run in November 2017 at Maastricht University. We invited a total of 2,487 students of which 163 (6.6%) participated in the experiment. Participants were on average 20.7 years old, 49.6% were female. After excluding those who never updated and/or updated in the wrong direction at least once, 121 subjects (74.2%) remain for our analysis. Those who are excluded do not systematically differ from those who remain in the sample in terms of their gender (excl: 60% female vs incl: 50% female; p-value = 0.2350) or age (excl: 21.0 vs incl: 20.7; p-value = 0.5838). Importantly, they also do not differ regarding their self-reported math skills. We asked subjects to answer the

statement *"My math skills are higher than those of \_\_\_ of my fellow students."* on a 11-point Likert scale from 0% to 100% in steps of ten percentage points. Those excluded reported to be better than 60.9%, those who remained in the sample reported to be better than 65.5% (p-value = 0.2692).

### 3.3.2 Updating behavior and calculation scores

The 121 subjects solved 8.3 calculations in 5 minutes (see Table A3.2). There was no significant difference in individual scores between the non-communicating and communicating treatment (NC: 8.2 vs C: 8.5; p-value = 0.6180) which shows that our randomization worked. The distribution of calculation scores is slightly positively skewed. When we exclude scores of 16 and higher (3.0% of all scores), the distribution is approximately normal (Shapiro-Wilk test, p-value = 0.1922). Pairs across both treatments on average score 16.6 on the calculation task with a median score of 17. The difference between team scores across both treatments is not significant (NC: 16.2 vs C: 17.0; p-value > 0.4122).

### 3.3.3 Subjects in non-communicating pairs barely react to their teammate's belief

The critical difference between the two treatment groups is the level to which they can resolve their disagreement. While communicating pairs can chat, subjects in the non-communicating pairs have no direct way to discuss opposing beliefs. Among communicating pairs, 23 of 26 pairs find a unanimous belief in all five rounds. The remaining three pairs find a unanimous belief in four of five rounds.<sup>6</sup> On the contrary, non-communicating pairs form unanimous beliefs in 1.1 of five rounds (difference between treatment groups: p-value < 0.0001).<sup>7</sup>

Nevertheless, also subjects in the no communication treatment have the possibility to react to the teammate's belief. Figure A3.3 shows the evolution of disagreement among non-communicating pairs. Specifically, it plots the difference between the beliefs of the two teammates at the end of each round (difference in beliefs C.2, D.2, etc.). Compared to communicating pairs, where this difference by construction is zero (unanimous beliefs), the beliefs of subjects in non-communicating pairs differ significantly. Notably, the difference decreases only slightly from 10.4 to 7.4 percentage points (p-value = 0.0575) with an average of 9.0 percentage points. To underscore the observation of no reaction, we find that subjects react to the belief of their teammate on average only in 1.9 of five rounds. In other words, most of the time, subjects in non-communicating pairs just re-enter their individual belief after the protocol phase. Yet,

---

<sup>6</sup> We exclude beliefs of communicating pairs which did not manage to find a unanimous belief from our main analysis.

<sup>7</sup> As coordination is impossible for non-communicating pairs, unanimous team beliefs are to a certain extent a product of luck.

especially in round one (directly after the calculation task), we observe a strong reaction to the teammate’s belief. While beliefs in non-communicating pairs differ by 17.8 percentage points before observing the teammate’s belief (belief C.1), the difference significantly decreases to 10.4 percentage points at the end of the round (belief C.2;  $p$ -value  $< 0.0001$ ). Over the course of five rounds, teammates react less and less to the belief of each other. Thus, non-communicating pairs do take the belief of the teammate into consideration—but only in the beginning. The restriction of not being able to discuss differing beliefs causes subjects to stick to their own beliefs.

As stated above, joint beliefs of communicating pairs do not differ by construction. However, teammates do not start with the same beliefs either. In the first round, where beliefs in non-communicating pairs differ by 17.8 percentage points, the beliefs of teammates in communicating pairs differ by a similar 18.1 percentage points (difference between treatment groups,  $p$ -value = 0.9146). But communicating pairs strongly react to each other, not only by finding a unanimous belief, but also by planning on how to react to a future signal. While in the first round (directly after the calculation), only 7.7% of all communicating pairs have the same belief before even getting the chance to chat (belief C.1), 42.3% have the same individual belief before chatting in the final round (after the fourth signal). Thus, we observe that a significant share of communicating pairs actually uses the chatting time to coordinate their actions even beyond the current round.

These developments of beliefs nicely show the intended difference between the two treatment groups, where disagreement can and has to be resolved in communicating pairs while non-communicating pairs remain in disagreement. In the next section, we analyze whether the disagreement and coordination of beliefs in communicating pairs is the result of a thorough examination of individual beliefs (*the force of disagreement*) or whether it gives way for biased beliefs to prevail (*the force of dominance*).

## 3.4 Results

### 3.4.1 Paring causes subjects to form underconfident prior beliefs

The average prior belief that the individual score will be in the top half of all individual scores (see Figure 3.1, belief A) is 55.8%. This belief is slightly but significantly higher than the unconditional prior of 50% ( $p$ -value = 0.0003). After subjects learn that they have been paired, they are asked to assess the likelihood that their team score is in the top half (belief B). The average belief B is with 52.7% significantly lower than belief A ( $p$ -value = 0.0222).<sup>8</sup> The decline in confidence is a rational response to the higher level of uncertainty given that no information on

<sup>8</sup>Beliefs between treatment groups do not differ as subjects did not learn about their treatment until after the calculation task (NC: 53.6% vs C: 51.5%;  $p$ -value = 0.3362).

the teammate was provided. This increased level of uncertainty can further be observed looking at the correlation between beliefs and the corresponding scores. While the correlation between belief A and the individual score (IS) is with  $\rho(A, TS) = 0.2435$  positive and highly significant (difference from zero, p-value = 0.0071), the correlation between belief B and the team score (TS) is lower and only marginally significant ( $\rho(B, TS) = 0.1457$ ; p-value = 0.1108). While the two correlations are not significantly different from each other (Fisher r-to-z transformation, p-value = 0.2177), subjects do seem to struggle when trying to assess the team performance.

The provision of additional information in form of the team score and the teammate's belief C.1 should reduce the level of uncertainty. This is exactly what we observe as the correlation between the first belief formed in pairs (belief C.2) and the team score is with  $\rho(C.2, TS) = 0.4649$  significantly higher than it was before the calculation task ( $\rho(B, TS) = 0.1457$ ; Fisher r-to-z transformation, p-value = 0.0031). However, the level of confidence further drops to an average belief C.2 of 43.7% among non-communicating beliefs and 44.8% among communicating pairs (difference between treatment groups, p-value = 0.7119). Both types of pairs hold beliefs which are significantly lower than the unconditional mean of 50% (p-values < 0.0045). Pairs appear to hold *underconfident* beliefs—regardless of whether they can communicate or not.

The empirical cumulative distribution functions of beliefs A, B, and C.2 in Figure A3.4 visualize the above-described constant decline in average beliefs. The distribution of belief A is furthest to the right indicating that subjects are most confident about their individual score. It is also then that subjects hold the most extreme beliefs to either side of the distribution. Upon receiving additional information on being paired (belief B) respectively the actual performance (belief C.2), the distribution shifts further to the left which translates into more respondents holding lower beliefs. All distributions are significantly different from each other (Kolmogorov-Smirnov test, p-values < 0.001).

### 3.4.2 Pairs update symmetrically

Of the four signals received, non-communicating pairs react to 3.2 and communicating pairs to 3.3 (difference between treatments: p-value = 0.5263). Thus, pairs respond to almost all signals irrespective of their treatment. While we have so far focused on the absolute and relative levels of beliefs, the analysis of how pairs respond to feedback is more critical (Benoît and Dubra, 2011). We can only talk about overconfidence (or underconfidence in our case) if pairs structurally overweigh positive (negative) feedback. We employ the empirical approach which has widely been used to study asymmetric updating (see equation 3.3, section 3.2.6; e.g. Möbius *et al.* 2007, 2014; Eil and Rao 2011; Coutts 2018; Cacaault and Grieder 2018). Figure 3.2 shows the difference of the regression coefficients from the response to positive and negative feedback,  $\beta_H - \beta_L$ , for non-communicating and communicating pairs separately (see Table 3.1 for the complete

regression results).<sup>9</sup> With  $\beta_H - \beta_L = 9.1$  percentage points, non-communicating pairs respond slightly stronger to positive signals, yet the difference is not significant (p-value = 0.1357). Communicating pairs update perfectly symmetrically with  $\beta_H - \beta_L = -0.2$  percentage points (p-value = 0.9800). Thus, resolving disagreement seems to have a positive impact on belief updates in communicating pairs. Importantly, however, not being able to resolve disagreement does not cause pairs to update worse as the difference in asymmetry between the two treatment groups is non-significant (p-value = 0.2784) suggesting that pairs respond to signals equally well regardless of whether communication is allowed.

This result is further corroborated when we look at how the average beliefs develop across all four updating rounds. Bayesian updating predicts that pairs which are in the top half increase their beliefs closer to 100% when receiving informative signals while pairs in the lower half decrease their beliefs towards 0%. Consequently, after an infinite number of signals, all pairs in the top half would hold beliefs of 100% and all in the bottom half of beliefs of 0% resulting in an average belief of 50% across pairs in either half. Figure 3.3 presents the evolution of beliefs for each treatment group. We recognize the beliefs after the calculation (called 'Prior' in Figure 3.3) which are significantly lower than 50%. In the following four rounds, non-communicating and communicating pairs work their ways to the 50% benchmark arriving at an average belief in round four which no longer differs from the true mean of 50% (NC: 51.4% vs C: 52.7%; p-values > 0.8176).

Thus, both, the conventional way of looking for overconfident behavior through the better-than-average effect (e.g. Hoelzl and Rustichini 2005) as well as by acknowledging the Benoît and Dubra (2011) critique provide the same result:

**Result 1:** Pairs show no sign of overconfidence in an ego-relevant task and update symmetrically regardless of whether communication is possible or not.

---

<sup>9</sup> An optimally Bayesian updating subject has to be invariant to his prior beliefs which is shown by the coefficient  $\delta = 1$  as well as fully responsive to new information which is the case if  $\beta_H = \beta_L = 1$ . The latter we come back to in section 3.4.4. Being invariant means that the response to positive or negative signals must not depend on the level of the prior belief but only on the past signal itself. In other words, the response to signals must not be a function of the prior and therefore subjects with a prior of 20% should not update more or less optimally Bayesian than subjects with a prior of 50% or 80%. We can clearly reject these properties suggesting that pairs of either treatment do not update optimally Bayesian as they are not invariant to their prior belief ( $\delta < 1$ ; e.g. Möbius *et al.* 2007, 2014; Coutts 2018; Ruof and Smeets 2018).

Overall, the correlation we find between actual posterior beliefs and optimal posterior beliefs calculated using Bayes' Law is  $\rho = 0.62$  (p-value < 0.0001) which lies in-between the correlation coefficients found by Coutts (2018,  $\rho = 0.89$ ), Möbius *et al.* (2014,  $\rho = 0.57$ ), and Ruof and Smeets (2018,  $\rho = 0.77$ ). Thus, apart from finding pairs to update symmetrically, the reaction to signals and the deviation from optimal Bayesian updates in the present paper is similar to previous findings.

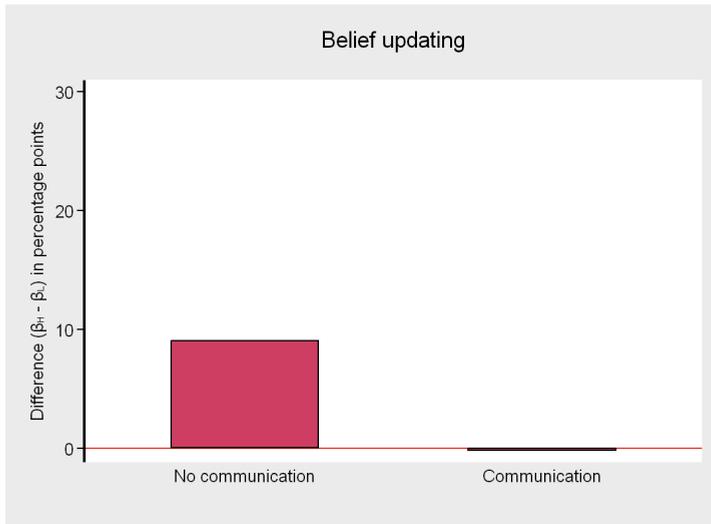


FIGURE 3.2 – **Asymmetry and communication.**

The graph plots the difference between coefficients  $\beta_H$  and  $\beta_L$  from Table 3.1. Separate regressions were run for each treatment. Updating according to Bayes' Law requires all bars to be equal to zero. A positive difference  $\beta_H > \beta_L$  implies asymmetric belief updating where positive feedback is valued stronger than negative feedback. Communicating pairs and individuals in non-communicating pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are excluded. Only the final decision per round is considered.

### 3.4.3 Dominance and updating

Since communicating pairs do not suffer from asymmetric belief updating, the force of dominance cannot be a driving force on the aggregate level. In this section, we therefore investigate whether dominance has any influence of the updating behavior. To do so, we look at pairs which host a dominant individual.

We asked respondents to rate the statement *“I see myself as a dominant person.”* on a 7-point Likert scale (1—fully disagree to 7—fully agree). The average rating was 4.3 and does not differ across treatments (NC: 4.3, C: 4.3; p-value = 0.8875). We define the teammate with the higher rating to be the dominant subject of a pair. The average of the higher dominance scores does not differ across treatments (NC: 5.4, C: 5.2; p-value = 0.3634). We group pairs in above- and below-median groups with regard to the higher dominance score (two regressions per treatment).

Table 3.2 shows that among communicating pairs the presence of an above-median dominant person leads to an asymmetric response of  $\beta_H - \beta_L = 22.9$  percentage points (p-value = 0.1618). For communicating pairs with a below-median dominant subject, the response turns negatively asymmetric indicating that they overweigh negative signals with  $\beta_H - \beta_L = -11.8$

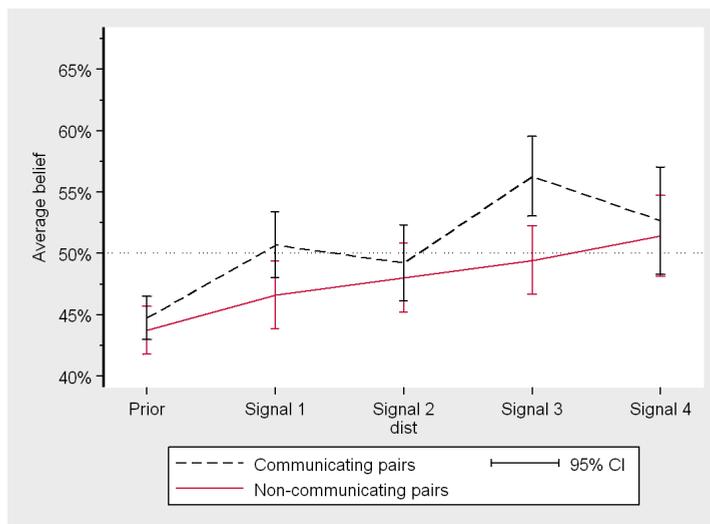


FIGURE 3.3 – Evolution of beliefs by treatment.

The graph plots the average beliefs of pairs that the team score is in the top half of all team scores for non-communicating and communicating pairs separately for each of the five experimental rounds. Communicating pairs and individuals in non-communicating pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are excluded. Only the final decision per round is considered. Non-overlapping error bars indicate a statistically significant difference at the 5% level.

percentage points ( $p$ -value = 0.2091). While neither difference is significant itself, the difference in asymmetry between above- and below-median pairs is highly significant ( $p$ -value = 0.0015)<sup>10</sup>.

This harmful effect is reduced among non-communicating pairs but does not fully disappear. Non-communicating pairs with an above-median dominant subject also update asymmetrically ( $\beta_H - \beta_L = 16.2$  percentage points; difference to zero:  $p$ -value = 0.0966) while those with a below-median dominant subject ( $\beta_H - \beta_L = -1.0$  percentage points; difference to zero  $p$ -value = 0.8800) do not. Similar to communicating pairs, this asymmetry is driven by those individuals who rate themselves as above-median dominant. The effect is likely less pronounced as subjects cannot actively dominate the decision of their teammates. When we then take the average of two beliefs where one is biased, the team belief necessarily also is biased.

Overall, our findings are in line with Brown and Sarma (2007) who find dominance to be positively correlated with overconfidence. We see that the force of dominance exists and does influence communicating pairs where a higher level of dominance leads to more asymmetric updating. Yet, the effects are not very strong and completely vanish once we include all observations (see section 3.5). Regardless, the results suggest that the upside of the ability to

<sup>10</sup> To test for the difference in asymmetry across treatment groups, we use a seemingly unrelated estimation using the error terms of the OLS regressions (Zellner, 1962).

TABLE 3.1 – Belief updates in pairs with and without communication.

This table presents pooled OLS estimates for each subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , where  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . Communicating pairs and individuals in non-communicating pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are excluded (for a more detailed explanation of exclusions, see section 3.2.7). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

	(1) No communication	(2) Communication	(3) p-values (1) vs (2)
$\delta$	0.9284** (0.0353)	0.7938*** (0.0537)	0.1486
$\beta_H$	0.6436*** (0.0406)	0.6864*** (0.0559)	0.5678
$\beta_L$	0.5530*** (0.0454)	0.6887*** (0.0736)	0.1170
$P(\beta_H = \beta_L)$	0.1357	0.9800	0.2784
Observations	267	206	
R-squared	0.8400	0.7412	

communicate, namely the force of disagreement, clearly outweighs the downside through dominant behavior. Even more, we find that pairs in general update beliefs strikingly well. They barely show any sign of asymmetry and follow optimal Bayesian beliefs very closely. The consequence is that the initially underconfident pairs end up with a belief that is non-differentiable from the true mean.

**Result 2:** Dominant individuals cause asymmetric updating. This effect is most pronounced in communicating pairs. However, the force of disagreement outweighs the force of dominance among communicating pairs and leads to symmetric updates.

### 3.4.4 Pairs update conservatively

While we find pairs to update symmetrically, the question is whether they can also overcome conservative belief updating. Conservative updating is an often-observed human trait (Huck and Weizsäcker, 2002; Möbius *et al.*, 2007, 2014; Tingley and Wang, 2010; Buser *et al.*, 2016; Coutts, 2018; Cacault and Grieder, 2018) and signifies that information is not fully incorporated into beliefs. Various causes have been shown to cause this conservatism, for example having

TABLE 3.2 – Belief updates and dominance.

This table presents pooled OLS estimates for each subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_H * I(x_{it} = h) * \lambda_h + \beta_L * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , where  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . *Dominance*: We asked respondents to rate the statement “I see myself as a dominant person.” on a 7-point Likert scale (1–fully disagree to 7–fully agree). We define the teammate with the higher rating to be the dominant subject of a pair. We then group pairs in above- and below-median groups regarding this higher dominance score (two regressions per treatment). Communicating pairs and individuals in non-communicating pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are excluded (for a more detailed explanation of exclusions, see section 3.2.7). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	No communication		Communication	
	(1) dominant	(2) not dominant	(3) dominant	(4) not dominant
$\delta$	1.0466 (0.0526)	0.7582*** (0.0405)	0.6127*** (0.0990)	0.9255 (0.0539)
$\beta_H$	0.7635*** (0.0676)	0.5244*** (0.0411)	0.8820 (0.0886)	0.4779*** (0.0583)
$\beta_L$	0.6015*** (0.0708)	0.5342*** (0.0490)	0.6532** (0.1321)	0.5960*** (0.0750)
$P(\beta_H = \beta_L)$	0.0966	0.8800	0.1618	0.2091
Observations	129	138	104	102
R-squared	0.8565	0.8608	0.6506	0.8760
p-values	(1) vs (2)		(3) vs (4)	
Diff. in asym.	0.0378		0.0015	

to assess unlikely events (Vlek, 1965; Vlek and van der Heijden, 1967) or having to aggregate multiple pieces of information (Edwards, 1968). Experience with a task, on the other side, can decrease conservative assessments (Wheeler and Beach, 1968). Since groups are hypothesized to aggregate information more efficiently (Snizek and Henry, 1989) and have been shown to profit from the experience of its members in making assessments (Bonner *et al.*, 2002), groups should consequently update less conservatively. Yet, we do not find support for this hypothesis among pairs.

Table 3.1 shows that the coefficients  $\beta_H$  and  $\beta_L$  are significantly smaller than one in both treatments (p-values < 0.0001; no difference between treatments: p-values > 0.1169) while the full incorporation of available information requires  $\beta_H = \beta_L = 1$ . The consequences of conservative updating can be quite significant. Consider a communicating pair. Suppose this pair

starts with a prior belief of 50% and receives a positive signal on its performance. Plugging the values for  $\delta$ ,  $\beta_H$  and  $\beta_L$  into equation (3.3) (for values, see Table 3.1), we obtain a posterior belief of 68.0% instead of the optimal Bayesian posterior  $\mu^1 = \frac{0.75 * 0.5}{0.75 * 0.5 + 0.25 * (1 - 0.5)} = 75\%$ . Thus, the pair updated its belief by only 18.0 of the required 25.0 percentage points, or by a mere 72%. Similarly, if a pair starting with a prior belief of 50% receives a negative signal, it would update its belief downwards to a belief of 31.9% instead of the optimal belief of 25%.

**Result 3:** Pairs update conservatively regardless of whether communication is allowed. Conservative updating means that pairs do not fully incorporate new information.

### 3.5 Robustness check of the main findings: including everyone

In the main analysis, we excluded 15 communicating pairs (i.e. 30 participants) and 15 individuals in non-communicating pairs who never updated or updated at least once contrary to their signal (see section 3.2.7). This procedure accounts for confusion, misunderstanding of the experiment, or factors potentially indicating that pairs were otherwise not committed to attentively follow the experiment. In addition, we excluded eleven extreme observations (beliefs below 10% and above 90%) to account for the lack of precision available for participants with extreme beliefs to adequately update their beliefs. In this section, we show that our main result is robust to including these pairs while the findings on the influence of dominance is not.

Table A3.3 presents the results of a pooled OLS regression of equation 3.3 where all pairs are included. The findings largely reproduce our main results from Table 3.1. For non-communicating pairs, the difference between their response to positive and negative signals is with  $\beta_H - \beta_L = 9.5$  percentage points similar in size and even marginally significant (difference to zero: p-value = 0.0990). Communicating pairs show a slightly positive but highly non-significant difference of  $\beta_H - \beta_L = 1.7$  percentage points (p-value = 0.8418). These findings confirm that while pairs in general update relatively symmetrically, especially communicating pairs update close to perfectly symmetrically. This supports our result that the force of disagreement causes a thorough examination of individual beliefs and results in a debiased team belief which does not lead to unjustifiably optimistic beliefs in the long run. We can conclude that our findings are robust to including all pairs and thus are not driven by participants who understood the experiment or were not confused.

In Table A3.4, we test the robustness of our results on the presence of a dominant individual within a pair. None of the results shown in Table 3.2 are robust to including all observations. This finding provides an indication that the force of dominance does not matter. We can show that the inclusion of participants who update in the wrong direction are the cause for the

disappearing effects. It therefore remains to be studied whether our results lack power or whether the theory is not robust to including participants who update in the wrong direction.

### 3.6 Discussion and Conclusion

Overconfidence has consistently been shown to negatively affect individual decisions (e.g. Camerer and Lovallo 1999; Barber and Odean 2001; Biais *et al.* 2005). However, decisions are not always taken by individuals, but often times in teams of two where overconfident beliefs can be equally harmful. We run an experiment in which pairs complete an ego-relevant task and receive feedback on their performance. We analyze whether communication has a significant impact on how pairs update their beliefs. Our results show that communication has no impact but that all pairs update symmetrically and thus do not develop unjustifiably optimistic beliefs in the long run.

These findings pose a stark contrast to the various studies looking at individuals and provide indicative evidence that pairs form more accurate beliefs in ego-relevant tasks than individuals. Especially the studies which use a design closest to ours find individuals to strongly overweigh positive feedback in ego-relevant tasks (see Möbius *et al.* 2007, 2014; Eil and Rao 2011; Cacault and Grieder 2018 as well as Buser *et al.* 2016, but only if subjects are included who update contrary to a signal). This comparison between individual and pair belief updates raises an interesting question. There are two critical differences between our design and these papers which could explain the difference between no asymmetry here and asymmetric updating patterns there. First and most obvious, our subjects are part of a team while the mentioned papers study individuals. Yet, Healy and Pate (2007) as well as Cacault and Grieder (2018) have found group identification to cause overconfidence. This difference would thus go against our finding. Second, subjects in our control group receive one additional piece of information per round, namely the belief of their teammate. The question thus is whether the provision of an additional belief suffices to make individuals update symmetrically? While our results could suggest this conclusion, it could also be that subjects feel being observed by the teammate ("What might he think about my belief?"). It will be for future research to investigate more deeply whether and how exactly pairs help to reach symmetric belief updates.

Further, it has to be noted that not all studies find positive asymmetric updating behavior. Ertac (2011) and Coutts (2018) find individuals to update asymmetrically by overweighing negative feedback. While it remains unclear under which circumstances individuals overweigh positive signals and under which they overweigh negative signals, one crucial design difference between studies of either finding has to be mentioned. Studies which report an overweighing of negative signals do not ask subjects to assess the likelihood of being in the top half, but of being in the top 20% (Ertac, 2011), respectively top 15% (Coutts, 2018). This change causes the average subject to receive more negative signals compared to a top 50% split where the average

subject receives an equal number of positive and negative signals. Coutts (2018) even shows that subjects overweight signals which are received less often and thus offers an explanation of why they find an overweighing of negative feedback. Further studies are needed to show whether this difference in the average feedback is causal for the difference in overweighing positive, respectively negative signals. More importantly, it remains to be shown whether pairs also update more accurately in experiments where receiving negative signals is more likely.

If the finding of pairs updating more symmetrically than individuals should materialize, these results have important implications for various daily situations where decisions are still made by individuals. This could include policy changes for trading on stock markets, where trades could be realized only if two traders agree. Or consider the search for a new research topic, where researchers could rely on external input to judge whether the next project is worth going for. These situations resemble ego-relevant decisions where individuals tend to focus more on positive signals in favor of their plans, but where a second person can help to make better decisions by incorporating new information more accurately.

Our paper provides evidence that pairs benefit from communication rather than suffering from dominant behavior. This finding provides ample room for further research to build upon. First and foremost, pairs in our experiment were randomly matched which is a weak representation of most situations in which teams take a joint decision. Consider team-managed mutual funds which have been vastly outgrowing individually-managed mutual funds (Bliss *et al.*, 2008). It is unrealistic to assume that the managers are selected randomly. Even more, the selected managers are likely to know each other or at least have been able to learn about each others' expertise, while we kept the group discussion phase purposefully sterile to elicit the effect of communication. Thus, teammates barely knew anything about each other. Future research could lift these restrictions and provide additional information on teammates such as their knowledge about or experience with the task at hand. This would enable us to step-by-step mirror real-life situations more closely where teams usually consist of people who know each other at least to a certain extent or have enough time to elicit relevant background information. Second, pairs resemble only the smallest form of a group. It would be highly interesting to see how groups of three or five form beliefs. Asch (1956), for example, finds group conformity to play the biggest role in groups of three. Third, bigger groups, especially those with an uneven number of members, test our external rule of taking the average belief as group belief. In bigger groups, unanimous beliefs can be the result of a majority decision where some teammates get overruled. Being overruled could have unique effects on the team dynamic leading to completely different outcomes. Finally, most teams have a natural or imposed hierarchy where one person is the boss and can weigh in on the final team decision more than others. The presence of hierarchical structures would allow us to study whether discussions are still thorough enough to lead to symmetric beliefs or whether these would rather resemble the presence of a dominant individual who causes asymmetric belief updates.

## A Appendix

### A.1 Tables

TABLE A3.1 – Simulated belief updates.

This table presents pooled OLS estimates for a simulated sample of 1,000 subjects of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , where  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . Analog to our experiment, 50% of the simulated subjects are placed into the top and 50% into the bottom half (in our simulation, placement is random). Signal accuracy is 75%. Dependent variable is the posterior belief. In column (1), all subjects update optimally Bayesian. In columns (2) to (4), subjects with beliefs in the specified decile respond one percentage point stronger to positive feedback than an optimal Bayesian; the response to negative feedback remains optimally Bayesian. All subjects with beliefs outside the specified decile still update optimally Bayesian. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

	(1)	(2)	(3)	(4)
	All update optimally Bayesian	All update optimally Bayesian except if belief $\in [0.5, 0.6)$	All update optimally Bayesian except if belief $\in [0.8, 0.9)$	All update optimally Bayesian except if belief $\in [0.9, 1.0)$
$\delta$	1.0000*** (0.0000)	0.9999*** (0.0000)	1.0028*** (0.0002)	1.0340*** (0.0043)
$\beta_H$	1.0000*** (0.0000)	1.0041*** (0.0003)	1.0213*** (0.0012)	1.1088*** (0.0068)
$\beta_L$	1.0000*** (0.0000)	1.0000*** (0.0000)	0.9992*** (0.0002)	0.9633*** (0.0049)
$P(\beta_H = \beta_L)$	0.0000	0.0000	0.0000	0.0000
Observations	4,000	4,000	4,000	4,000
R-squared	1.0000	1.0000	0.9997	0.9893

TABLE A3.2 – **Summary statistics.**

This table presents summary statistics for all 121 participants which are included in our analysis.

	<b>Mean</b>	<b>Median</b>	<b>Std Dev</b>	<b>Obs.</b>
Female	50.4%			60
Male	49.6%			61
Age	20.7	20	2.1	121
<b><i>Team calculation score</i></b>				
No Communication	16.2	16	5.5	69
Communication	17.0	16.5	4.7	52
p-value	0.4122			
<b><i>Updates (1-4)</i></b>				
No Communication	3.2	4	1.0	69
Communication	3.3	4	0.9	52
p-values	0.5263			
<b><i>Dominance (1-7)</i></b>				
No Communication	4.3	4	1.5	69
Communication	4.3	4	1.6	52
p-values	0.8875			

TABLE A3.3 – Belief updates in pairs with and without communication (all participants included).

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , where  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	(1) No communication	(2) Communication	(3) p-values (1) vs (2)
$\delta$	0.9344** (0.0306)	0.7553*** (0.0502)	0.1891
$\beta_H$	0.6064*** (0.0388)	0.5071*** (0.0605)	0.2466
$\beta_L$	0.5107*** (0.0436)	0.4803*** (0.0717)	0.6963
$P(\beta_H = \beta_L)$	0.0990	0.8418	0.4766
Observations	336	328	
R-squared	0.8369	0.7746	

TABLE A3.4 – Belief updates and dominance (all participants included).

This table presents pooled OLS estimates for each subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , where  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . *Dominance*: We asked respondents to rate the statement “I see myself as a dominant person.” on a 7-point Likert scale (1–fully disagree to 7–fully agree). We define the teammate with the higher rating to be the dominant subject of a pair. We then group pairs in above- and below-median groups regarding this higher dominance score (two regressions per treatment). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	No communication		Communication	
	(1) dominant	(2) not dominant	(3) dominant	(4) not dominant
$\delta$	1.0366 (0.0456)	0.7468*** (0.0372)	0.8782 (0.0761)	0.6724*** (0.0718)
$\beta_H$	0.6440*** (0.0671)	0.5225*** (0.0395)	0.5300*** (0.0773)	0.4679*** (0.0942)
$\beta_L$	0.5637*** (0.0714)	0.5035*** (0.0459)	0.5159*** (0.1025)	0.5194*** (0.1064)
$P(\beta_H = \beta_L)$	0.4010	0.7589	0.9131	0.7184
Observations	160	176	160	168
R-squared	0.8557	0.8382	0.5724	0.5710
p-values	(1) vs (2)		(3) vs (4)	
Diff. in asym.	0.5319		0.7334	

## A.2 Figures

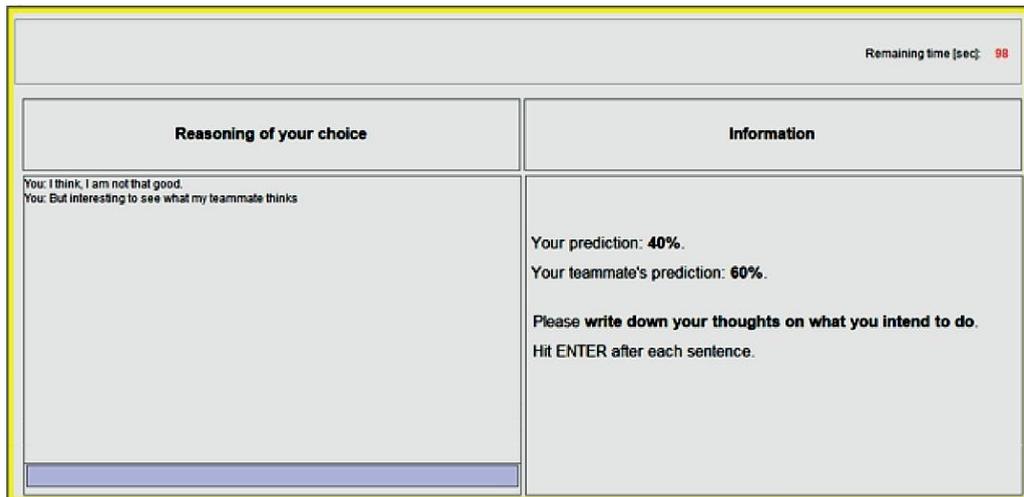


FIGURE A3.1 – Screenshot of the protocol.

On the left, the current protocol is visible including the blue entry field at the bottom. On the right, one's individual belief as well as the individual belief of the teammate is shown. Text and numbers are exemplary. The protocol was cleared after each round.

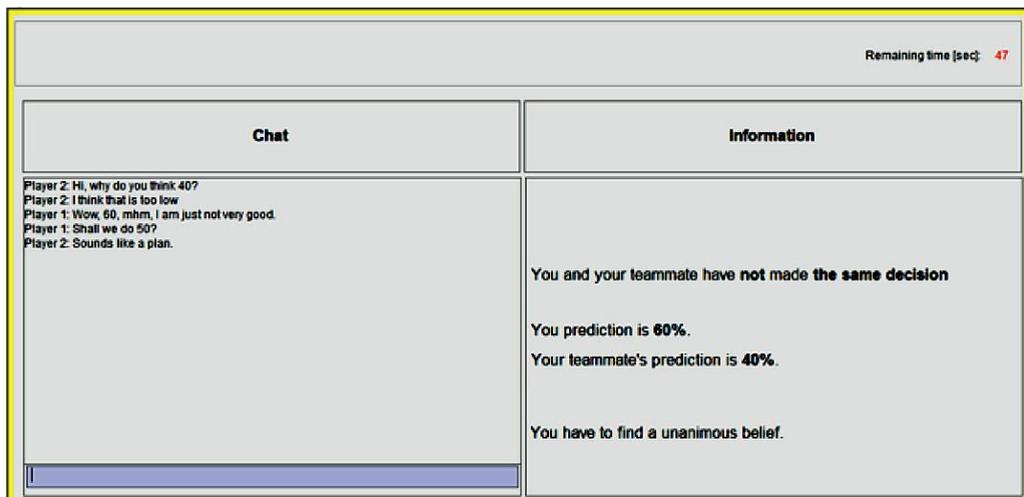
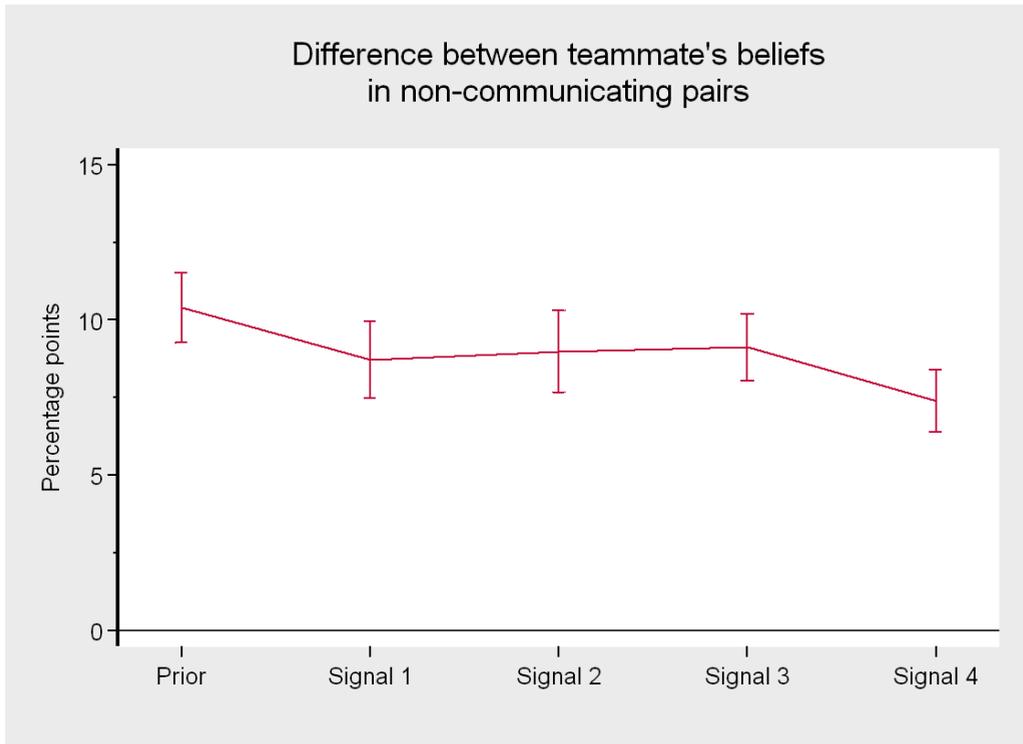


FIGURE A3.2 – Screenshot of the chat.

On the left, the current chat is visible including the blue entry field at the bottom. On the right, one's individual belief as well as the individual belief of the teammate is shown. Text and numbers are exemplary. The chat was cleared after each round.



**FIGURE A3.3 – Difference in beliefs within non-communicating pairs.**

The graph plots the difference between the final beliefs (beliefs C.2, D.2, etc.) of teammates after having seen each other's individual belief. The differences are presented for all five rounds. The graph presents the differences only for non-communicating pairs as the difference for communicating pairs by construction is zero. Individuals in non-communicating pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are excluded. Non-overlapping error bars indicate a statistically significant difference at the 5% level.

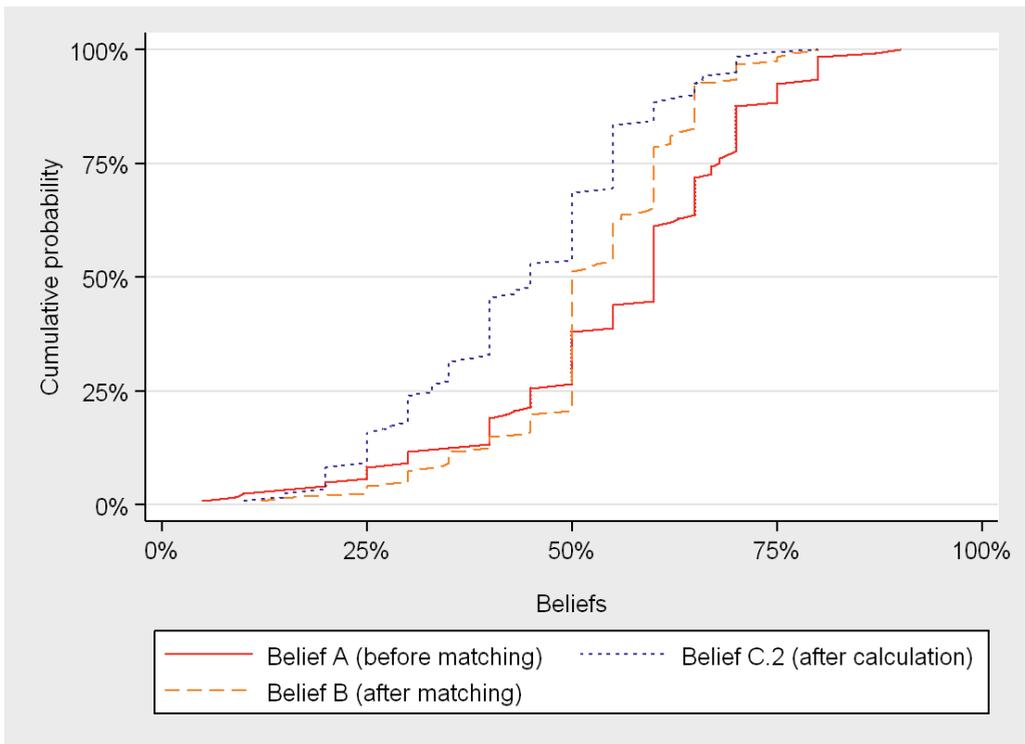


FIGURE A3.4 – **Empirical cumulative distribution functions.**

The graph plots three empirical cumulative distribution function. (1) *Belief A* (before matching and the calculation task): belief that the individual score is in the top half of all individual scores; (2) *Belief B* (after matching but before the calculation task): belief that the team score is in the top half of all team scores; (3) *Belief C.2* (after the calculation task): belief that the team score is in the top half of all team scores where non-communicating pairs could react to the belief of their teammate and communicating pairs have formed a unanimous belief. Two-sample Kolmogorov-Smirnov tests for equality of distributions indicate that all distributions are significantly different from each other ( $p$ -values  $< 0.001$ ). Communicating pairs and individuals in non-communicating pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Only the final decision per round is considered.

### A.3 Experimental instructions

**Welcome!**

Nice to have you here - and thanks for participating.

---

This session will approximately take 90 minutes.

You will have to do several tasks. In each task, you are asked to make decisions which can be relevant for your financial payoff.

One of all the decisions you make during these tasks throughout the session will be randomly chosen to determine your actual payoff. So, it is always optimal for you to answer as accurately as possible.

The payment will be in cash at the conclusion of the session.

You will be paid privately and anonymously.

All payoffs are guaranteed by Maastricht University (Netherlands).

Please click "Continue" to proceed.

---

**Before we start, quickly tell us:**

1. Where do you currently study?
  - (a) Maastricht University
  - (b) Zuid University of Applied Sciences
  - (c) RWTH Aachen
  - (d) Other university
  
2. What is your gender?
  - (a) Male
  - (b) Female

3. How old are you?

- (a) Below 18
- (b) 18-28
- (c) 29-39
- (d) 40-50

Please click "Continue" to proceed.

---

### **Example**

On the next few pages, we will go through a small experiment similar to our actual experiment.

The objective of this pre-experiment is to make you familiar with the choices you will have to make in the experiment to come.

It will be filled with a couple of examples to make life easier for you.

These examples will not be payoff-relevant but they will prepare and help you to earn more money in the actual experiment that will follow shortly.

Please click "Continue" to proceed.

---

### **Example**

The computer has just randomly drawn a number between 1 and 100 (each number has the same probability to be drawn).

The number assigned to you is \_\_\_

The computer has also randomly drawn a second number between 1 and 100. However, this number will not be shown to you.

On the next pages, we will tell you what to do. Click "Continue" to proceed.

---

**Example**

Remember the paper by Karni that was just shown to you and which proves that stating your true beliefs actually maximizes your expected payoffs?

In other words, **you maximize your expected earnings** when your submitted prediction equals your actual belief.

The details of the procedure are not crucial as long as you are aware that submitting a prediction that is not your actual belief will decrease your expected earnings. The mathematical proof is complex, but if you wish you can raise your hand and we come to your place and hand the paper to you. Otherwise, you can take a copy with you after the experiment.

Last but not least, we assure you that we are interested in your true beliefs. So for you as well as for us the best outcomes wait, if you state your true beliefs. Thanks!

Ready? Then click "Continue" to begin!

---

What is the probability that your number is higher than the second number randomly chosen by the computer?

**Your prediction (in %):** ---

Please click "Continue" to proceed.

---

### Reevaluation

We will now give you feedback on whether your number is larger or smaller than the second number. Specifically, we are going to give you a series of **four reports that say "lower" or "higher"**. Each report indicates whether your number is lower or higher than the second number but unfortunately they are not always correct.

Each of the reports you will get is **accurate with a 75% chance** and inaccurate with a 25% chance. In other words, if you get a report with the message "higher", there is a 75% chance that your number really is higher but a 25% chance that it is actually lower than the second number. Vice versa, if you get a report with the message "lower", there is a 75% chance that your number really is lower but a 25% chance that it is actually higher than the second number.

**You will receive your report on the next page.**

Please click "Continue" to proceed.

---

**Your first report:**

**lower/higher**

On the next page, you will again be asked to revise your prediction.

Make sure that you again you specify your true belief

Please click "Continue" to proceed.

---

What is the probability that your number is higher than the second number randomly chosen by the computer?

**Your prediction (in %): ---**

Please click "Continue" to proceed.

---

*The participants received three more reports following the same procedure*

---

**Now follows the real experiment!**

This was the experiment with the two random numbers.

By now you should be acquainted with the general procedure of the experiment to come.

On the next page, the actual experiment starts.

Please click "Continue" to proceed.

---

**Welcome to the first experiment!**

Please read the instructions carefully. Each decision can be relevant for your financial payoffs. Payoffs are as described on the next pages.

In a little bit, you're going to do a **calculation task**.

The task has **25 calculations** and you have **5 minutes** to do as many calculations as possible.

Please click "Continue" to proceed.

---

In all 25 calculations, you will be asked to add up five 2-digit numbers. For example:

$$17 + 24 + 38 + 70 + 28 =$$

$$89 + 55 + 13 + 11 + 22 =$$

You find a scratch paper and a pen on your desk which you can use to solve the calculations.

However, **only the result you enter into the input field will be counted**. This field can be reached by clicking in it with the mouse.

Your score will be the number of correctly solved calculation problems.

Please click "Continue" to proceed.

---

In addition, we will be hosting the **Top Half contest**.

In this contest, you try to end up in the **top half of all the scores in this calculation challenge**. In other words, try to correctly solve more calculations than 50% of all participants in this room.

As scores on the calculation tasks come in, our computer will rank all scores.

The computer will then be hosting the "Top Half contest", which gives out prizes for scores that rank in the top half.

Please click "Continue" to proceed.

---

### **Your decision in the Top Half contest**

1. We ask you to predict how likely it is that you end up in the top half of all participants. In other words, how likely do you think is it that you will have a score that is higher than at least half of the other scores in this classroom?
2. For example, if you think there is a 20% chance that you are going to be in the top half of all scores, your prediction should be 20%. Similarly, your prediction should be 65% if you think that the chance that you are going to be better than at least half of the participants in this room is 65%.

The higher the precision of your prediction is the higher is your chance of winning 10€ in the top half contest.

Please click "Continue" to proceed.

---

### **Your decision in the Top Half contest**

Remember: You maximize your expected earnings when your submitted prediction equals your actual belief that you end up in the top half.

Staying in the example stated above, if you believe that your true chance to end up in the top half is 65%, then deviating from this belief by stating more, e.g. 70%, or less, e.g. 64%, decreases your chance of being paid 10€. Only by stating your true belief of 65% are you most likely to win the 10€.

### 3 Dare to Pair!

---

In sum, it is in our interest to learn your true beliefs and in your interest to communicate your true beliefs to maximize your earnings. This holds also true for all remaining beliefs that we will ask of you.

Do you have any further questions? Then please raise your hand and we will come to you.

Please click "Continue" to proceed.

---

#### **Your Prediction**

What is the probability that you will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief A from Figure 3.1*

Please click "Continue" to proceed.

---

*Timer: 60 seconds*

#### **Reasoning of your choice**

Please state the reasons why you came to this prediction.

---

**You have been matched to another participant in this room.**

The two of you form a team for the remainder of this experiment.

In a bit, both of you will individually complete the calculation task. But there are some modifications to what you already read.

You receive one score as a team which is the sum of your score and the score of your teammate. Thus, your team score equals the total number of correctly solved problems between the two of you.

You will get to see your team score directly after the calculation task.

---

Note: everybody in this room is now in a team of 2.

Please click "Continue" to proceed.

---

### **Your decision in the Top Half contest**

1. We ask you to predict how likely it is that **your team** ends up in the top half of all participants. In other words, how likely do you think is it that **your team** will have a score that is higher than at least half of the other team scores in this classroom?
2. For example, if you think there is a 20% chance that **your team** is going to be in the top half of all scores, your prediction should be 20%. Similarly, your prediction should be 65% if you think that the chance that **your team** is going to be better than at least half of the teams in this room is 65%.

Please click "Continue" to proceed.

---

### **Your Prediction**

What is the probability that you as a team will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief B from Figure 3.1*

Please click "Continue" to proceed.

---

*Timer: 60 seconds*

### **Reasoning of your choice**

Please state the reasons why you came to this prediction.

---

### Your possible payoffs

There is a 50% chance for either one of the two following possibilities to be chosen for your payment.

1. You can be paid according to your team performance. For each point you as a team score (number of correct answers), you receive **0.50€**. Thus, if your team scores 20 points, you earn  $20 * 0.50 = 10€$ . The payment for your teammate is the same.
2. You can be paid 10€ in the Top Half contest according to how precisely you can **estimate the likelihood of your team to end up in the top half**.

So, with a 50% chance you get paid according to how well your team did in the calculation task. And with a 50% chance you get paid according to one of the beliefs you state.

Please click "Continue" to proceed.

---

Before the calculation task starts, remember:

You will be given **5 minutes** to calculate the correct sum of 25 calculations.

You cannot use a calculator but you are welcome to make use of the provided **scratch paper**. Don't forget to **enter your answers into the input field** on the screen. We cannot take into account any results on your scratch paper.

### Good luck!

Please click "Continue" to proceed.

---

*Timer: 300 seconds*

### Calculations (See Appendix A.4)

Please click "Continue" to proceed.

---

### Calculation score information

Your team score is \_\_\_ **out of 50**.

Please click "Continue" to proceed.

---

*Communicating Pairs***Chatting**

1. In a moment, you again have to predict how likely your team score is in the top half of all scores.
2. Once you have taken a decision by yourself, it will be **compared to the decision of your teammate**.
3. Directly afterwards, you and your teammate get to **chat** to each other where you see your and your teammate's prediction.
4. You and your teammate have to find a **unanimous prediction**.
5. Both of you have to **enter your unanimous prediction** directly after the chat.
6. If you do not manage to find a unanimous decision, you will not be paid for this prediction.
7. The chat round lasts **120 seconds**.

The decisions you make are **relevant for your own** as well as **for your teammate's payoffs**.

*Non-Communicating Pairs***Reasoning of your choices**

1. In a moment, you again have to predict how likely your team score is in the top half of all scores.
2. Directly afterwards, you will **get so see your teammate's prediction** (and your teammate can see yours).
3. Write down the reasons how you came up with your prediction and whether seeing your teammate's prediction affects you.
4. **If you want to reconsider your prediction** after having seen your teammate's prediction, you get the chance to do so directly after you have written down your thoughts.
5. **If you do not want to reconsider your prediction**, just enter the same prediction.
6. In total, your time to write down your reasoning is **120 seconds**.

Remember: **Only your prediction is payoff relevant for you - not the prediction of your teammate**. The more accurate your prediction is, the higher are your chances to win the 10 euro.

### Your Prediction

What is the probability that you as a team will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief C.1 from Figure 3.1*

Please click "Continue" to proceed.

---

### *Communicating Pairs*

On the next page, you have the possibility to chat to your teammate in the chat box in the middle.

You and your teammate have to find a unanimous prediction on how likely you as a team end up in the top half of all scores.

You can revise your choice **after the chat round**. If you by then **still do not have a unanimous decision**, you will **not be paid** for this stage.

### *Non-Communicating Pairs*

On the next page, you have to write down the reasons for your prediction.

You also get to see your teammate's prediction (and your teammate gets to see yours).

Write down the reasons how you came up with your prediction and whether seeing your teammate's prediction affects you.

If you want to reconsider your prediction, you can do so after you have written down your thoughts. If you do not want to change your prediction, just enter the same prediction a second time.

Please use this time to **think carefully** - and **write down your thoughts**.

Please click "Continue" to proceed.

---

*Timer: 120 seconds*

**Protocol** (see Figure A3.1)

**Chat** (see Figure A3.2)

---

*Communicating Pairs*

*(visible if Belief C.1 not unanimous)*

You will be able to submit your revised predictions on the next page.

This is your last chance to enter a unanimous decision.

If the two of you enter different predictions, you will not be paid for this decision.

Please click "Continue" to proceed.

---

*Communicating Pairs: visible if Belief C.1 not unanimous*

*Non-Communicating Pairs: always visible*

What is the probability that you as a team will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief C.2 from Figure 3.1*

Please click "Continue" to proceed.

---

*Communicating Pairs*

Your predictions after the chatting period:

You did not come to a unanimous decision. / You came to a unanimous decision.

Your prediction: ---

Your teammate's prediction: ---

*Non-Communicating Pairs*

Your predictions after the protocol period:

Your prediction: ---

Your teammate's prediction: ---

Please click "Continue" to proceed.

---

**Reevaluation**

Now that you have completed the calculation task and given us some information about how well you think you as a team did, we are going to give you some feedback on your team performance.

Specifically, we are going to give you a series of **four reports that say "top" or "bottom"**. The reports are meant to indicate whether your team performance was in the top or bottom half, but unfortunately they are not always accurate.

Each of the four reports you will get is **accurate with a 75% chance** and inaccurate with a 25% chance. In other words, if you get a report with the message "Top", there is a 75% chance that you really are in the top half but a 25% chance that you actually are in the bottom half. Vice versa, if you get a report with the message "Bottom", there is a 75% chance that you really are in the bottom half but a 25% chance that you actually are in the top half.

Please click "Continue" to proceed.

---

### Reevaluation

**You and your teammate get the same 4 reports.** Thus, if you receive a Top report, your teammate also receives a Top report; and if you receive a Bottom report, your teammate also receives a Bottom report.

**Signals are independent of each other.** Your second signal does not change just because your first signal was a Top or Bottom signal. All four signals are always accurate with a 75% chance.

**After each report we will ask you to reevaluate** the chances of your team to rank among the top half.

*Communicating Pairs:* Exactly as before, you have to find a unanimous decision together with your teammate after each report.

*Non-Communicating Pairs:* Exactly as before, you first make your own prediction. Then you get to see your teammate's prediction and are allowed to reconsider your choice.

Please click "Continue" to proceed.

---

Remember: The report you will see on the next page indicates whether you as a team are in the top half of all scores but is correct in only 75% of the cases.

Please reevaluate how likely it is that your team ends up in the top half of all participating teams.

Please click "Continue" to proceed.

---

**Your first team report** (both of you get the same report):

#### Top/Bottom

Please indicate on the next page **how likely** it is that your team score is in the top half of all scores.

Please click "Continue" to proceed.

---

**Reevaluation after the first report**  
**Your Prediction**

What is the probability that you as a team will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief D.1 from Figure 3.1*

Please click "Continue" to proceed.

---

*The pairs had to chat/write a protocol following the same procedure*

---

**Your second team report** (both of you get the same report):

**Top/Bottom**

The first report was: **Top/Bottom**

Please indicate on the next page **how likely** it is that your team score is in the top half of all scores.

Please click "Continue" to proceed.

---

*The pairs received two more reports following the same procedure*

---

## A.4 Calculations

Calculations were presented in this order. Participants could only answer one calculation at a time. Once they submitted an answer, they could not go back to a previous calculations.

1.  $66 + 95 + 85 + 50 + 82 =$
2.  $28 + 80 + 89 + 81 + 28 =$
3.  $87 + 56 + 23 + 26 + 77 =$
4.  $67 + 91 + 40 + 31 + 67 =$
5.  $13 + 10 + 63 + 45 + 79 =$
6.  $64 + 88 + 90 + 89 + 69 =$
7.  $25 + 59 + 57 + 54 + 82 =$
8.  $65 + 96 + 51 + 26 + 42 =$
9.  $33 + 57 + 87 + 51 + 98 =$
10.  $60 + 19 + 91 + 70 + 69 =$
11.  $84 + 38 + 24 + 83 + 88 =$
12.  $55 + 72 + 47 + 86 + 26 =$
13.  $81 + 38 + 55 + 34 + 47 =$
14.  $59 + 47 + 51 + 98 + 96 =$
15.  $89 + 22 + 27 + 23 + 84 =$
16.  $77 + 99 + 80 + 36 + 98 =$
17.  $80 + 21 + 46 + 43 + 43 =$
18.  $99 + 63 + 21 + 91 + 99 =$
19.  $91 + 28 + 86 + 46 + 76 =$
20.  $26 + 95 + 68 + 32 + 73 =$
21.  $96 + 77 + 90 + 86 + 50 =$
22.  $71 + 40 + 90 + 17 + 61 =$
23.  $65 + 60 + 54 + 46 + 81 =$
24.  $95 + 73 + 23 + 39 + 90 =$
25.  $89 + 35 + 94 + 65 + 19 =$



# Chapter 4

## Let's Meet in the Middle!

How gender-diversity can overcome asymmetric belief updating<sup>◇</sup>

### 4.1 Introduction

Jake and Chris just finished the final draft of their paper and are looking forward to submitting it to a top journal. Yet, they are unsure whether it is already polished enough for submission, so they decide to send the final draft to two colleagues for a last round of feedback. A week later, both their colleagues have given feedback. One did not like the paper, the other gave the thumbs up. Jake and Chris have a quick discussion about what to do and decide to submit. After all, one of their colleagues does like the paper. Is their decision objectively correct? Jake and Chris ignore that their colleagues have equal expertise and decide to acknowledge the positive feedback more than the negative. The incremental increase in confidence made them submit their paper. Such behavior is called asymmetric belief updating and leads to unjustifiably optimistic beliefs or, as commonly referred to, overconfidence (see e.g. Möbius *et al.* 2014; Eil and Rao 2011; Cacault and Grieder 2018).

Like in the case of Jake and Chris, many decisions are taken by two people, and, in case of business decisions, predominantly by two men. Men typically have overconfident individual beliefs, especially in stereotypically male domains (e.g. Beyer and Bowden 1997; Barber and

---

<sup>◇</sup>This chapter is co-authored by Paul Smeets (Maastricht University). We are especially grateful to the valuable feedback of Rob Bauer, Jaap Bos, Martin Strobel, and Antoon Pelsser. We are also grateful to the valuable comments of Inka Eberhardt, Piet Eichholtz, Peiran Jiao, Matthijs Korevaar, Joyce Mertens, Juan Palacios, Luuk Perik, Thomas Post, Paulo Rodrigues, Peter Schotman, Lidwien Sol, Sjir Uitdewilligen, and Pim van Vliet. We thank seminar participants at WHU Otto Beisheim School of Management and at Maastricht University, participants of the 2017 TIBER symposium and the 2017 Maastricht Behavioral and Experimental Economics Symposium. Paul Smeets is supported by a VENI grant from the Netherlands Organisation for Scientific Research (NWO) under grant number 016.Veni.175.019. The authors declare no conflict of interest.

Odean 2001; Niederle and Vesterlund 2007; Croson and Gneezy 2009). These beliefs serve as anchor points and are expected to cause male pairs to form an overconfident team belief (Tversky and Kahneman, 1974; Sniezek, 1992). On the other side, women do not form overconfident beliefs in stereotypically male domains (Beyer, 1990; Beyer and Bowden, 1997) which leads to the expectation of female pairs forming non-overconfident beliefs. But what happens in a gender-diverse pair? Theoretical predictions are not as clear as with single-gender pairs. The presence of a man and a woman leads to more dispersed individual beliefs in gender-diverse pairs. Different opinions are hypothesized to lead to disagreement and therefore to less overconfident joint beliefs (Tversky and Kahneman, 1974; Sniezek and Henry, 1989; Sniezek, 1992). However, men have been found to be more dominant than women (e.g. Sidanius *et al.* 1995; Pratto *et al.* 1997) which could lead to gender-diverse pairs forming overconfident beliefs.

Addressing the question of how gender-diversity affects the belief formation process in teams is highly relevant. Among European countries, several have introduced binding gender quota.<sup>1</sup> The proportion of female board seats within the largest listed European companies has consequently more than doubled between 2005 and 2017,<sup>2</sup> but the current 25.3% women among top executives show that boards are still dominated by men. Similar figures can be observed in academia where the share of women among full professors had increased by 8.5 percentage points to 23.5% between 2000 and 2014. In US companies, 12% of seats in executive boards are held by women (PwC, 2017),<sup>3</sup> while in US universities, women held 38.4% of tenured positions in 2015<sup>4</sup>.

Scientific debates about the consequences of a more gender-balanced board are controversial. Some authors find companies with more gender-diverse boards to perform significantly better (e.g. Campbell and Mínguez-Vera 2008; Julizaerma and Sori 2012; Liu *et al.* 2014; Terjesen *et al.* 2016), but others find a significantly negative effect (e.g. Adams and Ferreira 2009; Pande and Ford 2012).<sup>5</sup> The strength of these papers is that they analyze the effect of a gender-diverse board structure on the financial performance of the company using real market data. However, this approach has the downside that it is difficult to reliably identify causality. Consider the

---

<sup>1</sup> In 2003, Norway passed a bill which required a quota of 40% female executives in public and state-owned companies. The goal was reached in 2010. Countries like Spain (2007, 40%) France (2010, 40%), or the Netherlands (2010, 30%) followed (Pande and Ford, 2012). Failure to comply with gender quotas can result in voided nominations or even fines. Other countries like Germany and the UK have not yet passed legally binding quotas.

<sup>2</sup> See a report in the Guardian from 2017 (<https://www.theguardian.com/world/2017/nov/20/eu-to-push-for-40-quota-for-women-on-company-boards>) as well as the “2018 Report on equality between women and men in the EU” published by the European Commission.

<sup>3</sup> The share of 12% has stayed largely unchanged over the past decade. As the first of 50 states, California might pass a bill to mandate that corporations have at least one woman on their boards by the end of 2019.

<sup>4</sup> See <https://www.catalyst.org/knowledge/women-academia>.

<sup>5</sup> Some studies also find gender-diversity to have no effect (e.g. Marinova *et al.* 2010; Dale-Olsen *et al.* 2013; Chapple and Humphrey 2014) or a non-linear effect where a “critical mass” of about 30% women is needed before the effect of gender-diverse boards on firm performance turns from negative to positive (Joecks *et al.*, 2013).

example where a woman replaces a man in a company board. This replacement could be the outcome of a cultural change within the company raising questions of endogeneity. And even if the replacement can be referred to as an external shock, it has to be shown how the actions and decisions taken by the new female board member translate into stock price movement and how this differs compared to the male predecessor.

We run a laboratory experiment to study the causal influence of gender on belief updating in pairs. We restrict our research to pairs as there are multiple examples where decisions are made by two people. In the academic world, many papers are written by two co-authors; in the industry, many company decisions are taken by CEO and CFO. And even outside the professional world, couples take daily household decisions. On top, decision making in pairs allows us to better control for group dynamics which get significantly more complex as soon as we introduce more people to the group.

In this paper, we focus on the effect of gender diversity on overconfidence (asymmetric belief updating) for three reasons. First, the formation of beliefs resembles the very core of a decision-making process. Take the example of Jake and Chris. Incoming information will either support the submission or suggest to not submit. With each piece of information, the two have to form a belief about how likely a successful submission is. On the basis of this belief updating process, they eventually will decide. Second, overconfidence is one of the best documented differences between men and women and has also been found among professional executives (Huang and Kisgen, 2013; Mishra and Metilda, 2015). Boards could thus benefit from an increasing share of less overconfident women. Third, while some papers find overconfidence to have positive consequences,<sup>6</sup> the majority of papers find costly effects (e.g. Camerer and Lovo 1999; Barber and Odean 2001; Biais *et al.* 2005; Malmendier and Tate 2005, 2008; Möbius *et al.* 2007, 2014; Kim *et al.* 2016; Bordalo *et al.* forthcoming). This underlines the necessity to study how to reduce overconfidence when it is harmful. We therefore deliberately chose an experimental design where overconfidence is financially costly and where subjects are aware of this.

To mimic the male-dominated environment, we use a calculation task as a measure for performance. Men are perceived to be better calculators than women (e.g. Beyer 1990; Eccles 1993; Frome and Eccles 1998; Cvencek *et al.* 2011; Reuben *et al.* 2014). This adaptation is highly relevant to our study. Gender differences in beliefs about the own performance are largest in stereotypically male domains (Bordalo *et al.*, forthcoming). Importantly, however, there is no actual performance gap between men and women in simple calculation tasks (see e.g. Hyde *et al.* 1990; Catsambis 1994; Guiso *et al.* 2008; Hyde *et al.* 2008). Our findings in this study

---

<sup>6</sup> Bernardo and Welch (2001) argue that overconfident people convey private information to the public, Hirshleifer and Luo (2001) find overconfident traders to do better than purely rational traders at exploiting mispricing, Johnson and Fowler (2011) show that overconfident populations are evolutionarily stable, and Gervais *et al.* (2011) show that moderate levels of overconfidence make it cheaper for firms to motivate the worker to pursue valuable risky projects.

as well as the findings of Ruof *et al.* (2018) confirm these findings. Thus, the belief formation process is only affected by gender, not by an actual performance gap.

Experimental subjects are randomly assigned to purely male, gender-diverse, or purely female pairs. Upon being matched, subjects learn about the gender of their teammate. Then, participants individually complete a timed calculation task. The team score is the sum of correctly solved calculations between the two. Pairs get to know their score after the calculation task. Pairs then chat and have to find a unanimous belief. Specifically, a pair has to indicate how likely it is that their team score is in the top half of all team scores. Successively, noisy feedback on the performance is given in the form of a binary signal which is either positive (“Top half”) or negative (“Bottom half”). Signals are accurate with a 75% chance. After each of the four signals, pairs can update their beliefs. Given symmetric signals, Bayesian updating requires pairs to respond equally strongly to positive and negative signals. If pairs show a stronger response to positive and a weaker response to negative signals, they update asymmetrically<sup>7</sup>. In the long-run, asymmetric updating leads to unjustifiably optimistic beliefs.

Strikingly, we find gender-diverse pairs to update strongly asymmetrically. They respond 42% stronger to positive than to negative feedback. Such asymmetry has a considerable economic effect. A gender-diverse pair with a prior belief of 50% would end up with a posterior belief of 59.7% after having received two positive and two negative reports. This belief is 20% above the Bayesian benchmark. In line with theoretical predictions, we find male pairs to update highly asymmetrically and female pairs to update almost perfectly symmetrically. The asymmetric belief updating among gender-diverse pairs is the result of dominant behavior of men and women. But there is a cure. When both teammates decide to meet in the middle of their beliefs in order to form a joint belief, gender-diverse pairs update symmetrically. “Meet in the middle!” is a simple external rule for gender-diverse pairs to debias their beliefs. Remarkably, this rule does not apply to male pairs: they are overconfident either way.

With our work, we contribute to the literature in two important ways. First, we contribute to the literature on gender diversity while being able to address causality. Many field studies have tackled the gender effect on team decisions (e.g. Liu *et al.* 2014; Terjesen *et al.* 2016; Dale-Olsen *et al.* 2013; Chapple and Humphrey 2014; Adams and Ferreira 2009; Pande and Ford 2012; Joecks *et al.* 2013). These studies have the clear advantage that they can address the consequences of gender diversity using actual market data but face the challenge to cleanly identify what causes these effects. It mostly is difficult to argue that the occupation of an open position is an exogenous shock. Further, since most decisions of an executive are unobservable, their impact on to the financial performance of the company is ambiguous. Our lab experiment provides a critical contribution to cleanly identify the effects of gender on group decisions.

---

<sup>7</sup> For simplicity, when we write asymmetric updating, we refer to pairs responding stronger to positive than to negative feedback.

Second, we identify the behavioral traits which lead to overconfident updating among pairs. While previous studies have either studied whether there exists a gender difference in overconfidence (e.g. Möbius *et al.* 2007; Buser *et al.* 2016) or used this difference to explain a certain behavior (e.g. Estes and Hosseini 1988; Barber and Odean 2001; Niederle and Vesterlund 2007), we go one step further and analyze which behavioral traits cause overconfidence and how it can be overcome. This is an important contribution to the overconfidence literature as it helps to understand where overconfidence (in pairs) comes from and how it can be met.

## 4.2 Hypothesis development

Two major forces have been identified that can affect the formation of group beliefs, when communication is allowed: disagreement and dominance.

(1) The *force of disagreement*: Members of a group have different individual beliefs which serve as anchor points in the process of forming a team belief (Eriksen and Hake, 1957; Tversky and Kahneman, 1974; Sniezek and Henry, 1989; Kahneman, 1992; Sniezek, 1992; Whyte *et al.*, 1997). These individual beliefs compete with each other and lead to disagreement among teammates which in turn leads to a thorough discussion of available information. Sniezek and Henry (1989) hypothesize that competing individual beliefs thus increase the likelihood to detect asymmetric individual belief updates (the ‘competing anchors hypothesis’). This process, however, can be impeded if individual beliefs lie closely together or suffer from the same bias (Sniezek and Henry, 1989). In this case, there is less disagreement between the individual anchor beliefs and consequently a less thorough discussion. Hence, disagreement leads to more accurate (less asymmetric) group beliefs if individual beliefs are sufficiently diverse.

(2) The *force of dominance*: Next to disagreement, interaction in groups is affected by dominant behavior which is defined as the power to impose one’s will upon others (see e.g. Halebian and Finkelstein 1993; Brown and Sarma 2007). Dominant individuals are wrongly believed to be more competent (Kenny and LaVoie, 1985) and consistently gain a high level of influence in groups (Anderson and Kilduff, 2009). Two consequences of dominant behavior are especially relevant to our study on belief updating. First, dominance has been found to be positively related to overconfidence. Brown and Sarma (2007) study the impact of overconfidence and dominance on the acquisitiveness of CEOs and find a positive correlation between dominance and overconfidence. Burks *et al.* (2013) study the behavior of 1,000 truck drivers. They find dominant truck drivers to make more confident judgments about their own relative IQ score even when controlling for their actual ability. Ruof *et al.* (2018) find dominant behavior to be a slight cause for asymmetric belief updating among communicating pairs. Second, differences of dominant behavior between men and women are well documented. The majority of studies find men to score significantly higher on social dominance orientation (SDO) than women (e.g. Sidanius *et al.* 1995; Pratto *et al.* 1997; Sidanius *et al.* 2000, 2006; Adams and Funk 2012).

Scoring high on SDO resembles the preference for a hierarchy within a social system and for dominating individual group members. Hence, the force of dominance suggests that the presence of a dominant person in pairs leads to asymmetric belief updating.

In this paper, we address the question of how these two forces affect decision-making in pairs with a different gender composition.

**Female pairs.** Findings about women indicate that they are significantly less overconfident than men. In masculine-gender-typed tasks, women even tend to underestimate their performance (Beyer, 1990; Beyer and Bowden, 1997). Ring *et al.* (2016) do not find women to think they outperform their own kind while Bordalo *et al.* (forthcoming) report women to be less likely to believe their own ability exceeds that of a partner. As individual beliefs are not likely to be biased, the force of disagreement is unlikely to cause female pairs to update asymmetrically even though beliefs tend to not be very dispersed. The force of dominance should further play a minor role as women are significantly less likely to be dominant. Thus, female pairs are expected to show relatively symmetric belief updates.

**Male pairs.** The evidence of individual men being overconfident, especially in stereotypically male domains, is strong (Estes and Hosseini, 1988; Lundeberg *et al.*, 1994; Beyer, 1990, 1998; Beyer and Bowden, 1997; Pulford and Colman, 1997; Barber and Odean, 2001; Johnson *et al.*, 2006; Niederle and Vesterlund, 2007; Croson and Gneezy, 2009; Coffman, 2014; Bordalo *et al.*, forthcoming). Barber and Odean (2001) find men to be highly overconfident in investment decisions. Beyer (1990) finds the same for masculine-gender-typed tasks like a quiz on political and sports figures and a math-based task. Lundeberg *et al.* (1994) test item-specific confidence levels and find men to be overconfident especially with easy mathematical and statistical problems. In contrast to women, Ring *et al.* (2016) find that men even expect to perform significantly better than their own kind, other men. And Bordalo *et al.* (forthcoming) find that gender stereotypes are an important source for belief distortions which especially arise in stereotypically male-typed tasks such as mathematics. These findings even extend to professionals. Huang and Kisgen (2013) find male CEOs to be more overconfident than female CEOs where male CEOs show higher acquisitiveness and later exercising of stock options. Mishra and Metilda (2015) examine overconfidence among mutual fund investors and find men to be more overconfident than women. Hence, these findings suggest that anchor beliefs of male pairs are little dispersed and on average suffer from the same overconfidence bias which impedes the positive effect of the force of disagreement. The force of dominance would further deteriorate belief accuracy. Thus, we expect male pairs to update asymmetrically.

**Gender-diverse pairs.** The question we focus on in this paper is whether gender-diverse pairs suffer from asymmetric belief updates. According to the force of disagreement, pairing an overconfident man and a less overconfident or even underconfident woman creates a more dispersed belief space within gender-diverse pairs which is hypothesized to produce less asymmetric or even symmetric belief updates. On the other hand, gender-diverse pairs could suffer

from the force of dominance as men have a significantly higher dominance score than women leading to more asymmetric belief updates. Hence, it is not clear how gender-diverse pairs update. It could be that the force of disagreement weighs stronger than the force of dominance, but it could also be the other way around. Consequently, the influence of the two forces has to be studied empirically.

## 4.3 The experimental design

We randomly formed three distinct types of pairs: male, gender-diverse, and female pairs. Conditional on gender, the assignment of participants is random. A graphical illustration of the experimental procedure is provided in Figure 4.1. In the beginning of the experiment, all participants were told that they are about to do a timed calculation task. Subjects would receive a score which equals the number of correctly solved calculations. Subjects then should estimate how likely they are to end up in the 50 percent of all individual scores in their session (see belief A, Figure 4.1). Only then, we told subjects that they have been randomly paired with another subject of their session. The sum of the two individual scores would form their team score and individual scores would not be reported. Information was given about the teammate's current university, age bracket, and gender. Provided were several local universities and four age brackets (below 18; 18–28; 29–39; 40–50). The provision of several local universities as well as age brackets should make it less likely that participants considered gender to be of particular importance. We exclude pairs from our main analysis where the university and the age bracket are not the same between the two teammates (see section 4.3.8) to assure that the only source of variation stems from the different gender distributions. We then asked subjects to estimate how likely their team score is to end up in the top half of all teams (belief B). The two beliefs were elicited before the calculation task and for each subject individually. Participants did not know that they would have to interact with their teammate until after the calculation task was over. Thus, beliefs A and B are not influenced by the teammate. The calculation task itself will be explained in section 4.3.4.

### 4.3.1 The elicitation of beliefs

Participants were asked to state their true beliefs. In the beginning of the experiment, the experimenter instructed all participants that they would have to state beliefs throughout the experiment and that stating true beliefs would maximize their expected payoffs. The experimenter gave participants the paper by Karni (2009) whose payoff mechanism we use and who shows that stating true beliefs is incentive-compatible.<sup>8</sup> As participants were unlikely to be able

---

<sup>8</sup>The stated belief is assumed to represent the mean of someone's belief distribution. The elicitation method we employed follows the method developed by Karni (2009) and is strictly incentive-compatible. This method

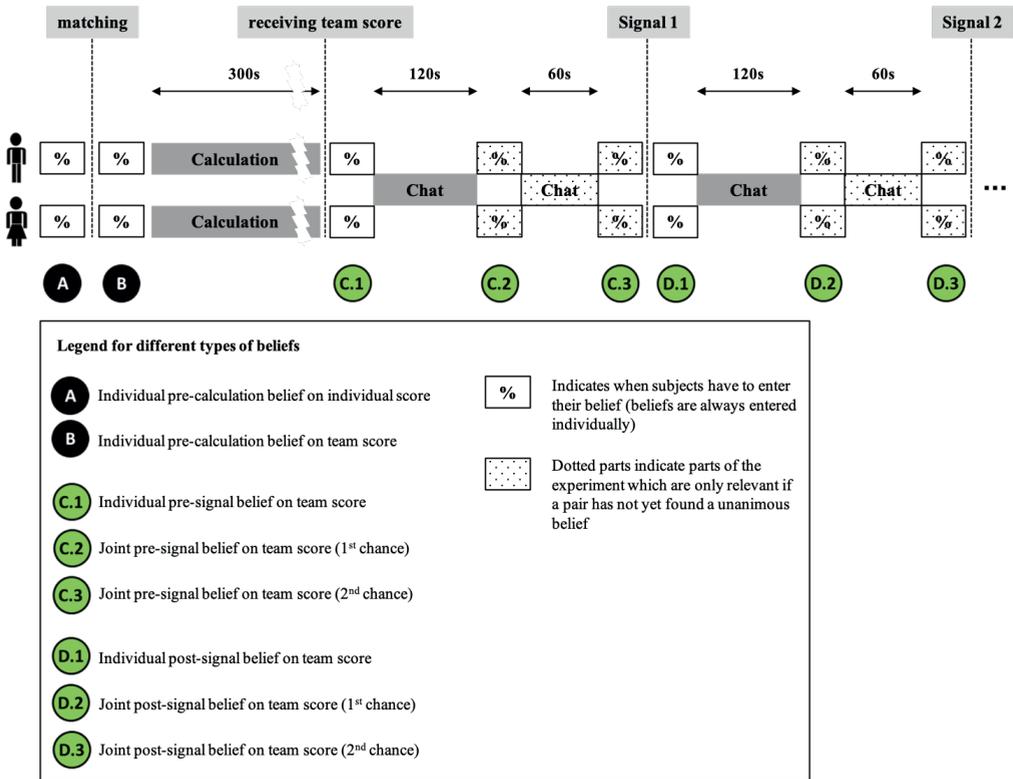


FIGURE 4.1 – Order of events in the experiment.

This figure presents the experimental procedure which is the same for all pairs. All subjects were told that they are about to do a timed calculation task. Subjects would receive a score which equals the number of correctly solved calculations. The experiment introduced the calculation task as an individual task and asked subjects to estimate how likely they are to end up in the top 50 percent of all individual scores in their session (belief A). Then, subjects were told that they had been randomly matched to another subject of their session and that the sum of the two individual scores will form their team score. Information was given about the teammate’s current university, age, and gender. Subjects were then asked to estimate how likely their team score is to end up in the top half of all teams (belief B). Then subjects individually completed the calculation task (300 seconds) and received their team score. Subjects first had to state their individual belief (belief C.1) that the team score would end up in the top half of all team scores in the session. Subjects knew that after entering their individual belief, they would get to chat to their teammate to find a unanimous belief. Once each teammate had submitted the individual prior, a chat window opened and the two could see the prior of their teammate. If the priors did not match, subjects had 120 seconds to find a unanimous belief. After the chat closed, subjects individually entered their revised belief (belief C.2). In case of a unanimous decision, they were done. If not, the chat box reappeared. After the second chatting period of 60 seconds, they got a final chance to enter a unanimous prior (belief C.3). Then, subjects received a sequence of four informative, but noisy signals on their team performance. After each signal, subjects had to find a unanimous belief following the same procedure as just described (beliefs D.1 to D.3, etc.).

to read and understand the paper during the experiment, we stressed that the abstract of the paper already stressed the main finding within two sentences, where especially one sentence is of importance: "The game induced by the mechanism has a dominant strategy equilibrium in which the players reveal their subjective probabilities." This procedure allowed us to make it as explicit as possible to participants that stating their true beliefs would yield them the highest expected payoffs. Of all participants, 12 participants requested to see the paper. The experimenter also gave a quick example which served to stress the fact that subjects should only state their true belief and that anything but their true belief would diminish their expected payoff.

In addition, we made sure that everybody understood the procedure of the experiment by thoroughly explaining the procedure before as well as during the experiment including several example questions. Subjects had to complete a similar non-incentivized updating game before the calculation task to get familiar with the theme.

### 4.3.2 The formation of joint beliefs

For simplicity, let us consider the following example. Jake and Lindsey have been matched. Jake knows he is matched to a girl, Lindsey knows she is matched to a guy. Assume they individually complete the calculation task (300 seconds) and receive the information that they correctly solved a total of 20 calculations. Individual scores are not provided. Now, Lindsey is asked to state her belief (belief C.1) that their team score is in the top half of all scores in her session. She is also informed that after she has entered her belief, she will get the chance to chat with her teammate to find a unanimous belief. After receiving all this information, she enters her individual belief. Jake follows the same procedure. Once each of them has submitted an individual prior (belief C.1), a chat window opens for 120 seconds. The two can see the prior of their teammate on the screen (see Figure A4.1). In case their beliefs do not match, Jake and Lindsey have up to two chatting rounds to find a unanimous belief. If they find a unanimous belief after the first chatting period (belief C.2), they are done. If the beliefs still do not match, the chat box reappears for 60 seconds after which they get a final chance to enter a unanimous belief (belief C.3). If the beliefs are still not the same, there is no possibility for revision. Both learn about the procedure of how to find a unanimous belief before they enter belief C.1. They are further told at the same time that disagreement after three rounds results in losing the entitlement to be paid in case this round is chosen for payment.

Suppose Jake estimates the likelihood to end up in the top half given their team score to be 60%. Lindsey on the other side believes that their score has a mere 40% chance to make it to the top half. After independently submitting these beliefs (C.1), they get to chat with each other to find a unanimous decision but cannot come to an agreement. On the next screen, Jake

---

does not allow for hedging behavior, which can happen with the quadratic scoring rule where under certain circumstances expected payoffs for subjects are marginally decreasing in solving more calculations leading to perverse incentives (for proof, see Möbius *et al.* 2014).

stays at 60% while Lindsey slightly increases her assessment to 50% (C.2). Hence, the chat box reappears while both know that this is the last chance to agree on a unanimous belief. Both agree on 55% and individually enter the belief on the next screen (C.3) enabling them to be eligible for payment in case this belief is chosen for payment. The procedure is the same for all gender subgroups and all receive equal information.

### 4.3.3 The updating of beliefs

After pairs have formed their joint prior beliefs, they are informed that they will receive a sequence of four conditionally independent, noisy, but informative signals. Teammates get to see the same signal. Signals provide an indication whether the team score is in the top or bottom half and have an accuracy of 75 percent. In other words, if the team score is in the top half, the signal says “Top” with a probability of 75 percent and “Bottom” with a probability of 25 percent. Equivalently, if the score is in the bottom half, the signal says “Bottom” with a probability of 75 percent and “Top” with a probability of 25 percent (symmetric signals). After each signal, pairs have to find a unanimous belief following the procedure described in section 4.3.2 (for the procedure after signal 1, see beliefs D.1 to D.3, Figure 4.1).

### 4.3.4 The calculation task

At the start of the experiment, participants were told that they would have to complete a timed calculation task. This task is a self-serving or ego-relevant task which requires ability and effort to be solved. This feature is important as self-serving tasks have been shown to cause asymmetric updating while self-irrelevant do not (see e.g. Eil and Rao 2011; Ertac 2011). This task had to be done by everybody individually to assure that we would not get any synergy effects. The task consisted of 25 calculations, for which participants had a maximum of five minutes (for the exact calculations, see Appendix A.4). Each calculation showed five randomly drawn two-digit numbers which had to be added (e.g.  $93 + 53 + 32 + 69 + 19 = \square$ ). Only one calculation was visible at a time and could not be revisited once an answer had been submitted. Participants were not allowed to skip calculations. The sequence of 25 calculations was the same for everyone.

We deliberately chose a calculation task for three reasons. First, a calculation task is considered to be a male-typed task in which men are perceived to be better than women (e.g. Beyer 1990; Eccles 1993; Frome and Eccles 1998; Cvencek *et al.* 2011; Reuben *et al.* 2014). Second, there is no actual performance gap between men and women in simple calculation tasks (see e.g. Hyde *et al.* 1990; Catsambis 1994; Guiso *et al.* 2008; Hyde *et al.* 2008) and neither do we find a performance gap in this study. Third, we picked a test which is ego-relevant but where English language skills do not matter. Most papers on belief updating use an IQ test which combines logical and English language skills (e.g. Möbius *et al.* 2007, 2014; Eil and Rao 2011;

Buser *et al.* 2016; Cacault and Grieder 2018) or mathematical tasks and English language skills (Coutts, 2018). As subjects in our sample mostly are not English natives, we decided to go with a purely mathematical task.

### 4.3.5 The matching

In the beginning of the experiment, we asked participants to provide information on their age, their gender, and the university they are currently enrolled in. The age was elicited using intervals (below 18, 18-28, 29-39, 40-50). The university was elicited by providing names of the three closest universities as well as the option ‘other’. The provision of several local universities as well as age brackets should make it less likely that participants considered gender to be of particular importance. If the two subjects of a pair were from a different university or within a different age-bracket, we excluded the pair from the analysis (see section 4.3.8) to assure that the only source of variation stems from the difference in gender distributions. Once participants were informed that they had been matched with another participant, they learned about the university, the age bracket, and the gender of their teammate. This approach allowed us to hide the fact that we wanted subjects to know the gender of their teammate to avoid demand effects.

### 4.3.6 The payoff

Subjects were explicitly told that throughout the experiment they would have to complete multiple payoff-relevant tasks of which only one would be chosen at random and implemented as their payoff-determining decision. With a likelihood of 50%, the calculation task was chosen to be payoff-relevant. For each correctly solved calculation as a team, subjects would earn 0.50 euro in case the calculation task was selected for payment. The remaining 50% were equally spread across all belief elicitation tasks (and two additional small experiments).

### 4.3.7 Methodology

With our methodology to elicit overconfidence through asymmetric updating behavior, we adhere to the critique of Benoît and Dubra (2011). They show that the better-than-average effect does not unambiguously show overconfident behavior even if the majority of people beliefs to be better than average as these beliefs could be rationalized in a Bayesian way given a certain distribution of priors. To infer unjustifiably optimistic beliefs according to Bayes’ Law, subjects would have to value positive feedback more than negative feedback. This behavior is called asymmetric updating as new information is valued asymmetrically and leads to overconfident beliefs in the long-run. Several papers have already looked at asymmetric updating (e.g. Möbius *et al.* 2007, 2014; Eil and Rao 2011; Cacault and Grieder 2018; Coutts 2018).

For the empirical analysis, we follow the approach by Möbius *et al.* (2014). Suppose a subject finishes the calculation task and has a prior belief  $\mu^t = P(A = H)$  that his true ability  $A$  is high  $H$  and assume this subject observes a high signal  $h$ . A subject receives a high signal  $x = h$  with probability  $p$  if his true ability is high and with probability  $q$  if his true ability is low. We employ symmetric signals with  $p = 0.75$  and  $q = 0.25$  which means that the signal accuracy is 75% irrespective of the true ability. People of high ability receive a high signal with a 75% chance and people of low ability receive a low signal with a 75% chance. The subject then derives his posterior belief  $\mu^{t+1}$  following Bayes' Law:

$$\mu_{t+1} = \frac{P(x = h|A = H) * \mu^t}{P(x = h|A = H) * \mu^t + P(x = h|A = L) * (1 - \mu^t)} \quad (4.1)$$

This translates into

$$\text{logit}(\mu_t) = \text{logit}(\mu_0) + X_h^t * \ln\left(\frac{p}{q}\right) + X_l^t * \ln\left(\frac{1-p}{1-q}\right) \quad (4.2)$$

where  $\text{logit}(z) = \ln\left(\frac{z}{1-z}\right)$  and  $X$  the number of high, respectively low signals received at time  $t$ .<sup>9</sup> This allows us to empirically test whether subjects update optimally by using the following specification:

$$\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_t = h) * \lambda_h + \beta_l * I(x_t = l) * \lambda_l + \epsilon_t \quad (4.3)$$

where  $\lambda_h = \ln\left(\frac{p}{q}\right)$  and  $\lambda_l = \ln\left(\frac{1-p}{1-q}\right)$  denote the loglikelihood ratios of positive and negative signals. For an optimal Bayesian updater, the coefficients equal one, i.e.  $\delta = \beta_h = \beta_l = 1$ .

### 4.3.8 Restriction of the sample

To account for confusion, misunderstanding of the experiment, or factors potentially indicating that subjects were otherwise not committed to attentively follow the experiment, we exclude all pairs which never update (10 pairs, 6.8%) or update in the wrong direction at least once (21 pairs, 14.3%). This is in line with the procedure of previous papers (Möbius *et al.*, 2007, 2014; Coutts, 2018). Further, we exclude pairs that do not have the same matching characteristics (age interval and university enrolled in) to assure that the only variation between pairs stems from the difference in the gender distribution. All in all, we exclude 40 pairs or 27.2% of the entire sample, which leaves us with 107 pairs (41 male, 33 gender-diverse, and 33 female pairs).

We further exclude observations from the analysis when the prior belief is either below 10% or above 90%. The rationale behind this exclusion is that minor deviations from optimal Bayesian updates have a huge impact on the updating function if priors lie within the extreme

<sup>9</sup> For the derivation, see Cacault and Grieder (2018).

part of the distribution but have almost no effect if priors are close to 50%. Consider Jake and Lindsey, who both deviate from optimal Bayesian updating by just one percentage point. Jake has a prior of  $\mu^0 = 50\%$  and receives a positive signal. According to Bayes' Law, he should update his belief to  $\mu^1 = \frac{0.75*0.5}{0.75*0.5+0.25*(1-0.5)} = 75\%$ , that is by 25 percentage points. Given his one percentage point deviation, he ends up with a posterior belief of  $\hat{\mu}^1 = 76\%$ . Jake therefore overreacts by  $\frac{26\%-25\%}{25\%} = 4\%$  compared to the optimal Bayesian updater. Now consider Lindsey with a prior belief of  $\mu^0 = 95\%$  who also receives a positive signal. Her optimal Bayesian posterior is  $\mu^1 = \frac{0.75*0.95}{0.75*0.95+0.25*(1-0.95)} \approx 98.3\%$ , that is she has to update by 3.3 percentage points. Lindsey also deviates by one percentage point which results in  $\hat{\mu}^1 = 99.3\%$  or a relative deviation of  $\frac{4.3\%-3.3\%}{3.3\%} \approx 30.3\%$ . In other words, while both overreact to a positive signal by one percentage point, we only observe a slight relative overestimation with Jake but a significant relative overestimation with Lindsey.

This algebraical effect has a strong impact on our regression analysis. For visualization, we ran simulations mimicking 1,000 individual belief updating processes over the course of four signals analog to the design of our experiment. Column (1) of Table A4.2 shows the pooled regression results when all participants update optimally Bayesian. All coefficients equal one as follows directly from equation 4.2.<sup>10</sup> In column (2), we simulate that individuals with a prior belief between 50% and 60% respond one percentage point stronger to positive feedback than an optimal Bayesian; the response to negative feedback remains optimally Bayesian. Subjects with priors below 50% and above 60% update optimally Bayesian for all signals. Take Jake from the example above. Given his prior belief of  $\mu^0 = 50\%$ , he falls within the relevant belief decile and updates to  $\mu^1 = 76\%$  upon receiving a positive signal (instead of the optimal 75%). One round later, he further updates  $\mu^1 = 76\%$  after receiving another signal. Since this belief does not lie within the decile from 50% to 60% anymore, he now updates optimally Bayesian. In other words, we create an environment which allows us to understand the impact that a one percentage point deviation from optimal Bayesian updating within a certain part of the belief distribution has on the *entire* regression. In columns (3) and (4), we change the asymmetric updating decile to the 9th, respectively 10th decile.

The results show that a one percentage point deviation in response to a positive signal has basically no effect if observed only in the sixth decile with a coefficient of  $\beta_H = 1.0041 \approx 1$ , while there is a small effect observable if subjects deviate in the ninth decile ( $\beta_H = 1.0213$ ). If belief updates are asymmetric by one percentage point in the tenth decile, however, we observe a huge effect. Even though, subjects by construction deviate in the tenth decile in absolute terms as much as in the 6th or 9th decile, the coefficient of  $\beta_H = 1.1088$  suggests that subjects now update strongly asymmetrically. This effect can equally be observed in the first decile and for overreacting to negative signals.

<sup>10</sup> Coefficients are significantly different from one as we include i.i.d. noise.

It is unrealistic to expect such updating precision of subjects in our experiment, especially since we limit their updates to integers. We therefore exclude observations where the prior belief lies within the first or the tenth decile from our analyses to avoid that extreme beliefs drive our results. Importantly, we only exclude those observations where the pre-signal belief of a certain round lies within the first or tenth decile. If, after an update, the belief lies within the range of 10% to 90% again, the next update is not excluded. Example: assume, a pair starts with a prior of 95%, receives a negative signal, and updates downwards to 80%. The second signal is positive, and the pair updates their belief back to 95%. Then we exclude their first update from our regression as the pre-signal prior (95%) lies within the tenth decile. Yet, we keep their second update as their pre-signal prior (80%) now does not lie within the tenth decile any longer. This leads to a total exclusion of 5.1% or 44 observations (30 male, 12 gender-diverse, and 2 female). We show in section 4.6.2 that the inclusion of all observations does not change our main result but does have an effect on the results regarding the effects of disagreement and dominance.

## 4.4 The data

### 4.4.1 The participants

The experiment was programmed using ztree (Fischbacher, 2007) and was run in May and November 2017 and in March 2018 at Maastricht University. We invited a total of 2,632 students of which 294 (9.0%) participated in the experiment. Participants were on average 21.3 years old, 49.3% were female. After excluding those who never updated and/or updated in the wrong direction at least once, 214 subjects (72.8%) remain for our analysis.

Those who are excluded differ only to a small extent from those who remain in the sample in terms of their gender (excl: 57.5% female vs incl: 46.3% female; p-value = 0.0868). Importantly, they do not differ regarding their self-reported math skills. We asked subjects to answer the statement *"My math skills are higher than those of \_\_\_ of my fellow students."* on a 11-point Likert scale from 0% to 100% in steps of ten percentage points. Those excluded reported to be better than 57.5%, those who remained to be better than 62.5% (p-value = 0.1079).

### 4.4.2 Updating behavior and calculation scores

On average, pairs update 3.3 of four times. Differences between the various gender distributions are non-significant (p-values > 0.5305). Thereby, all pairs find a unanimous belief in all five rounds.<sup>11</sup> The correlation between actual posterior beliefs and optimal posterior beliefs calculated using Bayes' Law is  $\rho = 0.77$  (p-value < 0.0001) which lies in-between the correlation

---

<sup>11</sup>In total, two of 147 pairs do not find a unanimous decision in one of five rounds. However, those are among the pairs which we already exclude from the sample.

coefficients found by Coutts (2018,  $\rho = 0.89$ ), Möbius *et al.* (2014,  $\rho = 0.57$ ), and Ruof *et al.* (2018,  $\rho = 0.62$ ). This shows that the reaction to signals in the present paper is similar to previous findings.

The 214 subjects solved 8.8 calculations in 5 minutes (see Table A4.3). There was no significant difference in scores between men and women (p-value = 0.8861) which is in line with findings in the literature (Hyde *et al.*, 1990; Guiso *et al.*, 2008). The distribution of calculation scores is slightly positively skewed. When we exclude scores of 16 and higher (3.4% of all scores), the distribution is approximately normal (Shapiro-Wilk test, p-value = 0.2546). Pairs on average score 17.6 on the calculation task with a median score of 17. Gender-diverse pairs solve slightly more than average, female slightly less. Yet, differences are not significant (p-values > 0.1439).

## 4.5 Results

### 4.5.1 Men and women wrongly belief in the gender-performance gap

As explained in section 4.3, subjects first were told that they would have to do the calculation task by themselves. They were asked to state how likely their individual score would be among the top 50% of all individual scores (see Figure 4.1, belief A). Then, subjects learned that they had been matched and that they would receive a team score instead of an individual score. Now they were asked how likely their team score would be among the top 50% of all team scores (belief B). Both beliefs were elicited before the calculation task and without interaction among teammates.

The average belief A for women is 52.1% which does not differ from the unconditional prior of 50% (p-value = 0.2283; see Figure 4.2(a)). Male subjects, on the other side, are more confident about their individual skills (Figure 4.2(b)). The average male has a belief A of 59.5% which is 9.5 percentage points or 14.2% higher than the average female belief (p-value = 0.0037). Upon learning that a woman has been matched to another woman, her belief stays the same (belief B = 53.4%; difference to belief A: p-value = 0.8608). Once a man is matched to another man, his belief slightly decreases to 57.7% (belief B; p-value = 0.1107) but is still higher than belief B of women in female pairs (p-value = 0.0643). These findings are in line with the literature showing that men are more confident about their ability than women (e.g. Beyer 1990; Barber and Odean 2001; Healy and Pate 2007; Bordalo *et al.* forthcoming).

But what about gender-diverse pairs? A woman who learns that she has been matched to a man significantly increases her belief by 5.1 percentage points or 10.2% (belief A: 52.1% vs belief B: 57.2%; p-value = 0.0012). Similarly, a man significantly decreases his belief by 6.4 percentage points or 10.8% (belief A: 59.5% vs belief B: 53.1%; p-value = 0.0604). Both sexes are clearly

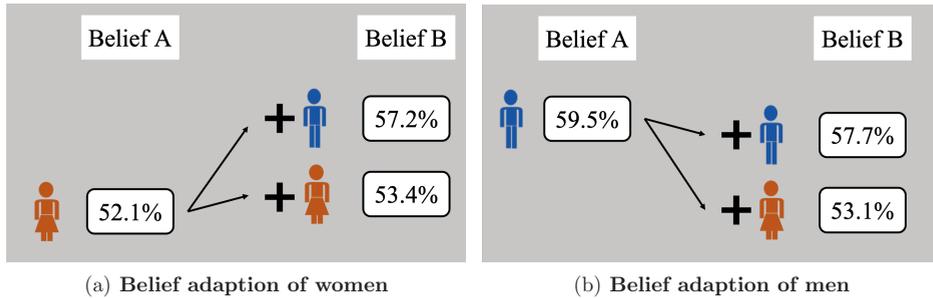


FIGURE 4.2 – **Belief adaption after being paired.**

This graph shows how women (left) and men (right) reacted to being matched. *Belief A*: Subjects were asked to state how likely their individual score would be among the top 50% of all individual scores (see Figure 4.1). *Belief B*: After subjects learned that they had been matched to a man or a woman, they were asked how likely their team score would be among the top 50% of all team scores. Both beliefs were elicited before the calculation task and without interaction among teammates.

affected by being matched to the opposite sex. This finding shows that the manipulation to put subjects into an environment which is believed to be dominated by men worked.

The question is whether these belief adaptations in gender-diverse pairs lead to more accurate beliefs. A first indication that this is not the case can be deduced from the fact that men and women wrongly believe in the existence of a gender performance gap. A second indication is provided by a look at the correlation between belief B and the team score (TS). Only female pairs show a significantly positive correlation of  $\rho(B, TS)|_{FF} = 0.2210$  (difference from zero: p-value = 0.0735) meaning that a higher level of confidence about the team score coincides with a higher team score. Male pairs with  $\rho(B, TS)|_{MM} = 0.1571$  (p-value = 0.1588) and gender-diverse pairs with  $\rho(B, TS)|_{MF} = 0.0681$  (p-value = 0.5871) however rather struggle with assessing the likelihood to be in the top half.

**Result 1:** In a stereotypically male-dominated environment, women display increasing confidence when matched to a man, while men show decreasing confidence when matched to a woman.

#### 4.5.2 Gender-diverse pairs update asymmetrically

The findings in the previous section provide indicative evidence that gender-diverse pairs update poorly. In this section, we study the causal relation between the gender distribution of pairs and their updating behavior. We therefore run pooled OLS regressions of equation 4.3 for each gender subgroup separately. According to our expectation that female pairs update symmetrically, the difference between their response to positive signals,  $\beta_H$ , and negative signals,  $\beta_L$ , should be zero. The left bar of Figure 4.3 shows that female pairs update symmetrically with a

difference of  $\beta_H - \beta_L = -1.3$  percentage points (difference from zero: p-value = 0.8361; for the complete regression results, see Table 4.1). This finding supports expectation. To understand the result, imagine a slightly adjusted version of the opening example with two professors, Lindsey and Chelsea, who have an R&R at a top journal. They have addressed all comments from the referees but are only 50% sure that their changes suffice. To get more information, they have sent out their current version of the paper to four colleagues who do not communicate with each other and have equal expertise in this field (this mimics our four symmetric signals). Two of the colleagues argue that the current version is insufficient, the other two think it is good to go. Lindsey and Chelsea would update their prior belief after reading the four answers and end up with the correct posterior belief of still only being 50% certain to re-submit the current version.<sup>12</sup>

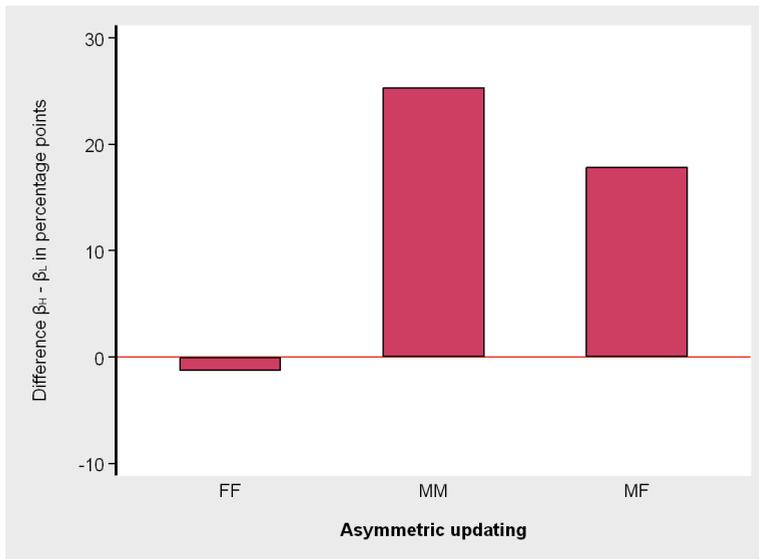


FIGURE 4.3 – Asymmetry and gender composition.

This graph plots the difference between coefficients  $\beta_H$  and  $\beta_L$  from Table 4.1. Separate regressions were run for each gender subgroup. Updating according to Bayes' Law requires all bars to be equal to zero. A positive difference  $\beta_H > \beta_L$  implies asymmetric belief updating where positive feedback is valued stronger than negative feedback. Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered.

The middle bar in Figure 4.3 shows that male pairs update beliefs in a highly asymmetric manner with  $\beta_H - \beta_L = 25.4$  percentage points (p-value = 0.0015) which is in line with our expectation for male pairs. The difference in asymmetry between male and female pairs is

<sup>12</sup>We assume that positive and negative feedback contains an equal amount of information.

highly significant (p-value = 0.0030)<sup>13</sup>. Male pairs respond 61% stronger to positive than to negative feedback ( $\beta_H/\beta_L = 0.6714/0.4177 \approx 1.61$ ; see column (2) of Table 4.1). The economic effect of such asymmetric updates is huge. Consider Jake and Chris to be in the same situation as Lindsey and Chelsea. A sequence of two positive and two negative feedback reports from colleagues would leave the two male professors with a posterior belief of 63.6% to be ready to submit. This belief is 27.2% too high which is detrimental to the authors as it could result in premature resubmission.

While these two results are in line with our expectation, the more interesting question is whether gender-diversity in areas dominated by men leads to symmetric updating. The right bar of Figure 4.3 shows that this clearly is not the case. We find gender-diverse pairs to update significantly asymmetrically (17.9 percentage points; difference from zero: p-value = 0.0187). They respond 42% stronger to positive than to negative feedback ( $\beta_H/\beta_L = 0.6087/0.4297 \approx 1.42$ ; see column (3) of Table 4.1). Suppose Jake and Lindsey co-author a paper and hold a belief of 50% to re-submit. After having received two positive and two negative reports from their colleagues, they would end at a posterior belief of 59.7%. This estimate is 19.4% above the optimal Bayesian posterior. While the extent of asymmetry is slightly smaller among gender-diverse than among male pairs, the difference in asymmetry is not significant (p-value = 0.4179). The difference in asymmetry between gender-diverse and female pairs is, however, highly significant (p-value = 0.0158). Thus, with gender being salient, gender-diversity does not lead to less overconfident beliefs.

**Result 2:** We find gender-diverse pairs to update asymmetrically.

Table 4.1 presents the complete regression results. A subject updating optimally Bayesian has to be invariant to his prior beliefs which is shown by the coefficient  $\delta = 1$ . Being invariant means that the response to positive or negative signals must not depend on the level of the prior belief but only on the past signal itself. In other words, the response to signals must not be a function of the prior and therefore subjects with a prior of 20% should not update more or less optimally Bayesian than subjects with a prior of 50% or 80%. We find all pairs to be approximately invariant to their prior belief ( $\delta \approx 1$ ). Optimal Bayesian updating further requires subjects to fully incorporate available information for which the condition  $\beta_H - \beta_L = 1$  has to be met. We find pairs to update conservatively in response to positive and negative feedback ( $\beta_H < 1, \beta_L < 1$ ; p-values < 0.0001) which means that subjects do not fully appreciate available information. These findings are in line with previous findings (e.g. Möbius *et al.* 2007, 2014; Eil and Rao 2011; Buser *et al.* 2016; Cacault and Grieder 2018; Coutts 2018).

For the remainder of the paper, we only focus on asymmetric belief updating as it has the largest impact on belief distortions in the long run. While inaccurate priors and a (slight) lack of

---

<sup>13</sup>To test for the difference in asymmetry across different gender subgroups, we use a seemingly unrelated estimation using the error terms of the OLS regressions (Zellner, 1962).

invariance mainly prolong the time it takes to arrive at the optimal belief, asymmetric updating distorts beliefs in a way which is not corrected for in the long run.

TABLE 4.1 – **Belief updates per gender subgroup.**

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , where  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	<b>FF</b> <b>(1)</b>	<b>MM</b> <b>(2)</b>	<b>MF</b> <b>(3)</b>
$\delta$	0.9150** (0.0388)	1.1110** (0.0457)	1.0286 (0.0468)
$\beta_H$	0.5101*** (0.0424)	0.6714*** (0.0554)	0.6087*** (0.0452)
$\beta_L$	0.5230*** (0.0457)	0.4177*** (0.0626)	0.4297*** (0.0558)
$P(\beta_H = \beta_L)$	0.8361	0.0015	0.0128
Observations	262	298	252
R-squared	0.7827	0.8193	0.7908
p-values	<b>(1) vs (2)</b>	<b>(1) vs (3)</b>	<b>(2) vs (3)</b>
Diff. in asym.	0.0030	0.0158	0.4179

### 4.5.3 Disagreement leads to asymmetric updating among gender-diverse pairs

One of the essential ingredients to communication is the possibility to disagree. We argued above that the likelihood to disagree increases with more dispersed individual beliefs. This should especially materialize in gender-diverse pairs where (under)confident women are matched to overconfident men. The dispersion of belief B (the belief which is not yet affected by new information or chatting), however, does not confirm this hypothesis. If we look at the dispersion of these individual beliefs per gender subgroup, the dispersion is the lowest for gender-diverse pairs with a standard deviation of 11.4 percentage points. This is significantly lower than the dispersion of beliefs in male (14.3 percentage points; one-sided two-sample variance comparison test: p-value = 0.0286) and female pairs (13.7 percentage points; p-value = 0.0723). This tells

us that beliefs among individuals in gender-diverse pairs are in general the least dispersed. In a second step, we look at the average spread between the two beliefs of a pair. With an average spread of 13 percentage points, we find the difference to be the lowest within gender-diverse pairs. This spread is lower than within male pairs (15 percentage points,  $p$ -value = 0.1736) and significantly lower than within female pairs (16.3 percentage points,  $p$ -value = 0.0537). Thus, it is not the case that gender-diverse pairs profit from a more dispersed belief space, but in contrast show the least dispersed belief space. According to the force of disagreement, gender-diverse pairs should be the least likely to disagree and should therefore profit the least from disagreement.

To causally test this effect, we measure disagreement by counting how many tries it took a pair to find a unanimous belief. Per round, a pair has a maximum of two tries to find a unanimous belief (after the first chat and after the second chat) which results in a maximum disagreement score of 10 across all five rounds. Pairs where both members have the same individual beliefs (i.e. beliefs C.1 and D.1 in Figure 4.1) before the first chat in all five rounds receive a score of zero. These pairs never disagreed as their individual beliefs always directly matched. A higher score implies more disagreement. Unsurprisingly, scores of zero or ten do not exist. Scores in-between are given to pairs that needed the respective amount of tries. The highest score in our sample is 6, the lowest 1. Again, it is not gender-diverse pairs that disagree the most. With an average of 4.3 tries, female pairs disagree the most and significantly more than male pairs (3.8;  $p$ -value = 0.0121). Gender-diverse pairs range in-between with a score of 4.1 (not significantly different from either of the two other subgroups,  $p$ -values > 0.2555).

We group pairs into above-median and below-median disagreement groups within each gender subgroup and run separate regressions for each (six regressions). Figure 4.4 presents the difference between the response to high and low signals for each of the six subgroups separately (for the full regression results, see Table 4.2). Symmetric updating requests all bars to be equal to zero. For female and male pairs, we find disagreement to help making less asymmetric belief updates. While female pairs with little disagreement even update negatively asymmetrically (“underconfident”;  $p$ -value = 0.0565), female pairs who disagree update symmetrically ( $p$ -value = 0.1520). The difference in asymmetry between low- and high-disagreement pairs is highly significant ( $p$ -value = 0.0008). We observe a similar pattern among male pairs, yet in the other direction. Those who tend to agree update strongly asymmetrically by overvaluing positive feedback ( $p$ -value = 0.0032), while those who disagree do not show any sign of asymmetric updating ( $p$ -value = 0.1892). The difference in asymmetry between high- and low-disagreement groups is significant ( $p$ -value = 0.0445). The findings among male and female pairs are thus in line with our hypothesized direction of the force of disagreement. When we look at gender-diverse pairs, we observe the opposite relation. While pairs which do not disagree update symmetrically ( $p$ -value = 0.2136), those that do disagree update strongly asymmetrically ( $p$ -value = 0.0079). The difference in asymmetry is significant ( $p$ -value = 0.0418) and indicates that disagreement

leads to more asymmetric updating. Hence, while disagreement helps within single-gender pairs, we surprisingly find it to be counterproductive in gender-diverse pairs.

TABLE 4.2 – **Belief updates and disagreement.**

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , with  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . *Disagreement* is a score between 0 and 10 and counts the number of chats which pairs needed to come to a unanimous belief in each of the five rounds of the experiment. Pairs where both members had the same individual belief before the first chat in all five rounds receive a score of zero. These pairs never had to find a unanimous belief as their individual beliefs were the same in each round. A score of ten is given to pairs that needed two chatting rounds in each round to find a unanimous belief. Consequently, scores between zero and ten indicate that pairs needed the respective number of chats. A higher score implies more disagreement. Pairs were grouped into above-median and below-median score groups within each gender subgroup. Separate regressions were run for above-median and below-median score groups for each gender subgroup (six regressions). Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	FF		MM		MF	
	(1) disagree	(2) agree	(3) disagree	(4) agree	(5) disagree	(6) agree
$\delta$	0.9639 (0.0428)	0.8021*** (0.0695)	1.1312** (0.0529)	1.0968 (0.0871)	1.0155 (0.0700)	1.0538 (0.0546)
$\beta_H$	0.5809*** (0.0464)	0.4151*** (0.0697)	0.6261*** (0.0592)	0.7526** (0.1157)	0.8538* (0.0773)	0.4158*** (0.0457)
$\beta_L$	0.4707*** (0.0596)	0.6086*** (0.0706)	0.5037*** (0.0737)	0.3250*** (0.1147)	0.5333*** (0.0896)	0.3231*** (0.0597)
$P(\beta_H = \beta_L)$	0.1520	0.0565	0.1892	0.0032	0.0079	0.2136
Observations	120	142	180	118	116	136
R-squared	0.8670	0.7253	0.8351	0.8070	0.8028	0.8317
p-values	(1) vs (2)		(3) vs (4)		(5) vs (6)	
Diff. in asym.	0.0008		0.0445		0.0418	

**Result 3:** Disagreement leads to asymmetric belief-updating in gender-diverse pairs.

An explanation why disagreement does not help gender-diverse pairs could lie in the presence of a dominant subject. The imposition of one's own view on the team could cause disagreement to arise without the expected positive effects.

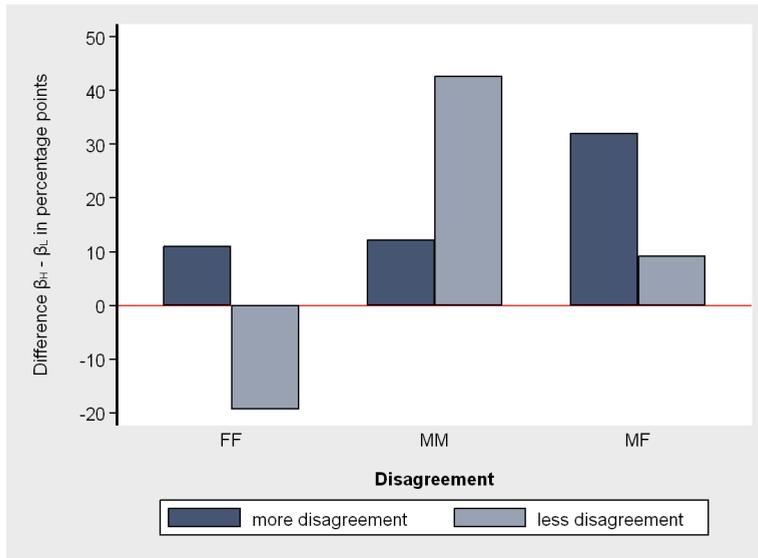


FIGURE 4.4 – Asymmetry and disagreement.

This graph plots the difference between coefficients  $\beta_H$  and  $\beta_L$  from Table 4.2. Updating according to Bayes' Law requires all bars to be equal to zero. *Disagreement* is a score between 0 and 10 and counts the number of chats which pairs needed to come to a unanimous belief in each of the five rounds of the experiment. Pairs where both members had the same individual belief before the first chat in all five rounds receive a score of zero. These pairs never had to find a unanimous belief as their individual beliefs were the same in each round. A score of ten is given to pairs that needed two chatting rounds in each round to find a unanimous belief. Consequently, scores between zero and ten indicate that pairs needed the respective number of chats. A higher score implies more disagreement. Pairs were grouped into above-median and below-median score groups within each gender subgroup. Separate regressions were run for above-median and below-median score groups for each gender subgroup (six regressions). A positive difference  $\beta_H > \beta_L$  implies asymmetric belief updating where positive feedback is valued stronger than negative feedback. Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered.

#### 4.5.4 Dominant behavior drives asymmetric updating in gender-diverse pairs

We define someone to be dominant if he or she has a greater influence on the team decision than the teammate. Thereby, we take advantage of the knowledge we have on individual priors before the first chat (belief C.1, see Figure 4.1) and how they differ from the resulting team belief (belief C.2 or C.3). It is important that we look at beliefs during the first interaction phase, as teammates do not know each other at that point in time. The subject with the bigger impact on the very first team decision is defined to be dominant. A pair is defined to host no dominant person if the team belief lies in the middle of the two individual beliefs. Subjects then agree to meet in the middle and neither of the two tries to impose his belief on the pair.

Analogously, we define a pair to host a dominant person if the team belief does not lie in the middle as the belief of one individual dominated the belief of the other. Suppose Jake has a pre-signal prior of 80% and Lindsey of 60% (belief C.1). During the chat, they decide to meet in the middle at 70%. Neither of the two would be considered dominant. Instead, imagine the two with the same priors as above decide on 75% as joint belief. Then Jake would be considered to have dominated Lindsey as the team belief lies closer to his individual prior than to Lindsey's.

We split gender subgroups into pairs which “meet in the middle”<sup>14</sup> and those that do not. Among male and female pairs, we find 69.5% respectively 72.7% to host a dominant person. In other words, roughly 30% of single-gender pairs meet in the middle at their first chatting encounter. With gender-diverse pairs, the number of pairs hosting a dominant individual is with 63.6% slightly lower. The remaining 36.4% meet in the middle.

To test the effect of dominance on belief updating, we run two separate regressions for each gender subgroup. One contains the pairs which host a dominant subject, the other regression those pairs which meet in the middle. Figure 4.5 presents the difference between the response to high and low signals for each of the six subgroups separately (for the full regression results, see Table 4.3). Symmetric updating requires all bars to equal zero. Dominance does not play a major role in explaining updating behavior among single-gender pairs. In line with our expectation, neither the presence nor the absence of a dominant woman causes asymmetry (difference in asymmetry within dominant and non-dominant subgroups: p-values > 0.3314). The presence of a dominant man in male pairs cannot explain asymmetric belief updating. Male pairs update asymmetrically regardless of the presence of a dominating individual (p-values < 0.0358). This finding suggests that a lack of disagreement is a stronger driver of asymmetric updating among male pairs than the presence of a dominant individual.

Among gender-diverse pairs, dominance plays a strong role. The presence of a dominant person can entirely explain asymmetric updating. A dominant subject causes gender-diverse pairs to value positive feedback 66.5% stronger than negative feedback ( $\beta_H/\beta_L = 0.7008/0.4208 \approx 66.5\%$ ; p-value = 0.0058; see column (3) of Table 4.3). Strikingly, gender-diverse pairs that meet in the middle update symmetrically (p-value = 0.6418). The difference in asymmetry between the two subgroups is highly significant (p-value = 0.0383). Thus, we indeed find dominant behavior to be the driving force behind asymmetric updating. But more importantly, a simple arithmetic mean of the two beliefs solves the problem of asymmetric belief updating in gender-diverse pairs. This is even more remarkable as this is against the theoretical prediction that taking the average of (partly) biased beliefs cannot de-bias the joint belief. We can see this among male pairs who update strongly asymmetrically whenever they meet in the middle. This is valuable information as it offers gender-diverse pairs a simple external rule of how to approach the formation of joint beliefs.

<sup>14</sup>This also includes pairs which (by chance) have the same individual pre-signal priors.

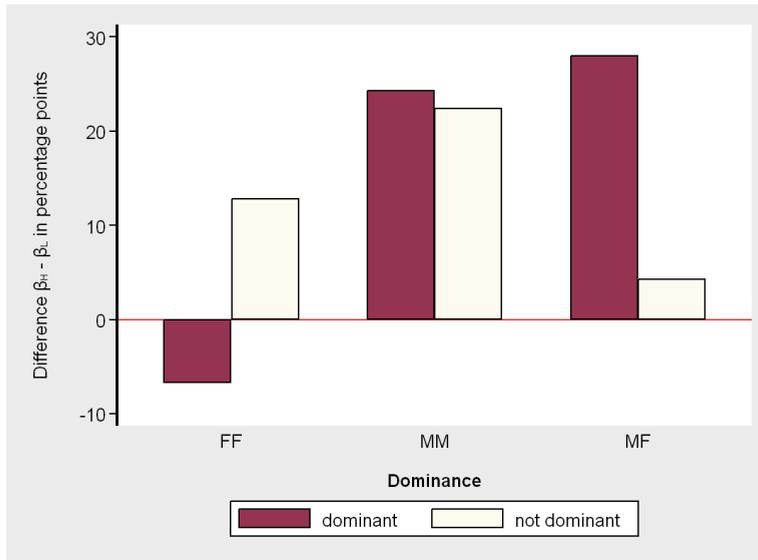


FIGURE 4.5 – **Asymmetry and dominance.**

This graph plots the difference between coefficients  $\beta_H$  and  $\beta_L$  from Table 4.3. Updating according to Bayes' Law requires all bars to be equal to zero. We define someone to be *dominant* if he or she has a greater influence on the team decision during the first chatting encounter than the teammate. Specifically, we compare individual beliefs (belief C.1, see Figure 4.1) and how they differ from the resulting team belief (belief C.2 or C.3). A pair is defined to host *no dominant* person if the team belief lies in the middle of the two individual beliefs. Analog, we define a pair to host a dominant person if the team belief does not lie in the middle but consequently closer to the individual belief of one of the two teammates. This teammate's belief dominated the belief of the other. Separate regressions are run for dominant and non-dominant groups for each gender subgroup (six regressions). A positive difference  $\beta_H > \beta_L$  implies asymmetric belief updating where positive feedback is valued stronger than negative feedback. Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded (for a more detailed explanation of exclusions, see section 4.3.8). Beliefs which are within the first or last decile prior to receiving a signal are also excluded. Only the final decision per round is considered.

**Result 4:** The presence of a dominant person is a strong cause for asymmetric updating among gender-diverse pairs. Gender-diverse pairs where subjects agree to meet in the middle overcome this bias and update symmetrically.

#### 4.5.5 Men and women are responsible for dominant behavior within gender-diverse pairs

To better understand, why dominance has such a strong effect on the updating behavior of gender-diverse pairs, we analyze whether this effect stems from the male or female teammate and why they act differently from men or women in single-gender pairs. We again employ the actual impact as proxy for dominance. Instead of comparing behavior across different gender

TABLE 4.3 – Belief updates and dominance.

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , with  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . We define someone to be dominant if he or she has a greater influence on the team decision during the first chatting encounter than the teammate. Specifically, we compare individual beliefs (belief C.1, see Figure 4.1) and how they differ from the resulting team belief (belief C.2 or C.3). A pair is defined to host *no dominant (not dom)* person if the team belief lies in the middle of the two individual beliefs. Analog, we define a pair to host a *dominant (dom)* person if the team belief does not lie in the middle but consequently closer to the individual belief of one of the two teammates. This teammate's belief dominated the belief of the other. Separate regressions are run for dominant (do not meet in the middle) and non-dominant pairs (do meet in the middle) for each gender subgroup (six regressions). Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	FF		MM		MF	
	(1) dom	(2) not dom	(3) dom	(4) not dom	(5) dom	(6) not dom
$\delta$	0.9255* (0.0433)	0.9160 (0.0920)	1.1233* (0.0662)	1.1216** (0.0545)	1.0073 (0.0623)	1.0525 (0.0654)
$\beta_H$	0.4897*** (0.0493)	0.5620*** (0.0855)	0.7130*** (0.0709)	0.5453*** (0.0817)	0.7008*** (0.0691)	0.5001*** (0.0491)
$\beta_L$	0.5568*** (0.0528)	0.4335*** (0.0952)	0.4701*** (0.0854)	0.3208*** (0.0803)	0.4208*** (0.0735)	0.4568*** (0.0789)
$P(\beta_H = \beta_L)$	0.3530	0.3315	0.0234	0.0357	0.0058	0.6418
Observations	190	72	202	96	156	96
R-squared	0.7981	0.7416	0.7781	0.9033	0.7794	0.8387
p-values	(1) vs (2)		(3) vs (4)		(5) vs (6)	
Diff. in asym.	0.2802		0.9050		0.0383	

subgroups, we now compare how the updating behavior within gender-diverse pairs is affected by whether the man or the woman is dominant.

We split pairs into two subgroups “dominated by male” (where the team belief lies closer to the male belief) and “dominated by female” (where the team belief lies closer to the female belief). The third group, where the two team members meet in the middle is the same as in section 4.5.4. We find pairs to be quite evenly spread across these two subgroups with 20 pairs being dominated by a male and 19 by a female subject. Figure 4.6 shows that gender-diverse pairs with a dominant man update highly asymmetrically ( $\beta_H - \beta_L = 28.9$  percentage points, p-value = 0.0303) while valuing positive feedback 1.7 times as high as negative feedback. The

same effect can be observed with pairs hosting a dominant woman ( $\beta_H - \beta_L = 27.6$  percentage points,  $\beta_H/\beta_L \approx 1.7$ ; p-value = 0.0825). And even though the significance level for the latter subgroup is lower, the coefficients for  $\beta_H$  and  $\beta_L$  barely differ across both subgroups (compare columns (1) and (2) of Table 4.4). Overall, we find supporting evidence that not only dominant men, but also dominant women are a root cause for asymmetric updating in gender-diverse pairs.

**Result 5:** Asymmetric updating in gender-diverse pairs is caused to an equal extent by dominant men and dominant women.

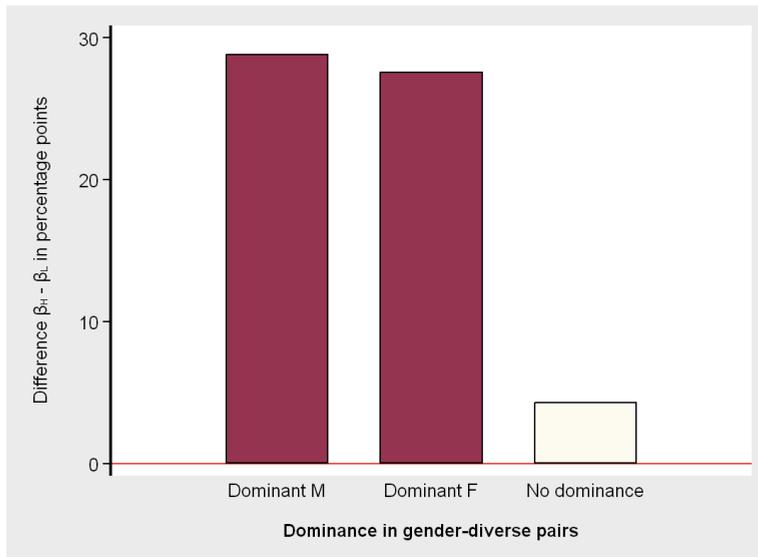


FIGURE 4.6 – Asymmetry and dominance in gender-diverse pairs.

This graph plots the difference between coefficients  $\beta_H$  and  $\beta_L$  from Table 4.4. Updating according to Bayes' Law requires all bars to be equal to zero. Only gender-diverse pairs are considered. We define someone to be *dominant* if he or she has a greater influence on the team decision during the first chatting encounter than the teammate. Specifically, we compare individual beliefs (belief C.1, see Figure 4.1) and how they differ from the resulting team belief (belief C.2 or C.3). A pair is defined to host no dominant person if the team belief lies in the middle of the two individual beliefs. Analog, we define a pair to host a dominant person if the team belief does not lie in the middle but consequently closer to the individual belief of one of the teammates. This teammate's belief dominated the belief of the other. Separate regressions are run for pairs where the male teammate dominates and for pairs where the female teammate dominates (two regressions). A positive difference  $\beta_H > \beta_L$  implies asymmetric belief updating where positive feedback is valued stronger than negative feedback. Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered.

As dominant women cause asymmetric updating in gender-diverse but not in female pairs, the question arises whether women change their behavior when put in a gender-diverse pair?

We find evidence of such a behavioral change when we compare the level of dominance of women in single-gender and gender-diverse pairs. In other words, how much does a dominant woman force another woman to come her way? And how much does she force another man to give in?

To analyze the degree of overconfidence, we use the conviction ratio.<sup>15</sup> A conviction ratio of 100% means that the dominant person's individual belief was taken as team belief, and a conviction ratio of 50% means that the two teammates meet in the middle. A percentage in-between 50% and 100% consequently signals, how close the team belief is towards the belief of the dominant subject.

Dominant women in female pairs on average convince their teammate to come 71% her way. Thus, the dominated woman gives in 71%, while the dominant woman concedes a mere 29% to find a unanimous belief. In gender-diverse pairs, the dominant woman manages to make the man come 87% her way (difference between conviction ratios:  $p$ -value = 0.0110). In other words, women are significantly more dominant towards men than towards women. On the contrary, a dominant man in gender-diverse pairs convinces his female teammate to come 73% his way which is significantly lower than the 87% of dominant women ( $p$ -value = 0.0689). Interestingly, it is also slightly lower than the conviction ratio men show when matched to another man (81%;  $p$ -value = 0.1412).

Hence, we find support for a behavioral change with women who become more dominant towards men which could explain why gender-diversity does not per se lead to more accurate belief updates.

## 4.6 Additional findings and robustness checks

### 4.6.1 Perceived math skills and asymmetric belief updating

Experience is an important reference point for confidence estimates (e.g. Fiore 1999; Marel *et al.* 2000; Brady and Bowd 2005). In our experiment, general mathematical skills help subjects to estimate how good they perform in the calculation task. Those who assess their mathematical skills to be above average are likely more confident about being in the top half than those with lower skills. In an optimal Bayesian world, subjects use their mathematical skills to form their prior belief. The processing of new information is, however, independent thereof. In other words, two subjects with a prior belief of 50% should respond in exactly the same way to a signal –

<sup>15</sup>The conviction ratio is calculated as follows: the numerator contains the distance between the teammate's individual prior and the unanimous team belief and the denominator the distance between the two individual priors. Take the example from above where Jake has a pre-signal prior of 80% and Lindsey of 60%. During the chat, they decide on 76%. Then, the ratio of Jake is  $(76 - 60)/(80 - 60) = 80\%$ . Analog, the ratio of Lindsey is  $(80 - 76)/(80 - 60) = 20\%$ . Jake convinced Lindsey to come 80% his way.

TABLE 4.4 – Belief updates per gender subgroup.

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , with  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . Only gender-diverse pairs are considered. We define someone to be dominant if he or she has a greater influence on the team decision during the first chatting encounter than the teammate. Specifically, we compare individual beliefs (belief C.1, see Figure 4.1) and how they differ from the resulting team belief (belief C.2 or C.3). We define a pair to host a *dominant (dom)* person if the team belief does not lie in the middle but consequently closer to the individual belief of one of the two teammates. This teammate's belief dominated the belief of the other. Separate regressions are run for dominant (do not meet in the middle) and non-dominant pairs (do meet in the middle) for each gender subgroup (six regressions). Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

	MF	
	dom M (1)	dom F (2)
$\delta$	0.9842 (0.0769)	1.0420 (0.1047)
$\beta_H$	0.7159*** (0.0933)	0.6807*** (0.1046)
$\beta_L$	0.4274*** (0.0910)	0.4048*** (0.1225)
$P(\beta_H = \beta_L)$	0.0303	0.0825
Observations	80	76
R-squared	0.7952	0.7657

irrespective of their mathematical skills. Yet, there are two rationales why updates could depend on math skills. First, Bénabou and Tirole (2002) argue that subjects do not only derive direct utility from holding the correct belief but indirect utility from believing that they are good at a specific task. We argue accordingly that subjects who assess their own mathematical skills to be high should derive indirect utility from a positive but not from a negative signal. Hence, subjects with higher mathematical skills would update asymmetrically. Second, asymmetric updating could stem from those with worse math skills. If this was the case, our results would not be driven by self-image concerns but by a task-inherent ability. This would prevent us from making inferences about the general updating behavior of pairs but reduce our results to a question of knowledge and ability regarding the task.

We elicit self-reported mathematical skills by asking subjects to answer the statement "*My math skills are higher than those of \_\_\_ of my fellow students.*" on a 11-point Likert scale from

0% to 100% in steps of ten percentage points. Since we look at pairs, we proxy the average mathematical skill of the pair by taking the average of the skills of each member. This average is positively and significantly correlated with the team score ( $\rho = 0.3551$ ; p-value  $< 0.0001$ ). We then group pairs in above-median and below-median skilled groups within each gender subgroup. The average male pair rates its mathematical skills to be better than that of 68% of their fellow students (see Table A4.3). The average female pair reports to be better than 58%, which is significantly lower (p-value = 0.0006). Gender-diverse pairs lie with 60% close to female groups (p-value = 0.3243) and significantly below male pairs (p-value = 0.0104). These findings showing a higher level of reported mathematical skills is in line with previous findings that men have higher confidence in their skills to execute a stereotypically male task even in the absence of a gender performance gap (Eccles, 1993; Cvencek *et al.*, 2011).

Figure A4.2 presents the difference between the response to high and low signals for each gender subgroup separately (for the full regression results, see Table A4.4). Thereby, we run different regressions for the above-median skilled and below-median skilled groups of each gender subgroup (in total six regressions). Each bar corresponds to the difference  $\beta_H - \beta_L$  of each regression. Symmetric updating requires all bars to be equal to zero. We observe that asymmetric belief updating is driven by pairs which report above-median mathematical skills. Male pairs with above-median mathematical skills react 28.2 percentage points stronger to positive than to negative signals (p-value = 0.0203), gender-diverse pairs with above-median mathematical skills 29.4 percentage points stronger (p-values = 0.0229). Male and gender-diverse pairs with below-median skills update symmetrically (p-values of asymmetry  $> 0.5011$ ). On the other side, reported mathematical skills do not have any significant effect on the updating behavior of female pairs. Pairs with high and low skills alike update symmetrically (p-values of asymmetry  $> 0.1865$ ).

These findings support the self-image explanation of asymmetric updating where those who report to be good at the calculation task derive additional utility from responding to positive signals. Importantly, this finding also shows that the ability to correctly update is not related to the reported math skills. This dilutes the concern that those who are bad at a calculation task also are bad at updating their beliefs. Our results much rather hold despite the fact that we use a mathematical task to elicit performance.

### 4.6.2 Robustness check of the main findings

Besides looking at the asymmetric updating stemming from the regression results presented in section 4.5.2, Figure A4.3 presents the development of the average beliefs throughout the five experimental rounds (the prior and four updates). The advantage of looking at the level of beliefs (and not the logit function) is that we can include extreme beliefs (see section 4.3.8). Bayesian updating predicts that pairs which are in the top half increase their beliefs closer

to 100% when receiving informative signals while pairs in the lower half decrease their beliefs towards 0%. Consequently, after an infinite number of signals, all pairs in the top half would hold beliefs of 100% and all in the bottom half of 0% resulting in an average belief of 50% across pairs in either half. Figure A4.3 depicts that male and gender-diverse pairs hold beliefs which are significantly higher than the benchmark of 50% while female pairs stay close to an average belief of 50% throughout the experiment. This is in line with the findings that male and gender-diverse pairs update asymmetrically which eventually leads to unjustifiably optimistic beliefs.

In a second step, we include pairs in our main regression which never update or update contrary to the received signal at least once. Table A4.5 presents the results of a pooled OLS regression of equation 4.3 where all pairs are included. The findings confirm our main results presented in Table 4.1. Male and gender-diverse pairs update highly asymmetrically (p-values < 0.0082) by valuing positive feedback 53%, respectively 62% more than negative feedback. Female pairs again update symmetrically (p-value = 0.3986). The difference in asymmetry between male and female pairs as well as between gender-diverse and female pairs stays highly significant (p-values < 0.0376). Thus, our findings are robust to extreme beliefs and to including all pairs.

In a third step, we test the robustness of the findings regarding disagreement. Table A4.6 shows the results of a pooled OLS regression of equation 4.3 where we again include all pairs. These pairs are put into above-median and below-median disagreement groups within each gender subgroup and separate regressions are run for each (six regressions). The general direction of these results stays similar to the results of Table 4.2. Yet, some differences deserve a closer look. Female pairs who tend to disagree show signs of asymmetric updating ( $\beta_H - \beta_L = 17.6$  percentage points, p-value = 0.0114). For male and gender-diverse pairs, results seem to be robust. The lack of robustness for our findings on female pairs might have to do with the fact that the percentage of those who update contrary to a signal is highest among female pairs. The fact that the regression coefficients are highly sensitive to incorrect updating might therefore be an explanation of why female pairs suddenly become overconfident. Yet, it is not clear to us what drives the high number of incorrect updates which seems to be a common problem when working with experiments on belief updating (Möbius *et al.*, 2014; Coutts, 2018).

Finally, we test the robustness of the impact of dominance on belief updates. Table A4.7 shows the results of a pooled OLS regression of equation 4.3 where we once more include all pairs. Analog to the procedure explained in section 4.5.4, pairs within each gender subgroups are split into those that “meet in the middle” (no dominant individual present) and those that do not (dominant individual present). Results again provide an indication that the impact of dominance is the same once we include all pairs. Yet, effects become smaller and partially non-significant.

With all three robustness checks on including all pairs, we observe that the explanatory power of the model drops which makes sense given that the pairs included update contrary to the model mechanism. Further, we observe that pairs become on average more conservative which is not entirely clear to us why this should be the case. Finally, we lack power to exhaustively address all of the above raised concerns and thus have to conclude that our results only seem to hold when we exclude those who never update or update in the wrong direction.

### 4.6.3 Using a different proxy for dominance

Next to looking at the actual impact of participants to measure dominant behavior, we analyze the subject's perception of who had the bigger impact on the team decisions. After the calculation and updating experiment was over, we asked subjects to rate the statement "*My word had a bigger impact on the joint decision than my teammate's word.*" on a 7-point Likert scale (from 1—fully disagree to 7—fully agree). The correlation of the conviction ratio and the impact rating is positive and highly significant ( $\rho = 0.1869$ ; p-value = 0.0061). Thus, an individual claiming to have had a higher impact on the team decision also was better at convincing the teammate to come his or her way to find a unanimous belief. This shows that the two proxies for dominance pick up—at least to a certain extent—the same underlying mechanism. The advantage of this rating is that it also considers what happened in the later rounds of the experiment. This additional control allows us to see whether our results using a measure for actual impact are in line with the results we find when using the team's perception of who had the bigger impact. We assign the higher of the two impact ratings within a pair as proxy for the presence of dominance. The average maximum impact ratings are slightly higher for male pairs than for mixed-gender pairs (5.3 vs 4.9; p-value = 0.0928). The average maximum rating among female pairs is with 5.1 not different from any of the other two scores (p-values > 0.3412).

Regression results are presented in Table A4.8 and are in line with our previous results on the real impact. Among female pairs, neither pairs with an above- nor below-median maximum impact rating show asymmetric updating ( $\beta_H - \beta_L < 2$  percentage points in either regression; testing difference from zero: p-values > 0.7520). Within male pairs, both, the presence and absence of high impact individuals results in asymmetric updating ( $\beta_H - \beta_L > 20$  percentage points in either regression; difference from zero: p-values < 0.0849). Among gender-diverse pairs, the presence of a dominant individual again entirely explains the existence of asymmetric updating. Pairs with dominant individuals value positive feedback two and a half times as much as negative feedback ( $\beta_H/\beta_L \approx 2.5$ ; p-value = 0.0028) while its absence leaves pairs updating almost perfectly symmetrically (p-value = 0.9154). The difference in asymmetry between these two subgroups is highly significant (p-value = 0.0001). Noteworthy is that the self-reported impact causes even stronger asymmetry than our proxy for actual impact. This could refer to the fact that our measure of actual impact only captures dominance at a single moment of time

while the self-rated impact measure was elicited after the experiment and thus captures the within-group dynamics of an additional four updating rounds.

In a second step, we check whether men and women are equally responsible for the dominant behavior that leads to asymmetric updating. Analog to our procedure in section 4.5.5, we split pairs into three subgroups where either the male or the female member claims to have had higher impact, or where the impact ratings were equal. Table A4.9 presents the results which are similar to our previous findings. Pairs with a dominant man update strongly asymmetrically (p-value = 0.0092; column (1)). Pairs with a dominant woman also update asymmetrically even though with only marginal significance (p-value = 0.1301; column (2)). However, both, male and female dominated pairs update significantly more asymmetrically than pairs where both teammates have equal impact (difference in asymmetry: p-values < 0.0906). And when pairs with a dominant man and those with a dominant woman are put in a single regression, the level of asymmetry is with 26.4 percentage points highly significant (p-value = 0.0162).

Those pairs where both felt that they had equal impact on the team decision update symmetrically (p-value = 0.6649; column (4)). This supports our finding that in gender-diverse pairs, the solution to updating beliefs is to meet in the middle and abstain from dominant behavior.

## 4.7 Conclusion

The body of literature studying the advantages of gender diversity is steadily growing. Especially research on how board decisions are affected by gender composition is a well-researched area. While it is highly relevant to compose field studies, it is of equal importance to show that there exists a causal link between gender diversity and the quality of decisions. In this paper, we close this gap and study the causal link between the gender composition of pairs and the overconfidence of their beliefs. But the topic of gender-diverse professional teams goes well beyond the question of how to compose company boards. Many academics write papers in teams of two where research topics should be chosen based on qualitative decisions and submissions done only after a thorough judgment of the quality of the paper. And even on a more general basis, couples make household decisions on a daily basis, which at times can include big investments like buying a car or building a house. All of these decisions have in common that protagonists have to form beliefs about the likelihood of success of their action.

We focus on poor decision quality through asymmetric belief updating, which in the long-run leads to unjustifiably optimistic beliefs or, in common terms, to overconfidence. We find causal evidence that gender-diversity in pairs causes asymmetric belief updates which are equal in size to those of male pairs. Only female pairs update symmetrically. Gender-diverse pairs on average respond 42% stronger to positive than to negative feedback. We show that asymmetric belief updating in gender-diverse pairs is driven by dominant teammates who try to impose

their belief on the team and thereby scotch the benefits of group interaction. Strikingly, men and women are equally dominant. Only gender-diverse pairs where both agree to meet in the middle of their individual beliefs update symmetrically.

These findings give valuable insights which can help to base the discussions around gender diversity on scientific grounds. With regard to overconfident decision-making, we find that the status quo where the majority of business decisions is taken by men in a male-dominated environment can cause seriously negative financial consequences. At the same time, purely female teams do not suffer from overconfident decisions in such an environment. Yet, it is highly unlikely that the majority of professional decisions will be taken by pure female teams in the near future. Thus, a conscious decision for gender-diverse pairs where a special focus is put on avoiding situations in which one team member dominates the other can be a viable and feasible solution to reduce overconfident decision-making.

Our study thereby provides ample directions for future research. First, we designed our experiment in such a way that we randomly assigned individuals to a pair, while people in the labor market self-select into their jobs (Dohmen and Falk, 2011; Fouarge *et al.*, 2014). For example, men and women in top positions have different characteristics than those of the average employees in a firm (Adams and Funk, 2012). It would be interesting to see how our results generalize to a setting with self-selection. Second, we only consider pairs where further decision-relevant dimensions like majority-rule or voting behavior are excluded. It is for future research to show how larger groups with a majority of men or a majority of women (e.g. three versus two) form and update joint beliefs. Group size also plays a vital role in the question of how far dominant behavior of individuals can affect group decisions. While Anderson and Kilduff (2009) show that the findings of dominant individuals dominating groups is quite consistent, it might make a difference whether someone has to impose his or her own belief on just one or an entire group of people. In addition, Asch (1956) shows that group-conform behavior starts to play a major role in group sizes of three or higher where it is not clear how group-conformity interacts with dominant behavior. Third, while we deliberately chose to avoid synergies regarding the task we used, teams in real-life situations mostly benefit from synergies in producing a certain good or coming to a certain decision faster. It is for future research to show how the gender composition of a group affects possible synergy-effects. Fourth, we only consider the overconfidence bias. It remains to be studied which effect gender diversity has on other impediments to optimal decision-making. The debate about gender diversity benefits from scientific findings providing causal evidence.

Last but not least, we have to leave some of the questions unanswered. We, for example, would like to address the question of how the force of disagreement and the force of dominance interact by dividing each gender subgroup into four types of pairs with high disagreement/high dominance, high disagreement/low dominance, low disagreement/high dominance, and low disagreement/low dominance. Yet, our sample size does not suffice for such deep levels of analysis.

This would help us to better understand why gender-diverse pairs suffer from disagreement. It could be that dominant behavior leads to heavy disagreement among teammates where one of the two eventually gives in. However, one can also think of a situation in which a dominant person decides by not allowing a discussion to arise. These interaction effects have not been studied and could shed further light into how dominance affects decision-making in pairs.

## A Appendix

### A.1 Tables

TABLE A4.1 – Variable definition

Variable	Description
Team score	The sum of the two individual scores form the team score. Individual scores equal the number of correctly solved calculations in the timed calculation task and can vary between 0 and 25. Thus team scores can vary between 0 and 50. Individual scores are not reported.
Disagreement	Disagreement is a score between 0 and 10 and counts the number of chats pairs needed to come to unanimous beliefs throughout the experiment. Pairs where both members had the same individual belief before getting to chat in all five rounds receive a score of zero. These pairs never had to find a unanimous belief as their beliefs were the same before getting to chat. A score of ten is given to pairs that needed two chatting rounds in each round to find a unanimous belief. Scores in-between are given to pairs that needed the respective number of chats. Consequently, a higher score implies more disagreement.
Dominance (actual impact)	Measured through the conviction ratio in the round prior to receiving the first signal. The conviction ratio indicates the degree to which a person convinced the teammate to come his or her way. The numerator of the conviction ratio contains the distance between the teammate's individual pre-signal prior belief and the pre-signal team belief. The denominator comprises the distance between the two individual pre-signal priors. The ratio yields information on whether the team belief lies in the middle of the individual beliefs (the ratio equals 0.5) or whether one of the two could convince the other that the team belief should come his or her way more than 50% (the ratio is greater than 0.5).
Dominance (perceived impact)	The participant's response to the statement "My word had a bigger impact on the joint decision than my teammate's word." on a 7-point Likert scale from 1—fully disagree to 7—fully agree.
Mathematical skills	The participant's response to the statement "My math skills are higher than those of ___ of my fellow students." on a 11-point Likert scale from 0% to 100% in steps of ten percentage points.

TABLE A4.2 – Simulated belief updates.

This table presents pooled OLS estimates for a simulated sample of 1,000 subjects of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , where  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . Analog to our experiment, 50% of the simulated subjects are placed into the top and 50% into the bottom half (in our simulation, placement is random). Signal accuracy is 75%. Dependent variable is the posterior belief. In column (1), all subjects update optimally Bayesian. In columns (2) to (4), subjects with beliefs in the specified decile respond one percentage point stronger to positive feedback than an optimal Bayesian; the response to negative feedback remains optimally Bayesian. All subjects with beliefs outside the specified decile still update optimally Bayesian. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

	(1) All update optimally Bayesian	(2) All update optimally Bayesian except if belief $\in [0.5, 0.6)$	(3) All update optimally Bayesian except if belief $\in [0.8, 0.9)$	(4) All update optimally Bayesian except if belief $\in [0.9, 1.0)$
$\delta$	1.0000*** (0.0000)	0.9999*** (0.0000)	1.0028*** (0.0002)	1.0340*** (0.0043)
$\beta_H$	1.0000*** (0.0000)	1.0041*** (0.0003)	1.0213*** (0.0012)	1.1088*** (0.0068)
$\beta_L$	1.0000*** (0.0000)	1.0000*** (0.0000)	0.9992*** (0.0002)	0.9633*** (0.0049)
$P(\beta_H = \beta_L)$	0.0000	0.0000	0.0000	0.0000
Observations	4,000	4,000	4,000	4,000
R-squared	1.0000	1.0000	0.9997	0.9893

TABLE A4.3 – Summary statistics.

This table presents summary statistics for the survey respondents. Variable definitions are provided in Table A4.1.

	Mean	Median	Std Dev	Obs.
Female	46.3%			99
Male	53.7%			115
Age	21.3	21	2.4	214
<i>Team calculation score</i>				
FF	16.8	16	4.9	66
MM	17.8	18	6.3	82
MF	18.1	18	5.5	66
p-values	>0.1439			
<i>Self-reported math skills (0 - 100%)</i>				
FF	58%	60%	15%	66
MM	68%	65%	17%	82
MF	60%	60%	14%	66
p-value (FF vs MM)	0.0006			
p-value (FF vs MF)	0.3243			
p-value (MM vs MF)	0.0104			
<i>Disagreement (0-10)</i>				
FF	4.3	4	1.1	66
MM	3.8	4	1.1	82
MF	4.1	4	1.3	66
p-value (FF vs MM)	0.0121			
p-value (FF vs MF)	0.2609			
p-value (MM vs MF)	0.2556			
<i>Rated impact (1-7)</i>				
FF	4.0	4	1.6	66
MM	4.1	4	1.8	82
MF	4.0	4	1.5	66
p-values	>0.5974			

TABLE A4.4 – **Belief updates and reported math skills.**

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $logit(\hat{\mu}_t) = \delta * logit(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , with  $logit(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . *Math skills* were elicited by asking subjects to answer the statement “My math skills are higher than those of \_ of my fellow students.” on a 11-point Likert scale from 0% to 100% in steps of ten percentage points. The math skill of a pair is calculated by taking the average of the math skills of its members. Pairs were grouped into above-median and below-median skilled groups within each gender subgroup. Separate regressions were run for above-median and below-median skilled groups for each gender subgroup (six regressions). Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	FF		MM		MF	
	(1) skilled	(2) not skilled	(3) skilled	(4) not skilled	(5) skilled	(6) not skilled
$\delta$	1.0334 (0.0687)	0.8723*** (0.0442)	1.1932*** (0.0654)	0.8941** (0.0529)	0.9982 (0.0744)	1.0169 (0.0548)
$\beta_H$	0.5334*** (0.0625)	0.4739*** (0.0537)	0.7084*** (0.0785)	0.4868*** (0.0635)	0.6979*** (0.0736)	0.4976*** (0.0478)
$\beta_L$	0.6659*** (0.0773)	0.4474*** (0.0538)	0.4267*** (0.0964)	0.5129*** (0.0636)	0.4042*** (0.0995)	0.4487*** (0.0552)
$P(\beta_H = \beta_L)$	0.1866	0.7263	0.0203	0.7649	0.0229	0.5012
Observations	134	128	186	112	136	116
R-squared	0.7264	0.8560	0.8198	0.8651	0.7670	0.8544
p-values	(1) vs (2)		(3) vs (4)		(5) vs (6)	
Diff. in asym.	0.1603		0.0103		0.0137	

TABLE A4.5 – Belief updates per gender subgroup (all participants included).

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , where  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . Beliefs which are within the first or last decile prior to receiving a signal are excluded (for a detailed explanation, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	FF (1)	MM (2)	MF (3)
$\delta$	0.8889*** (0.0343)	1.0857** (0.0426)	1.0421 (0.0408)
$\beta_H$	0.4294*** (0.0362)	0.6008*** (0.0527)	0.5330*** (0.0415)
$\beta_L$	0.3856*** (0.0372)	0.3939*** (0.0603)	0.3300*** (0.0496)
$P(\beta_H = \beta_L)$	0.3986	0.0081	0.0017
Observations	404	379	330
R-squared	0.7307	0.7677	0.7616
p-values	(1) vs (2)	(1) vs (3)	(2) vs (3)
Diff. in asym.	0.0375	0.0276	0.9627

TABLE A4.6 – Belief updates and disagreement (all participants included).

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , where  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . *Disagreement* is a score between 0 and 10 and counts the number of chats which pairs needed to come to a unanimous belief in each of the five rounds of the experiment. Pairs where both members had the same individual belief before the first chat in all five rounds receive a score of zero. These pairs never had to find a unanimous belief as their individual beliefs were the same in each round. A score of ten is given to pairs that needed two chatting rounds in each round to find a unanimous belief. Consequently, scores between zero and ten indicate that pairs needed the respective number of chats. A higher score implies more disagreement. Pairs were grouped into above-median and below-median score groups within each gender subgroup. Separate regressions were run for above-median and below-median score groups for each gender subgroup (six regressions). Beliefs which are within the first or last decile prior to receiving a signal are excluded (for a more detailed explanation, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

	FF		MM		MF	
	(1) disagree	(2) agree	(3) disagree	(4) agree	(5) disagree	(6) agree
$\delta$	0.9034** (0.0390)	0.8590** (0.0559)	1.0952* (0.0534)	1.1114 (0.0715)	0.9514 (0.0683)	1.1353*** (0.0432)
$\beta_H$	0.5784*** (0.0442)	0.2975*** (0.0532)	0.6548*** (0.0627)	0.4930*** (0.0936)	0.6982*** (0.0745)	0.4007*** (0.0413)
$\beta_L$	0.4022*** (0.0527)	0.3885*** (0.0516)	0.4969*** (0.0781)	0.2653*** (0.0957)	0.3707*** (0.0815)	0.2804*** (0.0535)
$P(\beta_H = \beta_L)$	0.0114	0.2251	0.1136	0.0693	0.0033	0.0736
Observations	160	244	221	158	156	174
R-squared	0.8359	0.6654	0.7802	0.7599	0.7058	0.8513
p-values	(1) vs (2)		(3) vs (4)		(5) vs (6)	
Diff. in asym.	0.0015		0.6048		0.0483	

TABLE A4.7 – Belief updates and dominance (all participants included).

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , with  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . We define someone to be dominant if he or she has a greater influence on the team decision during the first chatting encounter than the teammate. Specifically, we compare individual beliefs (belief C.1, see Figure 4.1) and how they differ from the resulting team belief (belief C.2 or C.3). A pair is defined to host *no dominant (not dom)* person if the team belief lies in the middle of the two individual beliefs. Analog, we define a pair to host a *dominant (dom)* person if the team belief does not lie in the middle but consequently closer to the individual belief of one of the two teammates. This teammate's belief dominated the belief of the other. Separate regressions are run for dominant (do not meet in the middle) and non-dominant pairs (do meet in the middle) for each gender subgroup (six regressions). Beliefs which are within the first or last decile prior to receiving a signal are excluded (for a more detailed explanation, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	FF		MM		MF	
	(1) dom	(2) not dom	(3) dom	(4) not dom	(5) dom	(6) not dom
$\delta$	0.8926*** (0.0389)	0.9088 (0.0704)	1.0691 (0.0550)	1.1200* (0.0658)	1.0863* (0.0510)	0.9176 (0.0684)
$\beta_H$	0.4440*** (0.0418)	0.3944*** (0.0697)	0.5805*** (0.0631)	0.6583*** (0.0943)	0.5991*** (0.0578)	0.4372*** (0.0565)
$\beta_L$	0.4829*** (0.0451)	0.2127*** (0.0650)	0.4194*** (0.0752)	0.3216*** (0.0981)	0.3848*** (0.0674)	0.2702*** (0.0704)
$P(\beta_H = \beta_L)$	0.5266	0.0653	0.0980	0.0082	0.0163	0.0637
Observations	272	132	276	103	194	136
R-squared	0.7652	0.6681	0.7175	0.8677	0.7932	0.6987
p-values	(1) vs (2)		(3) vs (4)		(5) vs (6)	
Diff. in asym.	0.0766		0.2560		0.6702	

TABLE A4.8 – Belief updates and dominance with a different proxy for dominance.

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , with  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . We proxy *dominance* using the variable rated impact which is elicited by asking participants to rate the statement “My word had a bigger impact on the joint decision than my teammate’s word.” on a 7-point Likert scale. We assign the higher of the two impact ratings within a pair as proxy for the presence of dominance within a group. Pairs were grouped into above-median and below-median impact rating groups within each gender subgroup. Separate regressions were run for above-median and below-median impact groups for each gender subgroup (six regressions). Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	FF		MM		MF	
	(1) dom	(2) not dom	(3) dom	(4) not dom	(5) dom	(6) not dom
$\delta$	0.9077 (0.0870)	0.9174** (0.0393)	1.1146* (0.0647)	1.1153* (0.0657)	1.0244 (0.0936)	1.0408 (0.0471)
$\beta_H$	0.5657*** (0.1036)	0.4882*** (0.0412)	0.6684*** (0.0799)	0.6704*** (0.0775)	0.7224*** (0.0786)	0.5033*** (0.0501)
$\beta_L$	0.5481*** (0.0970)	0.5081*** (0.0477)	0.3481*** (0.0916)	0.4761*** (0.0866)	0.2840** (0.1203)	0.4955*** (0.0539)
$P(\beta_H = \beta_L)$	0.9013	0.7521	0.0049	0.0848	0.0028	0.9154
Observations	86	176	142	156	104	148
R-squared	0.7041	0.8363	0.8392	0.8004	0.7299	0.8561
p-values	(1) vs (2)		(3) vs (4)		(5) vs (6)	
Diff. in asym.	0.7860		0.3773		0.0001	

TABLE A4.9 – Belief updates and dominance in gender-diverse pairs with a different proxy for dominance.

This table presents pooled OLS estimates for each gender subgroup separately of the following equation:  $\text{logit}(\hat{\mu}_t) = \delta * \text{logit}(\hat{\mu}_{t-1}) + \beta_h * I(x_{it} = h) * \lambda_h + \beta_l * I(x_{it} = l) * \lambda_l + \epsilon_{it}$ , with  $\text{logit}(z) = \ln(\frac{z}{1-z})$ . The dependent variable  $\mu^t$  is the posterior belief of each round,  $\delta$  shows whether the response to positive or negative information is invariant to the level of their prior belief,  $\beta_H$  indicates how strong pairs react to positive feedback and  $\beta_L$  how strong they react to negative feedback. Bayes' Law requires  $\beta_H = \beta_L = 1$ . We proxy *dominance* using the variable rated impact which is elicited by asking participants to rate the statement “My word had a bigger impact on the joint decision than my teammate’s word.” on a 7-point Likert scale. We assign the higher of the two impact ratings within a pair as proxy for the presence of dominance within a group. If the ratings among teammates are the same, we call it equal impact. Gender-diverse pairs are divided into higher male impact (column (1)), and higher female impact (2), higher male or female impact (3), and equal impact (4). Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered. Standard errors are in parentheses and clustered on an individual level. Asterisks indicate significance levels which test whether estimation coefficients are different from one; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	MF			
	(1) M dom	(2) F dom	(3) M or F dom	(4) No dom
$\delta$	0.8831 (0.0767)	1.1669 (0.1251)	1.0006 (0.0682)	1.0536 (0.0529)
$\beta_H$	0.7564*** (0.0740)	0.6150*** (0.1067)	0.6948*** (0.0648)	0.4573*** (0.0521)
$\beta_L$	0.4032*** (0.1033)	0.3508** (0.1482)	0.4314*** (0.0865)	0.4239*** (0.0574)
$P(\beta_H = \beta_L)$	0.0092	0.1301	0.0162	0.6649
Observations	76	78	154	98
R-squared	0.8182	0.7399	0.7652	0.8700
p-values	(1) vs (4)	(2) vs (4)	(3) vs (4)	(1) vs (2)
Diff. in asym.	0.0005	0.0905	0.0299	0.4885

## A.2 Figures

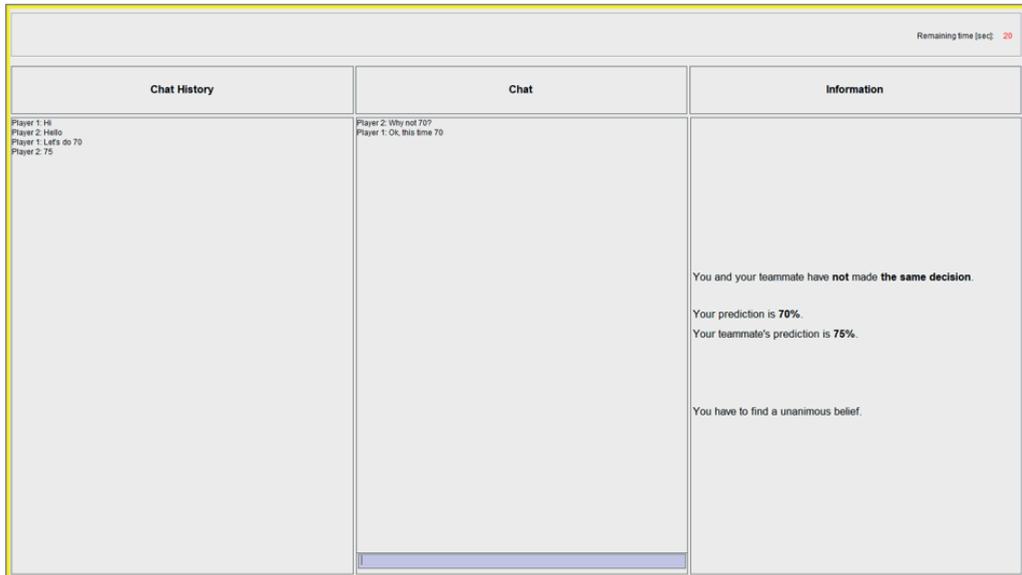


FIGURE A4.1 – Screenshot of the chat.

On the left, the chat history is shown in case a pair did not come up with a unanimous belief in the first chat. The chat history was cleared after a unanimous belief was found. In the middle, the current chat is visible including the blue entry field at the bottom. On the right, one's individual belief as well as the individual belief of the teammate is shown. Text and numbers are exemplary.

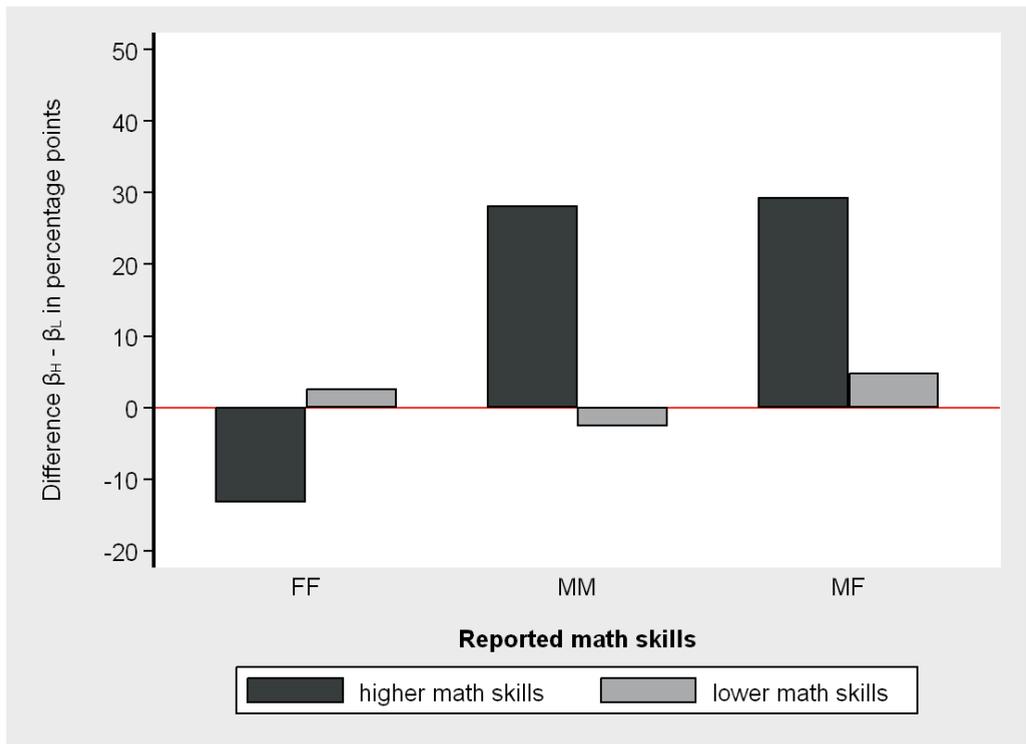


FIGURE A4.2 – Asymmetry and reported math skills.

This graph plots the difference between coefficients  $\beta_H$  and  $\beta_L$  from Table A4.4. Updating according to Bayes' Law requires all bars to be equal to zero. Reported math skills were elicited by asking subjects to answer the statement "My math skills are higher than those of \_\_\_ of my fellow students." on a 11-point Likert scale from 0% to 100% in steps of ten percentage points. We proxy the math skill of a pair by taking the average of the math skills of its members. Pairs were grouped into above-median and below-median skilled groups within each gender subgroup. Separate regressions were run for above-median and below-median skilled groups for each gender subgroup (six regressions). A positive difference  $\beta_H > \beta_L$  implies asymmetric belief updating where positive feedback is valued stronger than negative feedback. Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded. Beliefs which are within the first or last decile prior to receiving a signal are also excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered.

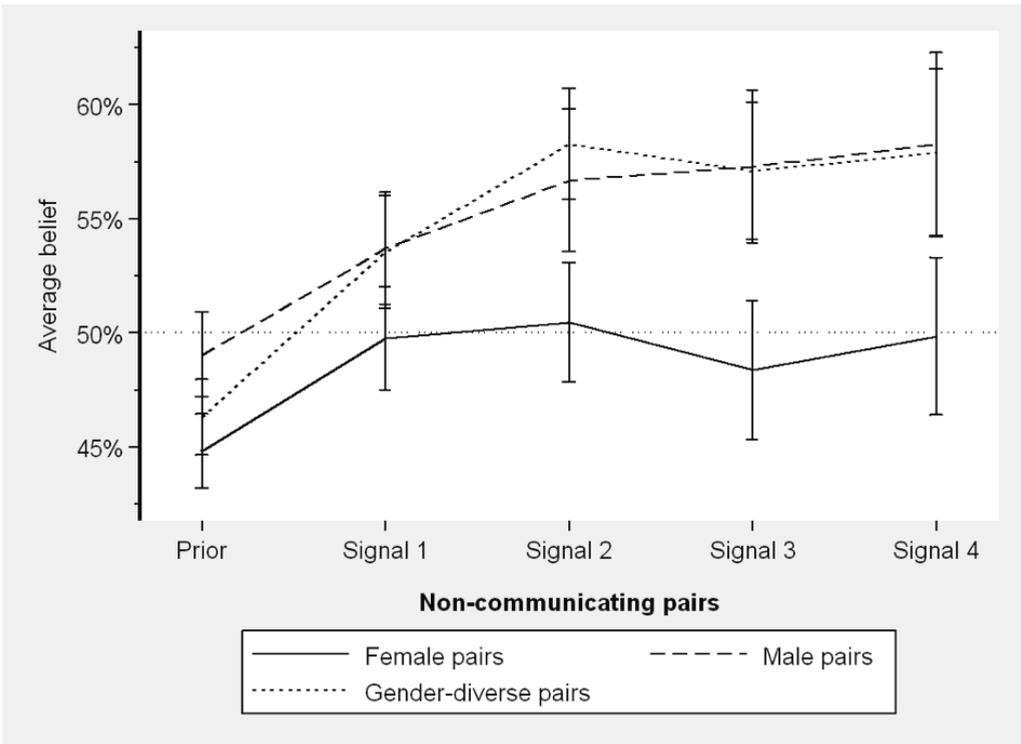


FIGURE A4.3 – Development of beliefs by round.

The graph plots the average beliefs of pairs that the team score is in the top half of all team scores for each gender subgroup separately for each of the five experimental rounds. Pairs which update contrary to a signal at least once or do not update at all throughout the four rounds are excluded (for a more detailed explanation of exclusions, see section 4.3.8). Only the final decision per round is considered.

### A.3 Experimental instructions

**Welcome!**

Nice to have you here - and thanks for participating.

---

This session will approximately take 90 minutes.

You will have to do several tasks. In each task, you are asked to make decisions which can be relevant for your financial payoff.

One of all the decisions you make during these tasks throughout the session will be randomly chosen to determine your actual payoff. So, it is always optimal for you to answer as accurately as possible.

The payment will be in cash at the conclusion of the session.

You will be paid privately and anonymously.

All payoffs are guaranteed by Maastricht University (Netherlands).

Please click "Continue" to proceed.

---

**Before we start, quickly tell us:**

1. Where do you currently study?
  - (a) Maastricht University
  - (b) Zuid University of Applied Sciences
  - (c) RWTH Aachen
  - (d) Other university
  
2. What is your gender?
  - (a) Male
  - (b) Female

3. How old are you?
- (a) Below 18
  - (b) 18-28
  - (c) 29-39
  - (d) 40-50

Please click "Continue" to proceed.

---

### Example

On the next few pages, we will go through a small experiment similar to our actual experiment.

The objective of this pre-experiment is to make you familiar with the choices you will have to make in the experiment to come.

It will be filled with a couple of examples to make life easier for you.

These examples will not be payoff-relevant but they will prepare and help you to earn more money in the actual experiment that will follow shortly.

Please click "Continue" to proceed.

---

### Example

The computer has just randomly drawn a number between 1 and 100 (each number has the same probability to be drawn).

The number assigned to you is \_\_\_

The computer has also randomly drawn a second number between 1 and 100. However, this number will not be shown to you.

On the next pages, we will tell you what to do. Click "Continue" to proceed.

---

**Example**

Remember the paper by Karni that was just shown to you and which proves that stating your true beliefs actually maximizes your expected payoffs?

In other words, **you maximize your expected earnings** when your submitted prediction equals your actual belief.

The details of the procedure are not crucial as long as you are aware that submitting a prediction that is not your actual belief will decrease your expected earnings. The mathematical proof is complex, but if you wish you can raise your hand and we come to your place and hand the paper to you. Otherwise, you can take a copy with you after the experiment.

Last but not least, we assure you that we are interested in your true beliefs. So for you as well as for us the best outcomes wait, if you state your true beliefs. Thanks!

Ready? Then click "Continue" to begin!

---

What is the probability that your number is higher than the second number randomly chosen by the computer?

**Your prediction (in %):** ---

Please click "Continue" to proceed.

---

### Reevaluation

We will now give you feedback on whether your number is larger or smaller than the second number. Specifically, we are going to give you a series of **four reports that say "lower" or "higher"**. Each report indicates whether your number is lower or higher than the second number but unfortunately they are not always correct.

Each of the reports you will get is **accurate with a 75% chance** and inaccurate with a 25% chance. In other words, if you get a report with the message "higher", there is a 75% chance that your number really is higher but a 25% chance that it is actually lower than the second number. Vice versa, if you get a report with the message "lower", there is a 75% chance that your number really is lower but a 25% chance that it is actually higher than the second number.

**You will receive your report on the next page.**

Please click "Continue" to proceed.

---

### Your first report:

**lower/higher**

On the next page, you will again be asked to revise your prediction.

Make sure that you again you specify your true belief

Please click "Continue" to proceed.

---

What is the probability that your number is higher than the second number randomly chosen by the computer?

**Your prediction (in %): ---**

Please click "Continue" to proceed.

---

*The participants received three more reports following the same procedure*

---

**Now follows the real experiment!**

This was the experiment with the two random numbers.

By now you should be acquainted with the general procedure of the experiment to come.

On the next page, the actual experiment starts.

Please click "Continue" to proceed.

---

**Welcome to the first experiment!**

Please read the instructions carefully. Each decision can be relevant for your financial payoffs. Payoffs are as described on the next pages.

In a little bit, you're going to do a **calculation task**.

The task has **25 calculations** and you have **5 minutes** to do as many calculations as possible.

Please click "Continue" to proceed.

---

In all 25 calculations, you will be asked to add up five 2-digit numbers. For example:

$$17 + 24 + 38 + 70 + 28 =$$

$$89 + 55 + 13 + 11 + 22 =$$

You find a scratch paper and a pen on your desk which you can use to solve the calculations. However, **only the result you enter into the input field will be counted**. This field can be reached by clicking in it with the mouse.

Your score will be the number of correctly solved calculation problems.

Please click "Continue" to proceed.

---

In addition, we will be hosting the **Top Half contest**.

In this contest, you try to end up in the **top half of all the scores in this calculation challenge**. In other words, try to correctly solve more calculations than 50% of all participants in this room.

As scores on the calculation tasks come in, our computer will rank all scores.

The computer will then be hosting the "Top Half contest", which gives out prizes for scores that rank in the top half.

Please click "Continue" to proceed.

---

### **Your decision in the Top Half contest**

1. We ask you to predict how likely it is that you end up in the top half of all participants. In other words, how likely do you think it is that you will have a score that is higher than at least half of the other scores in this classroom?
2. For example, if you think there is a 20% chance that you are going to be in the top half of all scores, your prediction should be 20%. Similarly, your prediction should be 65% if you think that the chance that you are going to be better than at least half of the participants in this room is 65%.

The higher the precision of your prediction is the higher is your chance of winning 10€ in the top half contest.

Please click "Continue" to proceed.

---

### **Your decision in the Top Half contest**

Remember: You maximize your expected earnings when your submitted prediction equals your actual belief that you end up in the top half.

Staying in the example stated above, if you believe that your true chance to end up in the top half is 65%, then deviating from this belief by stating more, e.g. 70%, or less, e.g. 64%, decreases your chance of being paid 10€. Only by stating your true belief of 65% are you most likely to win the 10€.

In sum, it is in our interest to learn your true beliefs and in your interest to communicate your true beliefs to maximize your earnings. This holds also true for all remaining beliefs that we will ask of you.

Do you have any further questions? Then please raise your hand and we will come to you.

Please click "Continue" to proceed.

---

### **Your Prediction**

What is the probability that you will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief A from Figure 4.1*

Please click "Continue" to proceed.

---

*Timer: 60 seconds*

### **Reasoning of your choice**

Please state the reasons why you came to this prediction.

---

**You have been matched to another participant in this room.**

The two of you form a team for the remainder of this experiment.

In a bit, both of you will individually complete the calculation task. But there are some modifications to what you already read.

You receive one score as a team which is the sum of your score and the score of your teammate. Thus, your team score equals the total number of correctly solved problems between the two of you.

You will get to see your team score directly after the calculation task.

---

The person that you have been matched with is a **18- to 28-year-old, male/female** student from **Maastricht University**.

---

Note: everybody in this room is now in a team of 2.

Please click "Continue" to proceed.

---

### **Your decision in the Top Half contest**

1. We ask you to predict how likely it is that **your team** ends up in the top half of all participants. In other words, how likely do you think is it that **your team** will have a score that is higher than at least half of the other team scores in this classroom?
2. For example, if you think there is a 20% chance that **your team** is going to be in the top half of all scores, your prediction should be 20%. Similarly, your prediction should be 65% if you think that the chance that **your team** is going to be better than at least half of the teams in this room is 65%.

Please click "Continue" to proceed.

---

### **Your Prediction**

What is the probability that you as a team will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief B from Figure 4.1*

Please click "Continue" to proceed.

---

*Timer: 60 seconds*

### **Reasoning of your choice**

Please state the reasons why you came to this prediction.

---

### Your possible payoffs

There is a 50% chance for either one of the two following possibilities to be chosen for your payment.

1. You can be paid according to your team performance. For each point you as a team score (number of correct answers), you receive **0.50€**. Thus, if your team scores 20 points, you earn  $20 * 0.50 = 10€$ . The payment for your teammate is the same.
2. You can be paid 10€ in the Top Half contest according to how precisely you can **estimate the likelihood of your team to end up in the top half**.

So, with a 50% chance you get paid according to how well your team did in the calculation task. And with a 50% chance you get paid according to one of the beliefs you state.

Please click "Continue" to proceed.

---

Before the calculation task starts, remember:

You will be given **5 minutes** to calculate the correct sum of 25 calculations.

You cannot use a calculator but you are welcome to make use of the provided **scratch paper**. Don't forget to **enter your answers into the input field** on the screen. We cannot take into account any results on your scratch paper.

### Good luck!

Please click "Continue" to proceed.

---

*Timer: 300 seconds*

### Calculations (See Appendix A.4)

Please click "Continue" to proceed.

---

### Calculation score information

Your team score is \_\_\_ **out of 50**.

Please click "Continue" to proceed.

---

### Chatting

1. In a moment, you again have to specify how likely it is that your team score is in the top half of all scores.
2. Once you have taken a decision by yourself, it will be **compared to the decision of your teammate**.
3. Directly afterwards, you and your teammate get to **chat** to each other.
4. You and your teammate have to find a **unanimous decision** regarding the likelihood that **you as a team** end up in the top half of all team scores.
5. In case the two of you do not share the same beliefs right away, you have up to 2 chat rounds to come to an agreement. After each chat round you can restate your beliefs.
6. The first chat round lasts **120 seconds**. In case you do not find a unanimous decision, a second chat round will open. This second chat round only lasts **60 seconds**.

The decisions you make are **relevant for your own** as well as **for your teammate's payoffs**.

Please click "Continue" to proceed.

---

### Your Prediction

What is the probability that you as a team will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief C.1 from Figure 4.1*

Please click "Continue" to proceed.

---

On the next page, you have the possibility to chat to your teammate in the chat box in the middle.

You and your teammate have to find a unanimous prediction on how likely you as a team end up in the top half of all scores.

You can revise your choices **two more times** (after each chat round) in case you do not find a unanimous decision right away. If you by then **still do not have a unanimous decision**, you will **not be paid** for this stage.

Please click "Continue" to proceed.

---

*Timer: 120 seconds*

**Chat Round 1** (see Figure A4.1)

---

*visible if Belief C.1 not unanimous*

You will be able to submit your revised predictions on the next page.

After that, you have **one more chat round remaining** to find a unanimous decision.

If you still **do not come to a unanimous decision**, you will **not be paid** for this decision.

Please click "Continue" to proceed.

---

*visible if Belief C.1 not unanimous*

What is the probability that you as a team will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief C.2 from Figure 4.1*

Please click "Continue" to proceed.

---

*visible if Belief C.2 not unanimous*

*Timer: 60 seconds*

**Chat Round 2** (see Figure A4.1)

---

*visible if Belief C.2 not unanimous*

What is the probability that you as a team will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief C.3 from Figure 4.1*

Please click "Continue" to proceed.

---

Your predictions after the chatting period:

You did not come to a unanimous decision. / You came to a unanimous decision.

Your prediction: ---

Your teammate's prediction: ---

Please click "Continue" to proceed.

---

### Reevaluation

Now that you have completed the calculation task and given us some information about how well you think you as a team did, we are going to give you some feedback on your team performance.

Specifically, we are going to give you a series of **four reports that say "top" or "bottom"**. The reports are meant to indicate whether your team performance was in the top or bottom half, but unfortunately they are not always accurate.

Each of the four reports you will get is **accurate with a 75% chance** and inaccurate with a 25% chance. In other words, if you get a report with the message "Top", there is a 75% chance that you really are in the top half but a 25% chance that you actually are in the bottom half. Vice versa, if you get a report with the message "Bottom", there is a 75% chance that you really are in the bottom half but a 25% chance that you actually are in the top half.

Please click "Continue" to proceed.

---

### Reevaluation

**You and your teammate get the same 4 reports.** Thus, if you receive a Top report, your teammate also receives a Top report; and if you receive a Bottom report, your teammate also receives a Bottom report.

**Signals are independent of each other.** Your second signal does not change just because your first signal was a Top or Bottom signal. All four signals are always accurate with a 75% chance.

**After each report we will ask you to reevaluate** the chances of your team to rank among the top half.

Exactly as before, you have to find a unanimous decision together with your teammate after each report.

Please click "Continue" to proceed.

---

Remember: The report you will see on the next page indicates whether you as a team are in the top half of all scores but is correct in only 75% of the cases.

Please reevaluate how likely it is that your team ends up in the top half of all participating teams.

Please click "Continue" to proceed.

---

**Your first team report** (both of you get the same report):

**Top/Bottom**

Please indicate on the next page **how likely** it is that your team score is in the top half of all scores.

Please click "Continue" to proceed.

---

**Reevaluation after the first report**  
**Your Prediction**

What is the probability that you as a team will end up in the top half of all scores in this calculation experiment?

**Your prediction (in %):** ---

*Belief D.1 from Figure 4.1*

Please click "Continue" to proceed.

---

*The pairs had to find a unanimous belief following the same procedure*

---

**Your second team report** (both of you get the same report):

**Top/Bottom**

The first report was: **Top/Bottom**

Please indicate on the next page **how likely** it is that your team score is in the top half of all scores.

Please click "Continue" to proceed.

---

*The pairs received two more reports following the same procedure*

---

## A.4 Calculations

Calculations were presented in this order. Participants could only answer one calculation at a time. Once they submitted an answer, they could not go back to a previous calculations.

1.  $66 + 95 + 85 + 50 + 82 =$
2.  $28 + 80 + 89 + 81 + 28 =$
3.  $87 + 56 + 23 + 26 + 77 =$
4.  $67 + 91 + 40 + 31 + 67 =$
5.  $13 + 10 + 63 + 45 + 79 =$
6.  $64 + 88 + 90 + 89 + 69 =$
7.  $25 + 59 + 57 + 54 + 82 =$
8.  $65 + 96 + 51 + 26 + 42 =$
9.  $33 + 57 + 87 + 51 + 98 =$
10.  $60 + 19 + 91 + 70 + 69 =$
11.  $84 + 38 + 24 + 83 + 88 =$
12.  $55 + 72 + 47 + 86 + 26 =$
13.  $81 + 38 + 55 + 34 + 47 =$
14.  $59 + 47 + 51 + 98 + 96 =$
15.  $89 + 22 + 27 + 23 + 84 =$
16.  $77 + 99 + 80 + 36 + 98 =$
17.  $80 + 21 + 46 + 43 + 43 =$
18.  $99 + 63 + 21 + 91 + 99 =$
19.  $91 + 28 + 86 + 46 + 76 =$
20.  $26 + 95 + 68 + 32 + 73 =$
21.  $96 + 77 + 90 + 86 + 50 =$
22.  $71 + 40 + 90 + 17 + 61 =$
23.  $65 + 60 + 54 + 46 + 81 =$
24.  $95 + 73 + 23 + 39 + 90 =$
25.  $89 + 35 + 94 + 65 + 19 =$



# Chapter 5

## Conclusion

The leading theme of this dissertation is the question of how to profit from group decisions. While individual decisions are important and deserve their role in academic research, many decisions depend on more than just one single decision-maker. I thereby built on the vast knowledge we have of how individuals behave, which biases they fall for and how these biases are affected or even how they can be erased, to understand how such personal traits and behavioral characteristics affect decisions in groups.

This book provides a **recipe with three key ingredients to successful group decision-making**. As is always the case, this recipe is anything but complete and offers ample space for future research. But similar to a recipe teaching how to bake chocolate cake, it is not intended to teach baking triple chocolate cookies. This recipe focuses on two main aspects of group decision-making: first, *decision-making on behalf* of a group (*Chapter 2*): consider a situation in which new laws are enacted on behalf of the people, or in which investment managers have to decide which stocks to buy for their trustors. And second, *decision-making within* a group (*Chapters 3 and 4*): take a situation where a couple decides on whether to buy a house, or where CEO and CFO decide on whether to acquire a new company.

So let us consider the suitable environment for this book to provide the best insights. Two investment managers are in charge of the assets of a large pension fund. They are the only ones responsible for investment decisions but both in equal parts—no one is more senior and they do not have an additional board to report to. Given their fiduciary duty towards their members, the decisions of these two managers have to reflect the preferences of these members. Yet, by construction, the decision is made between the two of them. What do they have to consider and which problems can they avoid?

In such a situation, this book provides a recipe which claims the following ingredients to be necessary for a scientifically valid and outcome-enhancing decision.

Ego-relevant decisions...

1. must have **real consequences**,
2. should be **made in pairs**, and
3. should be made by **at least one woman** (but only if no one dominates)—**better two**.

*Chapter 2* analyzes the necessity of real consequences in cases of representative decision-making based on surveys. Surveys offer the great advantage of being able to reach out to many people allowing the decision-makers to base their decisions on the preferences and beliefs of the majority. Yet, surveys are only as good as their incentives for respondents to answer truthfully. One way to assure truthful reporting is by granting respondents a consequential vote where they have to live with the consequences themselves.

The main question of this chapter is, whether people put their pension savings on the table to promote sustainability? To answer this question, I run a large-scale field experiment ( $n = 3,256$ ). The pension fund in our study gave its members a real vote for more or less sustainable investments. A comparison group made the same decision, but hypothetically. I find that 66.7% of the participants favor to invest their pension savings in a sustainable manner. This choice is driven by social preferences. As a result of these strong social preferences we find no difference between the real and hypothetical treatment groups. We rule out financial beliefs, confusion, or a lack of information as explanation. Institutional investors benefit from taking their clients' social preferences seriously, with consequences for asset prices and the fulfillment of the United Nations Sustainable Development Goals.

Chapters 3 and 4 take a different angle and look at how groups form joint group beliefs subject to the question of overconfident decision making. *Chapter 3* asks whether groups make better decisions if they can communicate. Specifically, this chapter studies whether overconfidence can be reduced through communication. I conduct an incentivized experiment which randomly assigns subjects to pairs. These form beliefs about their team performance in an ego-relevant task relative to competing pairs. Pairs in the communication treatment are allowed to chat and have to find a unanimous team belief. Pairs in the no communication control get to see the teammate's belief but communication is not allowed. Over a series of four noisy but informative signals about their relative performance, pairs can update their beliefs. Results show that, first, pairs react equally strong to positive and negative feedback regardless of whether communication is allowed. Thus, pairs update their beliefs symmetrically which implies that pairs do not form overconfident beliefs in the long run. Second, we find that dominant individuals impose their overconfidence on pairs. This harmful effect is reduced when pairs are not allowed to communicate, but it does not fully disappear. Third, pairs update their beliefs too conservatively, a bias independent of communication. Pairs are thus not overconfident in

---

assessing their performance but are still too slow in incorporating new information into their beliefs.

*Chapter 4* asks in how far knowing the gender of a teammate affects group decision-making. Correlational studies are undecided about the effects of gender diversity on decision quality. I study whether the gender composition of pairs has a causal influence on the formation of joint beliefs. I conduct an incentivized experiment which randomly assigns subjects to purely male, purely female, or gender-diverse pairs. Pairs jointly form beliefs about their performance in a stereotypically male-dominated, ego-relevant task relative to competing pairs. Then pairs receive a series of four noisy but informative signals about their relative performance after each of which they can update their beliefs. Our results show that gender-diverse pairs respond 42% stronger to positive than to negative feedback. Such asymmetric updating has a negative effect on expected profits and leads to overoptimistic beliefs in the long run. They do as bad as male and significantly worse than female pairs. Asymmetric updating in gender-diverse pairs is caused by dominant leaders. As soon as one teammate dictates the team belief, gender-diverse pairs update asymmetrically. Gender-diverse pairs can overcome asymmetric updating if their joint belief lies in the middle of their individual beliefs.

Coming back to our investment manager duo. For them, an ideal way to decide on which investment decision to take is to engage with their members and to learn about their preferences. This has to be done by giving them a real choice. Once they know the preferences of their members for a specific topic—for example sustainable investments—they will be able to decide on their behalf even for future decisions. Assume these members have strong preferences for sustainable investing similar to the respondents from Chapter 2 with the willingness to trade financial for social return. Then, the investment managers know that they can tweak their portfolio towards more sustainability without having to fear that a slightly less performing portfolio brings them in legal trouble or would cause their members to make the fur fly.

Nevertheless, once the two managers know what their members prefer, they themselves have to make the investment decisions. Here, the results of chapters 3 and 4 come into play. First, they should make those decisions jointly. Thereby, it is not so relevant whether they sit down and discuss their ideas or individually come up with their investment strategy. It is only important that the final decision takes both their ideas into account to avoid overconfident investment decision which eventually will lead to a lower financial performance. But how should "their decisions be taken into account"? This depends on the gender of the two managers. If the two are women, it does not matter. Neither of the two will bring overconfident individual beliefs to the table nor do they jointly come up with overconfident beliefs. It is more difficult, if the two are man and woman. Then, the two must value each idea, each belief equally—especially if they sit down together and engage in discussion. They have to make sure that neither of the two tries do impose the own belief on the team. If the two follow this approach, they can avoid

making overconfident investment decisions. Last and for once least, in case the two investment managers are men, chances are very low that their decisions do not suffer from overconfident beliefs. Instead, the two could consider outsourcing investment decisions or at least ask for outside advice. Even more promising might be to change the gender composition of the team.

# Chapter 6

## Research Impact

Policy advisors tend to rely on the explanatory power of hypothetical surveys. Take the European Commission which not only held public online consultations about whether to keep or abolish daylight saving time (see Chapter 1). Their approach to promote sustainable investing within the investment industry follows a similar pattern. A High-Level Expert Group (HLEG) instructed by the European Commission has recommended that “citizens with savings to invest should be empowered to invest in portfolios that reflect their sustainability and ethical preferences”<sup>1</sup>. But the question whether investors want to invest sustainably does not end on the individual level. In contrast, many third parties like institutional investors or pension funds invest on behalf of individuals who themselves have no say. How should these funds manage their fiduciary duty and decide on how sustainably to invest the savings of their investors? The European Commission states in its May 2018 Action Plan that “Investment firms... [should ask retail investors] about their preferences”. Further, the Commission will amend the Markets in Financial Instruments Directive (MiFID II) and the Insurance Distribution Directive to include ESG considerations into the advice that investment firms and insurance distributors offer to individual clients. While these intentions are clearly aimed at aligning investors’ preferences with the investment allocation of their funds, it will be highly relevant *how* the preferences are elicited. Given the possible difference between stated and revealed preferences, it is essential that this question is not simply asked hypothetically but in a way that entails actual consequences. Especially asset owners such as pension funds will be key to this process as they alone hold more than half of all assets under management (Willis Tower Watson, 2018)<sup>2</sup>. Any action to reveal the preferences of their participants could thus not only lead to a change in the sustainable investment policy of pension funds but also to a significant economic impact worldwide. It is of great importance that the elicitation of preferences to invest sustainably is

---

<sup>1</sup> See [https://ec.europa.eu/info/sites/info/files/180131-sustainable-finance-final-report\\_en.pdf](https://ec.europa.eu/info/sites/info/files/180131-sustainable-finance-final-report_en.pdf)

<sup>2</sup> According to an estimate by Willis Tower Watson (2018), pension funds worldwide hold USD 45 trillion assets under management.

done in a scientifically valid manner, especially because we still know very little about whether people are willing to put their pension savings on the table to promote sustainability.

The present study on the consequences of giving members of a pension fund a real choice on sustainable investing (see Chapter 2) is the first to provide insights into the possibly far reaching consequences of the May 2018 Action Plan. The strong preferences for sustainable investing in my sample show that, if executed correctly, the promotion of sustainability among individual as well as institutional investors has the potential to change the investment culture steering it towards including the social impact to a significantly larger extent. And even though my sample only consists of Dutch workers, my estimate for strong preferences of sustainable investments is likely to be conservative. Falk et al. (2018), whose question on social preferences I used in my survey, study the global variation of social preferences. They find that the social preferences of the Dutch population are 0.19 standard deviations below world average. At the same time, Germany (-0.05) and the UK (+0.03) are close to the world average while the US (+0.41) ranges clearly above. In other words, even though the Netherlands are one of the most sustainable investors in the world, they show below-average levels of social preferences. Given the even stronger social preferences in countries like the US, there is substantial growth potential for sustainable investments in these markets.

These findings have two important implications: First, policy advisors can set the legal framework to promote sustainability. Achieving the seventeen UN Sustainable Development Goals is in parts a question of unleashing the already existing but hidden power to go green among investors. And therefore, the European Commission should not only consider the option of making institutional investors ask their trustors about their preferences for sustainable investing. They should go the extra mile and require that this be done in a scientifically valid way—by giving real, consequential votes. My study shows that such votes do not require neither expert knowledge nor are they extremely costly or time-consuming. All they require is the will (or the legal pressure) to do it. The board of Pensioenfond Detailhandel, the pension fund I worked with for this study, did not only agree to an unprecedented study on the preferences of their members. More importantly, they now know what their members think about sustainable investing, they can even rest assured that the willingness to focus on social impact can even come at a certain cost.

Second, this study provides insights into other fields where hypothetical surveys or opinions deliver the basis for real actions. Consider the question of renewable energy and the opposition towards nuclear energy. While it is a hotly discussed topic filled with a mix of fear, ideology, and knowledge, it is difficult to retrieve the true preferences of people from within this mix. Or consider less politicized topics like the decision of a city council to spend a certain amount of money on a certain project. If such decisions are based on real preferences, on a real willingness to pay, politicians would know beforehand whether certain outcomes would be justifiable in the public or not. And while it should be the decision of the elected representatives, it should still

---

be a decision which represents the decisions of the people. This study tries to paint a feasible and reliable way of eliciting true preferences as the basis of representative decision-making—regardless of the topic.

My dissertation further contributes to the ongoing question of how to overcome overconfident decision-making. While this bias is largely and apparently invariably present among individuals, I show that pairs manage to form beliefs which are not unjustifiably optimistic. This is especially relevant in situations where overconfidence has a clearly negative impact on the decision maker as in the case in my study. A consequence could be that especially those situations where decisions are likely to be affected by overconfident individuals should be handed over to a team of two. This could affect the private trader who believes to be able to beat the market. He should seek the counsel of a second trader before submitting an order. Or consider a doctor who believes to be an outstanding expert in her field. She should have her patient get a second opinion from another doctor to verify whether her diagnosis was correct. It will be for future research to show in how far more natural pairs can overcome overconfident decision-making, for example pairs which know each other, possibly have worked together, and might even have a hierarchical difference in their say. Yet, my research strongly indicates that the road away from overconfidence leads to more group decisions.

Last but not least, I contribute to setting the question of gender diversity and its consequences on a scientific ground. While it is important to promote and uphold gender equality especially in a year which reminds of the introduction of the right to vote for women in many countries a century ago, it is equally important to understand how a more gender-diverse environment affects society. Studies which have examined the effect of gender-diversity in board rooms have been undecided about whether its consequences are positive or negative. The strength of my study is that I contribute causality to this debate. I show that there exists a causal link between the gender of team members and the tendency to form overconfident beliefs. As male pairs perform worst, this clearly shows the need of change in today's board rooms. Yet, my study also shows that merely establishing gender-diversity does not do the trick. Without any instructions, gender-diverse pairs form beliefs which are as overconfident as those of male pairs. And, while a simple rule of "Meet in the Middle" seems to cure this bias, my findings also suggest that female pairs do not have to even worry about forming overconfident beliefs.

The impact of research on gender-diversity could therefore lead to a world where people end up doing what they can do best. If we know that men perform poorly, it is only reasonable to have women do belief formation tasks which lead to a critical decision. And yes, gender-diversity is necessary as otherwise there would not be any women in the respective positions. But this research on gender equality should eventually help to find a ground where decisions can be made based on verifiable facts, be it performance, soft skills, or genetical differences. Our society can only profit from understanding itself and trying to establish a world which caters to the very best of each and every person.



# References

- ADAMS, R. B. and FERREIRA, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, **94** (2), 291–309.
- and FUNK, P. (2012). Beyond the glass ceiling: Does gender matter? *Management Science*, **58** (2), 219–235.
- AJZEN, I., BROWN, T. C. and CARVAJAL, F. (2004). Explaining the discrepancy between intentions and actions: The case of hypothetical bias in contingent valuation. *Personality and Social Psychology Bulletin*, **30** (9), 1108–1121.
- ALMÁS, I., CAPPELEN, A. W. and TUNGODDEN, B. (2016). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? *Working paper*.
- ALSERDA, G., DELLAERT, B., SWINKELS, L. and VAN DER LECQ, S. (2016). Pension risk preferences. *Netspar Industry Design Paper*, **62**.
- ANDERSON, C. and KILDUFF, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? the competence-signaling effects of trait dominance. *Journal of Personality and Social Psychology*, **96** (2), 491.
- ANDREONI, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, **100** (401), 464–477.
- , RAO, J. M. and TRACHTMAN, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy*, **125** (3), 625–653.
- ARIELY, D., BRACHA, A. and MEIER, S. (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review*, **99** (1), 544–55.
- ARROW, K. J. (2012). *Social choice and individual values*, vol. 12. Yale university press.
- ASCH, S. E. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, **70** (9), 1.
- BANDIERA, O., BARANKAY, I. and RASUL, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, **120** (3), 917–962.

## REFERENCES

---

- BARBER, B. M., MORSE, A. and YASUDA, A. (2018). Impact investing. *Working paper*.
- and ODEAN, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, **116** (1), 261–292.
- BARTLING, B., WEBER, R. A. and YAO, L. (2014). Do markets erode social responsibility? *The Quarterly Journal of Economics*, **130** (1), 219–266.
- BAUER, R. and SMEETS, P. (2015). Social identification and investment decisions. *Journal of Economic Behavior & Organization*, **117**, 121–134.
- BÉNABOU, R. and TIROLE, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, **117** (3), 871–915.
- and TIROLE, J. (2006). Incentives and prosocial behavior. *American Economic Review*, **96** (5), 1652–1678.
- BENOÎT, J.-P. and DUBRA, J. (2011). Apparent overconfidence. *Econometrica*, **79** (5), 1591–1625.
- BERNARDO, A. E. and WELCH, I. (2001). On the evolution of overconfidence and entrepreneurs. *Journal of Economics & Management Strategy*, **10** (3), 301–330.
- BESHEARS, J., CHOI, J., CLAYTON, C., HARRIS, C., LAIBSON, D. and MADRIAN, B. (2014). Optimal illiquidity. *Paper for the Retirement Research Consortium, September*, **26**.
- , CHOI, J. J., LAIBSON, D. and MADRIAN, B. C. (2008). How are preferences revealed? *Journal of Public Economics*, **92** (8-9), 1787–1794.
- BEYER, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, **59** (5), 960.
- (1998). Gender differences in self-perception and negative recall biases. *Sex Roles*, **38** (1-2), 103–133.
- and BOWDEN, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, **23** (2), 157–172.
- BIAIS, B., HILTON, D., MAZURIER, K. and POUGET, S. (2005). Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. *The Review of Economic Studies*, **72** (2), 287–312.
- BIALKOWSKI, J. and STARKS, L. T. (2016). Sri funds: Investor demand, exogenous shocks and esg profiles. *Working paper*.
- BLISS, R. T., POTTER, M. E. and SCHWARZ, C. (2008). Performance characteristics of individually-managed versus team-managed mutual funds. *Journal of Portfolio Management*, **34** (3), 110.
- BOLLEN, N. P. (2007). Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis*, **42** (3), 683–708.

- 
- BOLTON, G. E. and KATOK, E. (1995). An experimental test for gender differences in beneficent behavior. *Economics Letters*, **48** (3-4), 287–292.
- and OCKENFELS, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, **90** (1), 166–193.
- BOND, R. and SMITH, P. B. (1996). Culture and conformity: A meta-analysis of studies using asch's (1952b, 1956) line judgment task. *Psychological Bulletin*, **119** (1), 111.
- BONNER, B. L., BAUMANN, M. R. and DALAL, R. S. (2002). The effects of member expertise on group decision-making and performance. *Organizational Behavior and Human Decision Processes*, **88** (2), 719–736.
- , SILLITO, S. D. and BAUMANN, M. R. (2007). Collective estimation: Accuracy, expertise, and extroversion as sources of intra-group influence. *Organizational Behavior and Human Decision Processes*, **103** (1), 121–133.
- BORDALO, P., COFFMAN, K. B., GENNAIOLI, N. and SHLEIFER, A. (forthcoming). Beliefs about gender. *American Economic Review*.
- BRADY, P. and BOWD, A. (2005). Mathematics anxiety, prior experience and confidence to teach mathematics among pre-service education students. *Teachers and Teaching*, **11** (1), 37–46.
- BROWN, R. and SARMA, N. (2007). Ceo overconfidence, ceo dominance and corporate acquisitions. *Journal of Economics and Business*, **59** (5), 358–379.
- BULLOCK, W., IMAI, K. and SHAPIRO, J. N. (2011). Statistical analysis of endorsement experiments: Measuring support for militant groups in pakistan. *Political Analysis*, **19** (4), 363–384.
- BURKS, S. V., CARPENTER, J. P., GOETTE, L. and RUSTICHINI, A. (2013). Overconfidence and social signalling. *Review of Economic Studies*, **80** (3), 949–983.
- BUSER, T., GERHARDS, L. and VAN DER WEELE, J. J. (2016). Measuring responsiveness to feedback as a personal trait. *Working paper*.
- CACAULT, M. P. and GRIEDER, M. (2018). How group identification distorts beliefs. *Working paper*.
- CAMERER, C. and LOVALLO, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, **89** (1), 306–318.
- CAMPBELL, K. and MÍNGUEZ-VERA, A. (2008). Gender diversity in the boardroom and firm financial performance. *Journal of Business Ethics*, **83** (3), 435–451.
- CAPPELEN, A. W., HALVORSEN, T., SØRENSEN, E. Ø. and TUNGODDEN, B. (2017). Face-saving or fair-minded: What motivates moral behavior? *Journal of the European Economic Association*, **15** (3), 540–557.

## REFERENCES

---

- CARROLL, G. D., CHOI, J. J., LAIBSON, D., MADRIAN, B. C. and METRICK, A. (2009). Optimal defaults and active decisions. *The Quarterly Journal of Economics*, **124** (4), 1639–1674.
- CARSON, R., GROVES, T., LIST, J. and MACHINA, M. (2006). Probabilistic influence and supplemental benefits: a field test of the two key assumptions behind using stated preferences. *Unpublished manuscript*, **2**.
- CARSON, R. T. and GROVES, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, **37** (1), 181–210.
- CATSAMBIS, S. (1994). The path to math: Gender and racial-ethnic differences in mathematics participation from middle school to high school. *Sociology of Education*, pp. 199–215.
- CHAPPLE, L. and HUMPHREY, J. E. (2014). Does board gender diversity have a financial impact? evidence using stock portfolio performance. *Journal of Business Ethics*, **122** (4), 709–723.
- CHARNESS, G. and GNEEZY, U. (2010). Portfolio choice and risk attitudes: An experiment. *Economic Inquiry*, **48** (1), 133–146.
- and RABIN, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, **117** (3), 817–869.
- COFFMAN, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, **129** (4), 1625–1660.
- COHN, A., JESSEN, L., KLASJNA, M. and SMEETS, P. (2018). How do the rich think about redistribution? *Working paper*.
- COUTTS, A. (2018). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, pp. 1–27.
- CROSON, R. and GNEEZY, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, **47** (2), 448–74.
- CUMMINGS, R. G., ELLIOTT, S., HARRISON, G. W. and MURPHY, J. (1997). Are hypothetical referenda incentive compatible? *Journal of Political Economy*, **105** (3), 609–621.
- , HARRISON, G. W. and RUTSTROM, E. E. (1995). Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *American Economic Review*, **85** (1), 260.
- and TAYLOR, L. O. (1998). Does realism matter in contingent valuation surveys? *Land Economics*, pp. 203–215.
- and — (1999). Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *American Economic Review*, **89** (3), 649–665.
- CVENCEK, D., MELTZOFF, A. N. and GREENWALD, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development*, **82** (3), 766–779.

- DALE-OLSEN, H., SCHØNE, P. and VERNER, M. (2013). Diversity among norwegian boards of directors: Does a quota for women improve firm performance? *Feminist Economics*, **19** (4), 110–135.
- DE-MAGISTRIS, T., GRACIA, A. and NAYGA JR, R. M. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *American Journal of Agricultural Economics*, **95** (5), 1136–1154.
- DELLAVIGNA, S., LIST, J. A. and MALMENDIER, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, **127** (1), 1–56.
- DERWALL, J., GUENSTER, N., BAUER, R. and KOEDIJK, K. (2005). The eco-efficiency premium puzzle. *Financial Analysts Journal*, pp. 51–63.
- DOHMEN, T. and FALK, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, **101** (2), 556–90.
- DONKERS, B., LOURENÇO, C. and DELLAERT, B. G. (2012). Measuring and debiasing consumer pension risk attitudes. *Netspar panel paper*, **28**.
- DOWNES, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, **65** (2), 135–150.
- DUFWENBERG, M., HEIDHUES, P., KIRCHSTEIGER, G., RIEDEL, F. and SOBEL, J. (2011). Other-regarding preferences in general equilibrium. *The Review of Economic Studies*, **78** (2), 613–639.
- DYCK, A., LINS, K. V., ROTH, L. and WAGNER, H. F. (2018). Do institutional investors drive corporate social responsibility? international evidence. *Journal of Financial Economics*.
- ECCLES, J. (1993). Psychological and social barriers to women’s participation in mathematics and science. In *Biennial Meeting of the American Educational Research Association, Atlanta, GA*.
- ECKEL, C. C. and GROSSMAN, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal*, **108** (448), 726–735.
- EDMANS, A. (2011). Does the stock market fully value intangibles? employee satisfaction and equity prices. *Journal of Financial Economics*, **101** (3), 621–640.
- EDWARDS, W. (1968). Conservatism in human information processing. *Formal Representation of Human Judgment*.
- EIL, D. and RAO, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, **3** (2), 114–38.
- ELLINGSEN, T. and JOHANNESSON, M. (2008). Pride and prejudice: The human side of incentive theory. *American Economic Review*, **98** (3), 990–1008.
- ELLSBERG, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, pp. 643–669.

## REFERENCES

---

- ERIKSEN, C. W. and HAKE, H. W. (1957). Anchor effects in absolute judgments. *Journal of Experimental Psychology*, **53** (2), 132.
- ERTAC, S. (2011). Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, **80** (3), 532–545.
- ESTES, R. and HOSSEINI, J. (1988). The gender gap on wall street: an empirical analysis of confidence in investment decision making. *The Journal of Psychology*, **122** (6), 577–590.
- EUROSIF (2016). European sri study 2016.
- FALK, A., BECKER, A., DOHMEN, T., HUFFMAN, D. and SUNDE, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Working paper*.
- and SZECH, N. (2013). Morals and markets. *Science*, **340** (6133), 707–711.
- FEHR, E. and FISCHBACHER, U. (2002). Why social preferences matter—the impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal*, **112** (478), C1–C33.
- and GÄCHTER, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, **14** (3), 159–181.
- and SCHMIDT, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, **114** (3), 817–868.
- FELDMAN, J., MIYAMOTO, J. and LOFTUS, E. F. (1999). Are actions regretted more than inactions? *Organizational Behavior and Human Decision Processes*, **78** (3), 232–255.
- FELDMANHALL, O., MOBBS, D., EVANS, D., HISCOX, L., NAVRADY, L. and DALGLEISH, T. (2012). What we say and what we do: the relationship between real and hypothetical moral choices. *Cognition*, **123** (3), 434–441.
- FIGORE, G. (1999). Math-abused students: are we prepared to teach them? *The Mathematics Teacher*, **92** (5), 403–406.
- FISCHBACHER, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, **10** (2), 171–178.
- FISMAN, R., JAKIELA, P. and KARIV, S. (2017). Distributional preferences and political behavior. *Journal of Public Economics*, **155**, 1–10.
- FOURGE, D., KRIECHEL, B. and DOHMEN, T. (2014). Occupational sorting of school graduates: The role of economic preferences. *Journal of Economic Behavior & Organization*, **106**, 335–351.
- FOWLER III, A. R. and CLOSE, A. G. (2012). It ain't easy being green: Macro, meso, and micro green advertising agendas. *Journal of Advertising*, **41** (4), 119–132.

- FREY, B. S. and MEIER, S. (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review*, **94** (5), 1717–1722.
- FROME, P. M. and ECCLES, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of personality and social psychology*, **74** (2), 435.
- GERVAIS, S., HEATON, J. B. and ODEAN, T. (2011). Overconfidence, compensation contracts, and capital budgeting. *The Journal of Finance*, **66** (5), 1735–1777.
- GINGERICH, D. W. (2010). Understanding off-the-books politics: Conducting inference on the determinants of sensitive behavior with randomized response surveys. *Political Analysis*, **18** (3), 349–380.
- GLAZER, A. and KONRAD, K. A. (1996). A signaling explanation for charity. *The American Economic Review*, **86** (4), 1019–1028.
- GNEEZY, A., GNEEZY, U., NELSON, L. D. and BROWN, A. (2010). Shared social responsibility: A field experiment in pay-what-you-want pricing and charitable giving. *Science*, **329** (5989), 325–327.
- GNEEZY, U. and LIST, J. A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica*, **74** (5), 1365–1384.
- GODA, G. S., LEVY, M. R., MANCHESTER, C. F., SOJOURNER, A. and TASOFF, J. (2015). *The role of time preferences and exponential-growth bias in retirement savings*. Tech. rep., National Bureau of Economic Research.
- GRACIA, A., LOUREIRO, M. L. and NAYGA JR, R. M. (2011). Are valuations from nonhypothetical choice experiments different from those of experimental auctions? *American Journal of Agricultural Economics*, **93** (5), 1358–1373.
- GRIMM, P. (2010). Social desirability bias. *Wiley International Encyclopedia of Marketing*.
- GUISSO, L., MONTE, F., SAPIENZA, P. and ZINGALES, L. (2008). Culture, gender, and math. *Science*, **320** (5880), 1164.
- , SAPIENZA, P. and ZINGALES, L. (2006). Does culture affect economic outcomes? *Journal of Economic perspectives*, **20** (2), 23–48.
- GÜTH, W., SCHMIDT, C. and SUTTER, M. (2007). Bargaining outside the lab—a newspaper experiment of a three-person ultimatum game. *The Economic Journal*, **117** (518), 449–469.
- HALEBLIAN, J. and FINKELSTEIN, S. (1993). Top management team size, ceo dominance, and firm performance: The moderating roles of environmental turbulence and discretion. *Academy of Management Journal*, **36** (4), 844–863.
- HARRISON, G. W. (2006a). Experimental evidence on alternative environmental valuation methods. *Environmental and Resource Economics*, **34** (1), 125–162.
- (2006b). Making choice studies incentive compatible. In *Valuing environmental amenities using stated choice studies*, Springer, pp. 67–110.

## REFERENCES

---

- , HARSTAD, R. M. and RUTSTRÖM, E. E. (2004). Experimental methods and elicitation of values. *Experimental Economics*, **7** (2), 123–140.
- and RUTSTRÖM, E. E. (2008). Experimental evidence on the existence of hypothetical bias in value elicitation methods. *Handbook of Experimental Economics Results*, **1**, 752–767.
- HARTZMARK, S. M. and SUSSMAN, A. B. (2018). Do investors value sustainability? a natural experiment examining ranking and fund flows. *Working paper*.
- HEALY, A. and PATE, J. G. (2007). Overconfidence, social groups, and gender: Evidence from the lab and field. *Working paper*.
- HILL, G. W. (1982). Group versus individual performance: Are  $n+1$  heads better than one? *Psychological bulletin*, **91** (3), 517.
- HINSZ, V. B. (1990). Cognitive and consensus processes in group recognition memory performance. *Journal of Personality and Social Psychology*, **59** (4), 705.
- HIRSHLEIFER, D. and LUO, G. Y. (2001). On the survival of overconfident traders in a competitive securities market. *Journal of Financial Markets*, **4** (1), 73–84.
- HOELZL, E. and RUSTICHINI, A. (2005). Overconfident: Do you put your money on it? *The Economic Journal*, **115** (503), 305–318.
- HONG, H. and KACPERCZYK, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, **93** (1), 15–36.
- and KOSTOVETSKY, L. (2012). Red and blue investing: Values and finance. *Journal of Financial Economics*, **103** (1), 1–19.
- HUANG, J. and KISGEN, D. J. (2013). Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of Financial Economics*, **108** (3), 822–839.
- HUCK, S. and WEIZSÄCKER, G. (2002). Do players correctly estimate what others do?: Evidence of conservatism in beliefs. *Journal of Economic Behavior & Organization*, **47** (1), 71–85.
- HYDE, J. S., FENNEMA, E. and LAMON, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin*, **107** (2), 139.
- , LINDBERG, S. M., LINN, M. C., ELLIS, A. B. and WILLIAMS, C. C. (2008). Gender similarities characterize math performance. *Science*, **321** (5888), 494–495.
- JOECKS, J., PULL, K. and VETTER, K. (2013). Gender diversity in the boardroom and firm performance: What exactly constitutes a “critical mass?”. *Journal of Business Ethics*, **118** (1), 61–72.
- JOHNSON, D. D. and FOWLER, J. H. (2011). The evolution of overconfidence. *Nature*, **477** (7364), 317.

- 
- , McDERMOTT, R., BARRETT, E. S., COWDEN, J., WRANGHAM, R., MCINTYRE, M. H. and ROSEN, S. P. (2006). Overconfidence in wargames: experimental evidence on expectations, aggression, gender and testosterone. *Proceedings of the Royal Society of London B: Biological Sciences*, **273** (1600), 2513–2520.
- JP MORGAN (2018). Sustainable investing is moving mainstream.
- JULIZAERMA, M. K. and SORI, Z. M. (2012). Gender diversity in the boardroom and firm performance of malaysian public listed companies. *Procedia-Social and Behavioral Sciences*, **65**, 1077–1085.
- KAHNEMAN, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, **51** (2), 296–312.
- , KNETSCH, J. L. and THALER, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, **5** (1), 193–206.
- KARNI, E. (2009). A mechanism for eliciting probabilities. *Econometrica*, **77** (2), 603–606.
- KENNY, D. and LAVOIE, L. (1985). Interpersonal relations and group processes. *J. Personality Soc. Psych*, **48** (2), 339–348.
- KERSCHBAMER, R. and MÜLLER, D. (2017). *Social preferences and political attitudes: An online experiment on a large heterogeneous sample*. Tech. rep.
- KESSLER, J. B., MILKMAN, K. L. and ZHANG, C. Y. (2018). Getting the rich and powerful to give. *Working paper*.
- KIM, J.-B., WANG, Z. and ZHANG, L. (2016). Ceo overconfidence and stock price crash risk. *Contemporary Accounting Research*, **33** (4), 1720–1749.
- KLINK, J. and LANGEN, N. (2015). Are animal welfare aspects of relevance in consumers' purchase decision? *Proceedings in Food System Dynamics*, pp. 328–346.
- KOLLMUSS, A. and AGYEMAN, J. (2002). Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, **8** (3), 239–260.
- KRÜGER, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics*, **115** (2), 304–329.
- KRÜGER, P., SAUTNER, Z. and STARKS, L. T. (2018). The importance of climate risks for institutional investors. *Working paper*.
- LIST, J. A. (2001). Do explicit warnings eliminate the hypothetical bias in elicitation procedures? evidence from field auctions for sports cards. *American Economic Review*, **91** (5), 1498–1507.
- (2006). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy*, **114** (1), 1–37.

## REFERENCES

---

- and GALLET, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics*, **20** (3), 241–254.
- and SHOGREN, J. F. (1998a). Calibration of the difference between actual and hypothetical valuations in a field experiment. *Journal of Economic Behavior & Organization*, **37** (2), 193–205.
- and — (1998b). The deadweight loss of christmas: comment. *The American Economic Review*, **88** (5), 1350–1355.
- and — (2002). Calibration of willingness-to-accept. *Journal of Environmental Economics and Management*, **43** (2), 219–233.
- LIU, Y., WEI, Z. and XIE, F. (2014). Do women directors improve firm performance in china? *Journal of Corporate Finance*, **28**, 169–184.
- LOOMIS, J. B. (2014). 2013 waea keynote address: Strategies for overcoming hypothetical bias in stated preference surveys. *Journal of Agricultural and Resource Economics*, pp. 34–46.
- LUNDEBERG, M. A., FOX, P. W. and PUNČOHAŘ, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of educational psychology*, **86** (1), 114.
- LUSK, J. L. and NORWOOD, F. B. (2009). Bridging the gap between laboratory experiments and naturally occurring markets: an inferred valuation method. *Journal of Environmental Economics and Management*, **58** (2), 236–250.
- MALMENDIER, U. and TATE, G. (2005). Ceo overconfidence and corporate investment. *The Journal of Finance*, **60** (6), 2661–2700.
- and — (2008). Who makes acquisitions? ceo overconfidence and the market's reaction. *Journal of Financial Economics*, **89** (1), 20–43.
- MAREL, G. M., LYON, P. M., FIELD, M. J., BARNESLEY, L., HIBBERT, E. and PARISE, A. (2000). Clinical skills in early postgraduate medical trainees: patterns of acquisition of confidence and experience among junior doctors in a university teaching hospital. *Medical Education*, **34** (12), 1013–1015.
- MARINOVA, J., PLANTENGA, J. and REMERY, C. (2010). Gender diversity and firm performance: Evidence from dutch and danish boardrooms. *Discussion Paper Series/Tjalling C. Koopmans Research Institute*, **10** (03).
- MCKINSEY (2017). From 'why' to 'why not': Sustainable investing as the new normal.
- MISHRA, K. and METILDA, M. J. (2015). A study on the impact of investment experience, gender, and level of education on overconfidence and self-attribution bias. *IIMB Management Review*, **27** (4), 228–239.
- MÖBIUS, M. M., NIEDERLE, M., NIEHAUS, P. and ROSENBLAT, T. (2007). Gender differences in incorporating performance feedback. *Working paper*.

- , —, — and ROSENBLAT, T. (2014). Managing self-confidence: Theory and experimental evidence. *NBER working paper*.
- MOORE, D. A. and HEALY, P. J. (2008). The trouble with overconfidence. *Psychological Review*, **115** (2), 502.
- MUKERJI, S. and TALLON, J.-M. (2001). Ambiguity aversion and incompleteness of financial markets. *The Review of Economic Studies*, **68** (4), 883–904.
- MULLIGAN, C. B. and HUNTER, C. G. (2003). The empirical frequency of a pivotal vote. *Public Choice*, **116** (1-2), 31–54.
- MURPHY, J. J., ALLEN, P. G., STEVENS, T. H. and WEATHERHEAD, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, **30** (3), 313–325.
- NICOLLE, A., FLEMING, S. M., BACH, D. R., DRIVER, J. and DOLAN, R. J. (2011). A regret-induced status quo bias. *Journal of Neuroscience*, **31** (9), 3320–3327.
- NIEDERLE, M. and VESTERLUND, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, **122** (3), 1067–1101.
- NORWOOD, F. B. and LUSK, J. L. (2011). Social desirability bias in real, hypothetical, and inferred valuation experiments. *American Journal of Agricultural Economics*, **93** (2), 528–534.
- O'DONOGHUE, T. and RABIN, M. (1998). Procrastination in preparing for retirement. *University of California-Berkeley Working Paper*.
- PANDE, R. and FORD, D. (2012). Gender quotas and female leadership. *Working paper*.
- PRATTO, F., STALLWORTH, L. M. and SIDANIUS, J. (1997). The gender gap: Differences in political attitudes and social dominance orientation. *British Journal of Social Psychology*, **36** (1), 49–68.
- PULFORD, B. D. and COLMAN, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and Individual Differences*, **23** (1), 125–133.
- PUNCOCHAR, J. M. and FOX, P. W. (2004). Confidence in individual and group decision making: When “two heads” are worse than one. *Journal of Educational Psychology*, **96** (3), 582.
- PWC (2017). Pwc director survey 2017.
- RENNEBOOG, L., TER HORST, J. and ZHANG, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, **32** (9), 1723–1742.
- REUBEN, E., SAPIENZA, P. and ZINGALES, L. (2014). How stereotypes impair women’s careers in science. *Proceedings of the National Academy of Sciences*, p. 201314788.
- RIEDL, A. and SMEETS, P. (2017). Why do investors hold socially responsible mutual funds? *The Journal of Finance*, **72** (6), 2505–2550.

## REFERENCES

---

- RING, P., NEYSE, L., DAVID-BARETT, T. and SCHMIDT, U. (2016). Gender differences in performance predictions: evidence from the cognitive reflection test. *Frontiers in Psychology*, **7**, 1680.
- RUOF, T. and SMEETS, P. (2018). Let's meet in the middle! how gender-diversity can overcome asymmetric belief updating. *Working paper*.
- , SMEETS, P. M. and VAN VLIET, P. (2018). Dare to pair. pairs update beliefs symmetrically with and without communication. *Working paper*.
- SAMUELSON, W. and ZECKHAUSER, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, **1** (1), 7–59.
- SIDANIUS, J., LEVIN, S., LIU, J. and PRATTO, F. (2000). Social dominance orientation, anti-egalitarianism and the political psychology of gender: an extension and cross-cultural replication. *European Journal of Social Psychology*, **30** (1), 41–67.
- , PRATTO, F. and BRIEF, D. (1995). Group dominance and the political psychology of gender: A cross-cultural comparison. *Political Psychology*, pp. 381–396.
- , SINCLAIR, S. and PRATTO, F. (2006). Social dominance orientation, gender, and increasing educational exposure 1. *Journal of Applied Social Psychology*, **36** (7), 1640–1653.
- SNEZEK, J. A. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes*, **52** (1), 124–155.
- and HENRY, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, **43** (1), 1–28.
- and — (1990). Revision, weighting, and commitment in consensus group judgment. *Organizational Behavior and Human Decision Processes*, **45** (1), 66–84.
- SOBEL, J. (2015). Do markets make people look selfish? *Working paper*, **43** (2), 392–436.
- TERJESEN, S., COUTO, E. B. and FRANCISCO, P. M. (2016). Does the presence of independent and female directors impact firm performance? a multi-country study of board diversity. *Journal of Management & Governance*, **20** (3), 447–483.
- TERLAU, W. and HIRSCH, D. (2015). Sustainable consumption and the attitude-behaviour-gap phenomenon-causes and measurements towards a sustainable development. *International Journal on Food System Dynamics*, **6** (3), 159–174.
- TINGLEY, D. H. and WANG, S. W. (2010). Belief updating in sequential games of two-sided incomplete information: An experimental study of a crisis bargaining model. *Quarterly Journal of Political Science*, **5** (3), 243–255.
- TONIN, M. and VLASSOPOULOS, M. (2013). Experimental evidence of self-image concerns as motivation for giving. *Journal of Economic Behavior & Organization*, **90**, 19–27.
- TVERSKY, A. and KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185** (4157), 1124–1131.

- 
- and — (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, **106** (4), 1039–1061.
- VERMEIR, I. and VERBEKE, W. (2006). Sustainable food consumption: Exploring the consumer “attitude–behavioral intention” gap. *Journal of Agricultural and Environmental ethics*, **19** (2), 169–194.
- VLEK, C. A. (1965). *The use of probabilistic information in decision making*. University, Psychological Institute.
- and VAN DER HEIJDEN, L. H. (1967). *Subjective likelihood functions and variations in the accuracy of probabilistic information processing*. University of Leiden, psychological institute.
- WHEELER, G. and BEACH, L. R. (1968). Subjective sampling distributions and conservatism. *Working paper*.
- WHYTE, G., SEBENIUS, J. K. *et al.* (1997). The effect of multiple anchors on anchoring in individual and group judgment. *Organizational Behavior and Human Decision Processes*, **69** (1), 74–85.
- WILLIS TOWER WATSON (2018). Global pension assets study 2018.
- ZELLNER, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, **57** (298), 348–368.



# Biography

Tobias Ruof was born on September 26, 1986 in Weifenburg i. Bayern, Germany. In 2012, he obtained a B.Sc. in Economics from Bonn University. He continued his studies at Maastricht University, where he completed his M.Sc. in Financial Economics with distinction. Tobias joined the Finance Department of Maastricht University as a Ph.D. student in September 2014 under the supervision of Rob Bauer and Paul Smeets. He taught courses on Institutional Investors, Behavioral Finance, and Market-Based Assets and Shareholder Value. Tobias research interests include behavioral finance, investment allocation decisions, and sustainable investing. He presented his work at numerous academic conferences and most notably at the European Commission. For his work on sustainable investing (Chapter 2), he received the INQUIRE Europe Award 2019.



