THREE ESSAYS ON POWER AND DIVERSITY

De Saint Priest Oriana

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE COMPORTEMENT ORGANISATIONNEL

THREE ESSAYS ON POWER AND DIVERSITY

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences Économiques, mention « Management »

par

Oriana de SAINT PRIEST

Directeur de thèse
Prof. Christian Zehnder

Co-directrice de thèse
Prof. Franciska Krings

Jury

Prof. Marianne Schmid Mast, Présidente
Prof. Rafael Lalive, Expert interne
Prof. Eva Derous, Experte externe

LAUSANNE
2022

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE COMPORTEMENT ORGANISATIONNEL

THREE ESSAYS ON POWER AND DIVERSITY

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences Économiques, mention « Management »

par

Oriana de SAINT PRIEST

Directeur de thèse
Prof. Christian Zehnder

Co-directrice de thèse
Prof. Franciska Krings

Jury

Prof. Marianne Schmid Mast, Présidente
Prof. Rafael Lalive, Expert interne
Prof. Eva Derous, Experte externe

LAUSANNE
2022

UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

I M P R I M A T U R

Sans se prononcer sur les opinions de l'autrice, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Madame Oriana de SAINT PRIEST, titulaire d'un bachelor en Mathématiques et Philosophie de King's College London, d'un master en Finance et Economie de l'Université de Cambridge, et d'un master en Management de la London Business School, en vue de l'obtention du grade de docteure ès Sciences économiques, mention « management ».

La thèse est intitulée :

THREE ESSAYS ON POWER AND DIVERSITY

Lausanne, le 24 juin 2022

La Doyenne

Marianne SCHMID MAST

UNIL | Université de Lausanne

Members of the thesis committee

Prof. Christian Zehnder
Professor of Organizational Behavior and Decision Making &
Vice-Dean of Faculty
University of Lausanne
Thesis supervisor

Prof. Franciska Krings
Professor of Organizational Behavior and Decision Making &
Department Director
University of Lausanne
Thesis co-supervisor

Prof. Rafael Lalive
Professor of Economics & Vice-Dean of Research
University of Lausanne
Internal member of the thesis committee

Prof. Eva Derous
Professor of Personnel Management, Work and Organizational Psychology
University of Ghent
External member of the thesis committee

University of Lausanne
Faculty of Business and Economics

Ph.D. in Economics
Subject area “Management”

I hereby certify that I have examined the doctoral thesis of

Oriana de SAINT PRIEST

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 10.6.2022

Prof. Christian ZEHNDER
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

Ph.D. in Economics
Subject area “Management”

I hereby certify that I have examined the doctoral thesis of

Oriana de SAINT PRIEST

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 10.06.2022

Prof. Franciska KRINGS
Thesis co-supervisor

University of Lausanne
Faculty of Business and Economics

Ph.D. in Economics
Subject area “Management”

I hereby certify that I have examined the doctoral thesis of

Oriana de SAINT PRIEST

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: June 9, 2022

Prof. Rafael LALIVE
Internal member of the doctoral committee

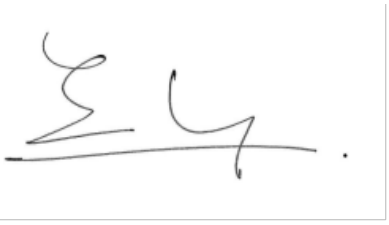
University of Lausanne
Faculty of Business and Economics

Ph.D. in Economics
Subject area “Management”

I hereby certify that I have examined the doctoral thesis of

Oriana de SAINT PRIEST

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 09.06.2022

Prof. Eva DEROUS
External member of the doctoral committee

Acknowledgements

Never could I have written this thesis on my own. While the PhD journey is above all a personal one, it would not be the same - or dare I say humanly possible, without the input and support of others. First and foremost, I would like to thank my supervisors, Professor Franciska Krings and Professor Christian Zehnder, who have constantly inspired me with their expertise, enthusiasm and rigor. Thanks to their continued guidance and support, I have not only grown as a scholar but as a woman, rich with skills that I trust will last me a lifetime. I will also never forget the fun moments we spent together, at conferences, winter and summer schools, as well as department parties and outings. You have both been role models for me in many areas and I dearly hope our collaboration continues in years to come.

I would also like to thank my colleagues and friends, who brightened my days in and out of the office. While I will not go through an exhaustive list, I would specifically like to thank my office mates Justin and Sirio for their feedback and support, as well as Sandra and Alessandro who also embarked on the PhD journey after our Masters in Cambridge and could best understand the joys and challenges of my daily life. I would also like to thank my friends outside of the academic world for their support, insights and inspiration. Special thanks to my Geneva based friends Enguerrand, Margaux, Delphine, Philippe, Benedetta and Agathe and to my close friends abroad, who have always been a phone call away from me. I think specifically of my dear bridesmaid and former Reuters colleague Alba, as well as my childhood friends Anais, Thomas and Yves. I consider myself extremely lucky to be friends with such passionate, inspiring and kind people and hope our friendships will last a lifetime, regardless of geographical distance.

Of course there is also no way I could ever have succeeded in my PhD without feeling supported by both my parents. May I continue to make both of you proud and raise your granddaughters as I was lucky to grow up, trusting that everything was possible with hard work. Thank you to both of you for inspiring me with your tireless determination, relentless motivation to move forward and passion for success. Special

thanks also go to my husband Alex, for always being next to me and bearing with the ups and downs of academic life. May we continue to grow together throughout life, in harmony and joy.

Last but definitely not least, I dedicate this thesis to our two daughters, Stella and Laetitia. Thank you for indirectly encouraging me to hone skills that could not have been more useful in my thesis: patience and optimism. You will always have my infinite love.

Synthesis

In my thesis, I study two management topics that are currently of pivotal importance in scholarly debate and practice alike. In fact, power is a foundational force of all individual, social and organizational life and it has long been at the center of attention in businesses and societies across the world. Research on power has thus been very active and it has resulted in many conclusions being drawn about power, with one of the most famous being "power corrupts". In my first chapter, I seek to contribute to a deeper understanding of power and leadership by establishing methodological best practices in the field of power. In particular, I aim to elucidate whether it might not be misguided to establish conclusions about behaviors associated with power in general, while power in practice has many different forms which could lead to diametrically opposed judgments, actions, and behaviors.

In the rest of my thesis, I study another issue that has recently been set as a priority in many countries and businesses worldwide: diversity. In my second thesis chapter, I focus more particularly on investigating from a macro perspective the trend of increased workforce gender diversity. My analysis reveals new insights on the causes of the gender wage gap in Switzerland, suggesting relevant practical changes both at the country and firm level. In my third thesis chapter, I build on the conclusions of my second chapter to suggest ways in which leaders can devise effective practices aimed at improving diversity inside their organizations. This chapter targets particularly age diversity, a topic within diversity that is currently gaining momentum in view of global demographic aging, but that is still understudied in the literature.

Overall my thesis contributes to advancing our knowledge of two central topics in management, making practical recommendations based on strong research methods to both scholars and practitioners.

An Investigation Into the Sensitivity of Power Studies to Subtle Differences in Experimental Manipulations

Thesis Paper

Oriana de Saint Priest, Christian Zehnder, Mahshid Khademi, Marianne Schmid Mast

Department of Organizational Behavior

University of Lausanne

Switzerland, 2022

Contents

Abstract

Power comes in many different forms and there is evidence that these various forms of power may impact behaviors in different ways. Nevertheless, many studies on power still present their conclusions in a very broad and general manner that does not take into account the multi-dimensionality of power. In this paper we explore if this is problematic by investigating the behavioral effects of two power manipulations designed to induce different forms of power. Our power manipulations both present the same choice set to the power holder and they are tested to have a very similar effect on felt power. The only dimension in which the manipulations differ is the framing with which the choice options are presented, either as reward or punishment. Against our hypothesis that reward power fosters prosocial behavior while punishment power stimulates antisocial behavior, we find that both power manipulations reduce prosocial choices. Our conclusions are positive for the power literature, as they indicate that the effects of power do not depend on subtle changes in the way in which power is manipulated.

Introduction

Power is heavily researched and the literature on power has established many interesting behaviors associated with power, such as increased corruption and risk-taking (Bendahan et al., 2015; Galinsky et al., 2015). Yet most of these results are stated in a general way, with little regard to the fact that power is a multi-dimensional concept (Schaerer et al., 2018). In fact, power comes in many different forms (Landells & Albrecht, 2013) as it stems from different bases (Raven, 1993). These forms of power can in turn have varying implications for behavior, which can even be diametrically opposed (Lammers et al., 2012; Willis et al., 2010). This variety in power types implies that power manipulations need to be adapted to the forms and sources of power that are relevant for a particular research question. Overall it might be misguided to generalize experimental results about the behavioral effects of power, if the research design manipulates exclusively one type of power. As there are almost infinitely many ways to manipulate power, it is important that the experimental manipulations of power chosen by researchers fit closely the context they are interested in. Still, notwithstanding these insights, many power scholars continue to present their results in a general way (Schaerer et al., 2018). In an attempt to shed further light on whether this is problematic, our paper contributes to a better understanding of how the detailed characteristics of power moderate its effects on behavior.

To reach this goal, we investigate the behavioral effects of two power manipulations which only differ minimally from each other. In fact, the only dimension in which our two power manipulations vary from each other is the framing with which the different choice options are presented to the power holder. We choose such similar power manipulations for the sake of methodological rigor. Indeed, from this point of view, the comparison between different forms of power raises important questions. To be able to identify precisely the causes of a behavioral change, the different forms of power need to be as similar as possible in all dimensions, except the dimension of interest. In particular, the compared power forms cannot differ in the perceived intensity of power, nor in their scope of impact, nor in the available resources over which

the power holder has control. In fact, if there were variations in these dimensions, then it would be hard to distinguish between the behavioral effects stemming from the form of power and those coming from these other dimensions. Hence we focus on assessing the behavioral impact of two power manipulations which grant power holders the possibility to change the payoffs of others after having observed their behavior, and only differ in whether the payoff changes are framed as reward or punishment. More precisely, participants in our study are split randomly into three conditions: reward power, punishment power and neutral power. Participants in the reward power condition are given the right to reward financially ten randomly selected participants from a pre-study, who displayed various degrees of honesty in this pre-study. In contrast, participants assigned to the punishment power condition are given the right to punish financially the same ten randomly selected participants from the pre-study. The two power conditions are designed such that participants in both power conditions can create the exact same outcomes for others. Moreover, we also ensured that both manipulations have the same effect on participants' perceived or felt power, using an extensive pre-test conducted before the main study with a separate sample of participants. This point is crucial to ensure that any differential behavioral effects induced by the two manipulations cannot be attributed to a difference in the strength of the power manipulations. In the neutral power condition, participants are passive observers who are presented with a randomly chosen and neutrally framed decision pattern of another participant in one of the power conditions.

We hypothesize that these two power manipulations may have different effects on pro- and antisocial behavior, as power has been shown to influence not only primary cognition - the content and amount of thoughts, but also secondary cognition - the way people think about their thoughts. In fact, high power has been linked to increased reliance on one's current thoughts, magnifying their influence on behavior (DeMarree et al., 2014). Since punishment power decreases social surplus, it is likely to give rise to antisocial thoughts, which might lead to further antisocial behavior. On the opposite, since reward power increases social surplus, it is likely to encourage prosocial thoughts, which might result in further prosocial

behavior. Hence we hypothesize that our two power manipulations which only differ in how they are framed, will impact the extent to which power impacts behavior. Specifically, we hypothesize that punishment power will make it more likely for participants to engage in antisocial behavior while reward power will make it more likely for participants to display prosocial behavior.

Our results show that the power manipulation framed as punishment significantly reduces the frequency of prosocial choices and therewith decreases social surplus, consistent with previous literature emphasizing the negative consequences of power. However, the reward power manipulation does not increase social surplus as we expected. In fact, we find that the effects of reward power are not statistically significantly different from the effects of punishment power, as both manipulations induce participants endowed with power to shift away from the prosocial allocation towards the antisocial and default allocations. We further investigate the extent to which our power manipulations induce long-lasting behavioral changes in participants, collecting extensive data on the emotions felt by the participants in our study. Our analysis confirms that participants in both high power conditions report feeling significantly more powerful than participants in the neutral power condition. In addition, the perceived level of power is very similar and not statistically distinguishable across the high power conditions. These results replicate the outcomes of our pre-test studying the effect of our power manipulations on felt power. However, in contrast to the pre-test where felt power was measured directly after the manipulations, felt power in the main study was measured after eliciting the dependent variable. These results therefore suggest that our manipulations of actual power have a long lasting effect on felt power and may therefore be a useful methodology that can be used confidently in future research aiming to manipulate actual power in experimental settings.

Our contributions to the literature are twofold. Firstly, we show that the behavioral effects of power do not seem to be strongly dependent on subtle elements of the experimental manipulations, yielding good news for the power literature. Yet, it is important to remember that the differences in behavioral effects

stemming from the two power manipulations we investigated need to be regarded as a lower bound, since the two power manipulations we analyzed only differ minimally from each other. Finding a difference in the behavioral effects stemming from these two similar power manipulations would have been striking. Thus one should not conclude from our paper that power studies are not sensitive to differences in power manipulations in general. This research should be seen as a first attempt to elucidate this question and future work should investigate whether the behavioral effects of power vary when the experimental manipulations used change in more fundamental ways.

Secondly, we also contribute to a new strand in the power literature which is moving away from power priming towards manipulations of actual power (Khademi et al., 2021; Schaerer et al., 2020). In fact, we build on our own prior work to devise two versions of the same manipulation of actual power, which were designed to affect two different power bases. Hence we study carefully the multi-dimensionality of power, in line with suggestions formulated in recent literature (Khademi et al., 2021; Schaerer et al., 2020, 2018). Furthermore, we provide evidence that the impact of our power manipulations on felt power is significant and robust across time, which constitutes a strong indication of their reliability. Thus we devise power manipulations which can be used in future research further interested in studying the effects associated with actual power.

Related Literature & Hypotheses

In this section we discuss how our paper relates to several different strands of existing literature and we develop our hypotheses.

Forms of Power

Our paper builds on research establishing that power comes in many different forms (Landells & Albrecht, 2013) and has many possible sources (Raven, 1993). This literature has long recognized that the behavioral implications of power may depend on the form that power takes (Wang & Sun, 2016; Caza et

al., 2011; Lammers et al., 2009) as well as on the sources from which power stems (Lammers et al., 2010; Lammers & Galinsky, 2008; Hornsey et al., 2003; Lammers et al., 2012, 2008). However, these important insights notwithstanding, many researchers in the power literature still present their results in a broad and general way that does not address the multi-dimensionality of power (Schaerer et al., 2018, for a review). Our paper attempts to make a first step towards a better and more systematic understanding of how the detailed characteristics of power moderate its effects on behavior.

From a methodological point of view, the comparison between different forms of power raises difficult questions. In order to strictly identify the determinant of a behavioral change, the different forms of power studied have to be as similar as possible in all dimensions, except the one of interest. In particular, the compared power forms cannot differ in the perceived intensity of power, nor in the scope of impact held by the power holders, nor in the resources available to them. In fact, variations in these dimensions would make it hard to distinguish between behavioral effects stemming from the form of power and between those coming from the strength of power. We therefore manipulate power in a very subtle and minimal way, as the only dimension in which our two power manipulations differ from each other is the framing with which the different choice options are presented to the power holders.

Power Manipulations

Existing experimental research investigating causal effects of power mainly relies on a technique called power priming (Schaerer et al., 2018). This power manipulation technique aims to activate a high or low power mindset in participants. To this end, participants are asked to think of either a high power or a low power role. They subsequently either act out this role in a role-play, imagine being this person, typically a manager or subordinate, or describe a situation in which they happened to be in the corresponding role (Schaerer et al., 2018; Tost, 2015; Anderson & Berdahl, 2002). The most widely used priming technique is the recall task, where participants have to describe a situation that occurred to them in

which they had power over others (high power priming) or in which someone else had power over them (low power priming) (Schaerer et al., 2018). In the neutral power condition, participants write about an event that happened to them the day before (Galinsky et al., 2003).

In some ways, power priming manipulations would be ideal to investigate subtle variations between different forms of power. As power priming targets "only" felt power rather than actual power, it would indeed be possible to change the hypothetical situation that the power holder should imagine being in, without varying the scope of power nor the resources available to the power holder. The only important dimension to be held constant would then be participants' felt power, which could be ensured using a manipulation check. Yet, power priming is unfortunately subject to at least two serious methodological concerns.

First, recall priming assumes that all study participants have appropriate power-related memories based on personal experience (Galinsky et al., 2003). However, since the researcher has no control over which memories participants choose to recall, this power manipulation is necessarily very subjective. In fact, different participants may remember very different situations, associated with various forms of power. Hence the recall prime manipulation might cause behavioral effects which depend on the type of power that participants choose to remember, which is impossible to perfectly control for. Researchers might be able to increase control when using versions of power priming in which participants are asked to recall an incident in a specific leader/follower situation (Dubois et al., 2010; Schmid, 2018; Rucker et al., 2011, for instance), but such attempts may encounter the problem that at least some participants may not have any relevant prior experience. Second, power priming methods require that the experimenter explicitly mentions a power role or a power related incident in the instructions. As participants are given salient cues about the manipulated variable, this might guide them in the experiment, to the extent that they hold expectations about how a powerful or a powerless person ought to behave. Hence the behavioral impact of power might not reflect a change in actual power, but may be driven by demand effects and simply reflect

what people think the effects ought to be (Khademi et al., 2021; Sturm & Antonakis, 2015; Schaerer et al., 2018; Lonati et al., 2018; Zizzo, 2010).

To avoid these issues, we build on our own previous research (Khademi et al., 2021), in which we have developed a simple power manipulation that exogenously changes actual power and grants researchers full control over the type of power they are manipulating. In our previous work, we manipulated actual power by giving participants control over valued resources in a setting where individuals in positions of power were free to decide about the allocation of valued resources to others. We avoided deception by giving participants control over real financial resources, raising the ecological validity of the design and the internal validity of the manipulation. Our previous work shows that this method substantially reduces the potential for demand effects. In our present research we use a similar manipulation, but instead of using a pure allocation task, we provide participants with the opportunity to change other participants' payoffs in response to their behavior in another study.

Pro- & Antisocial Behavior

One of the most widely studied hypotheses in the power literature is that power corrupts and induces power holders to engage in antisocial behavior. Previous empirical work has claimed that power increases the likelihood of punishment (Wiltermuth & Flynn, 2013), abuse (Foulk et al., 2018), objectification (Gruenfeld et al., 2008) and aggression (Bendahan et al., 2015; Zimbardo, 1973). Yet, this perspective has recently been nuanced, as it has been shown that under certain conditions, power can also increase prosocial behaviors. For instance, studies have claimed that power can increase helping (Frieze et al., 2001), generosity (Anderson et al., 2012), interpersonal sensitivity toward others (Schmid Mast et al., 2009), and moral awareness (DeCelles et al., 2012). Therefore it seems that social power can lead to corruption and unethical behavior in the pursuit of self-interest (Lammers et al., 2010), but can also result in a heightened concern for others and more generous contributions to collective resources (Chen et al.,

2001; Galinsky et al., 2003).

To reconcile the positive and negative effects of power on behavior, several mechanisms have been proposed. The one that has gathered most support in the literature suggests that power is associated with increased disinhibition (Foulk et al., 2020; Hirsh et al., 2011; DeMarree et al., 2012). In fact, power results in heightened activity of the behavioral approach system, which reduces the salience of competing responses and encourages goal directed behavior (Galinsky et al., 2011). In other words, power biases people's thoughts to be consistent with their initial preferences and values, reinforcing these inclinations (Copeland, 1994; Fischer et al., 2011). In practice, people endowed with high power increasingly use low rather than high thought routes, relying on stereotypes more than on effortful thought (Fiske, 1993; Guinote et al., 2002). High power is indeed typically linked to increased confidence in one's thoughts (Brinol et al., 2007), influencing not only the content and amount of thoughts (primary cognition), but also the way people think about their thoughts (secondary cognition) (DeMarree et al., 2012). High power can thus increase the impact of any accessible thought on behavior relative to low power. If accessible thoughts are antisocial in nature, power will lead to increased antisocial judgment and behavior. However, if accessible thoughts are prosocial in nature, power will result in increased prosocial judgment and behavior. This is true both if thoughts are chronically present, such as stemming from differences in personality, or situationally present, such as originating from a current goal, need, or informational context. Hence, power could amplify the tendency to endorse both prosocial and antisocial thoughts with action (DeMarree et al., 2014).[1] Thus there is evidence that power can encourage both antisocial and prosocial behaviors, depending on the way it is construed.

Building on this observation, we created two manipulations of actual power which both give the decision maker the exact same right to change other participants' financial payoffs after observing their behavior. Our manipulations differ only in whether the power holder's actions are framed as punishment or reward. In fact, the strategy space and payoff implications are completely identical in both our power

manipulations. This approach focused on varying framing makes sure that the scope of power and the resources available to the power holder are kept constant across power conditions. In addition, an extensive pre-test performed with an independent sample of participants before conducting the main experiment ensures that the two power manipulations do not differ with respect to perceived or felt power. As a control condition we use a version of the manipulation in which participants observe neutrally framed choices of another participant in one of the two high power conditions.

After having been subject to the power manipulation, participants face an allocation choice in which they can deviate in two ways from the initial default allocation. They can either give up some money to benefit a charity or they can get additional money for themselves at the expense of the charity. The prosocial choice is designed such that it increases social surplus, as the charity gains more than the participant foregoes, while the antisocial choice reduces social surplus, as the participant wins less than the charity looses.

Hypotheses

Based on the reasoning discussed in the previous section, we hypothesize that the framing of the power manipulation will affect its behavioral effects on power. More specifically, we expect that, as punishing decreases social surplus, endowing participants with punishment power will lead participants to form accessible antisocial thoughts. In turn, we expect these antisocial thoughts to lead to increased antisocial behavior.

Hypothesis 1: Punishment power increases antisocial behavior

Furthermore, as rewarding increases social surplus, we believe endowing participants with reward power will encourage participants to form accessible prosocial thoughts. In turn, we expect these prosocial thoughts to result in further prosocial behavior.

Hypothesis 2: Reward power increases prosocial behavior

Hence we hypothesize that our two power manipulations, which only differ in how they are framed, can impact the behavioral effects associated with power. We are aware that our research strategy is risky to an extent, as the difference between our manipulations is very subtle. On one hand, this feature makes our design particularly rigorous from a methodological point of view. On the other hand, there is a risk that the difference is too subtle to have an effect. Thus, if we find that our power manipulations yield different behavioral effects, the result would be striking in the sense that it would show that even the most subtle difference in induced form of power already makes a difference. The absence of an effect, in contrast, would constitute positive news for the power literature. In fact, such a result would mean that the conclusions drawn by power research are not sensitive to subtle changes in the way power is manipulated, though future work should then further investigate whether this holds when the power manipulations vary in more fundamental ways.

Method

We ran three separate studies - a pre-study, a pre-test and a main experiment, which occurred in this chronological order. Before describing the setup of these studies in more detail, we now provide a brief overview of our procedure. In a first stage, we ran a pre-study, which was used as an input in our power manipulations. In this pre-study, participants took part in a die rolling game where they could display various degrees of honesty. Ten participants from this pre-study were randomly selected and their behavior in the die rolling game was subsequently reported to participants in the pre-test and main experiment, as part of the power manipulations. In a second stage, we ran a pre-test which was designed to test the impact of our power manipulations on felt power. In this pre-test, participants were randomly allocated to three different power conditions: reward power, punishment power and neutral power. After the power manipulation, their felt power was measured. In a third stage, we ran the main experiment which was used to test whether our power manipulations triggered different degrees of prosocial behavior. In this main

experiment, participants were randomly allocated to the same three power conditions (reward, punishment and neutral power) as in the pre-test. After the power manipulation, participants' propensity to choose between a prosocial, antisocial and default option in an allocation task was measured.

Pre-Study

Our goal in the pre-study was to generate a list of people who displayed different degrees of honesty in a die rolling game, to use as an input for our power manipulations.

Participants. We recruited 105 participants on MTurk (Amazon Mechanical Turk). Out of this sample, 56 participants (68% men; mean age 36, SD 10.41) passed the attention and quality checks, which were the same as those used in the main study.[2] We also considered the full sample and provide all associated results in Appendix A.1. In the restricted sample, 70% of participants were employed full-time, 9% were employed part-time and the remaining 21% were unemployed.

Task. Participants in our pre-study took part in a version of the die rolling game (Fischbacher & Follmi-Heusi, 2013), in which they had to roll an electronic fair-sided die and report the outcome of their first die roll. They knew that their bonus in the study was a function of the self-reported first die roll ($0.1 multiplied by the reported die number if it is between 1 and 5 and $0 if the die number is 6). To psychologically facilitate cheating, participants were told that they should roll the electronic die several times "to check that it is fair". Cheating could be perfectly detected as the software recorded the true outcome of each participant's first die roll. Participants were aware of this recording feature which was explicitly described in the instructions. Participants were further informed that with probability 0.9 they would get the money associated with the number they chose to report, irrespective of whether the number they reported corresponded to the actual die roll. With probability 0.1, in contrast, their choice was going to be shown to a series of participants in another study, who were going to have the choice to change their payoffs. The precise instructions were the following: "The research team who organizes this study will be

able to see both the true outcome of your first die roll and the number that you report. However, the calculation of your bonus payment will only depend on the number that you report and will NOT depend on the true outcome of the die roll. Roughly 100 people participate in this study. Out of the 100 people who participate in this study the researchers will randomly select 10 people for passive participation in another study. The die reports of the 10 people selected will be shown to participants in the other study, who will then have the opportunity to change their bonus if they want to do so. The other 90 people who participate in this study will directly receive their unchanged bonus once this study has been completed by all 100 participants." There was no misrepresentation of information to the pre-study participants insofar as they were aware that, for 10 of them, participants in another study would have the ability to observe their behavior and change their bonus accordingly. However, participants in the pre-study were not given further details about the mechanism through which the participants in the other study would be able to change their bonus, i.e. through a reward or punishment feature. Participants in the pre-study were paid a fixed participation fee of $1, as well as their bonus if any, unless they were part of the 10 people selected for passive participation in the other study. If they were part of these 10 people, then they were initially only paid their participation fee, until data collection of the next study had been completed at which point they were paid their bonus too, if any. For each of the 10 pre-study participants selected, a random device picked one participant in the other study and we impacted the bonus of the pre-study participant accordingly.

Outcomes. Results show that 27% of participants chose to misreport their die roll, with the remaining 73% reporting their die roll accurately. The average bonus distributed was $0.32 per participant, while it would have been $0.22 had all participants reported their die roll truthfully. For a full description of actual and reported die rolls, please refer to Figure 1.

Pre-Test

After the pre-study and prior to running the main study, we performed a pre-test to measure the impact of our power manipulations (reward power and punishment power) on felt power. In fact, we needed to ensure that both our power manipulations had the same effect on felt power. The reason for doing so is related to our setting: both our power manipulations should be comparable in strength so that if we find that they have different behavioral effects, these cannot be interpreted as coming from a variation in their strength. In fact, if we find differences between the effects of the two power manipulations on pro- and antisocial behavior and one manipulation is substantially stronger with respect to felt power, we do not know whether the difference in effects is driven by the manipulation type, which we are interested in, or simply by the strength of the manipulation. We therefore needed to make sure that both the punishment power and reward power manipulations resulted in similar effect sizes on felt power. To this end, we devised an extensive pre-test in which participants were randomly allocated to our three different power conditions: reward power, punishment power and neutral power. After going through the power manipulation, participants were asked to report how their role in the study made them feel.

Participants. We recruited 400 participants on MTurk and there remained 279 participants after eliminating the people who failed our attention and understanding checks, though we also analyzed the full sample in Appendix A.2.[3] The clean sample had mean age 40 years old (SD 11.47) and it was 59% male. 72% of participants were employed full-time, 15% unemployed and the rest employed part-time. Out of these 279 participants, 84 were in the reward power condition, 94 in the punishment power condition and 101 in the neutral power condition. Participants were paid a fixed participation fee of $3.

Power Manipulations. We implemented three treatments in a between-subjects design (power manipulation: reward vs punishment vs neutral). Participants were first informed about the rules of the pre-study and were given full information on the instructions of the pre-study. They were then shown a report of the behavior displayed by the ten randomly selected pre-study participants. More precisely, the

information they were given included the first die roll of each of the ten pre-study participants, the die roll they reported, the bonus they got, whether they cheated or not, and how much money they made by cheating. For instance, person 1 was described as "Rolled a 6, reported a 6, got a bonus of $0. Did not cheat. Got a total payment of $1." Participants randomly allocated to the reward power condition then had the ability to increase each of the ten pre-study participants' payoff by any amount between $0 and $1. In contrast, participants randomly allocated to the punishment power condition could decrease each of the ten pre-study participants' payoffs by any amount between $0 and $1. To keep the strategy space and the payoff consequences identical across the two conditions, we simply added the maximal amount that participants in the reward power condition could attribute to pre-study participants ($1) to the payoffs of all pre-study participants shown in the punishment power condition. For instance, the description of person 1 in the punishment power condition read "Rolled a 6, reported a 6, got a bonus of $0. Did not cheat. Got a total payment of $2." This is in contrast with this person's description in the reward power condition, where his total payment was $1 instead of $2.

Thus, not punishing at all created the same payoff as maximally rewarding and full punishment created the same payoff as not rewarding. Hence the two conditions were calibrated such that the strategy space of participants in both power conditions was equivalent. All participants in a power condition could create the exact same outcomes for others so that the only difference between the two manipulations was the framing of the choices. Participants in the neutral power condition were passive observers who got to see a randomly chosen decision pattern of another participant in one of the power conditions, called the allocator. The decisions of the allocator in the neutral power condition were neither framed as rewards nor as punishments, but were presented in a neutral way. For instance, person 1 was described as "Rolled a 6, reported a 6, got a bonus of $0. Did not cheat. Got $1 by the allocator. Got a total payment of $2." The participants in the different conditions were not informed about the final payments of pre-study participants to avoid salient comparisons.

We ran our experiment on Qualtrics. To facilitate the logistics, we conducted data collection in two steps, first collecting data on the two high power conditions and then on the neutral power condition. This procedure allowed us to randomly select the decisions of a participant in the power conditions as an input in the neutral power condition. Lastly, a random device selected one allocation decision per pre-study participant and this person's bonus was affected accordingly. For instance, if for pre-study participant 1, the random device selected a participant in this study who decided to reward pre-study participant 1 by \$0.5, then pre-study participant 1 was paid his-her initial bonus plus \$0.5. However, if for pre-study participant 2, the random device selected a participant in this study who decided to punish pre-study participant 2 by \$0.5, then pre-study participant 2 was paid his initial bonus plus \$1 minus \$0.5.

Manipulation Check. After the power manipulation, participants answered several questions about how their role in this study made them feel using a Likert scale from 1 (strongly disagree) to 7 (strongly agree). Amongst 17 other emotions[4], participants were asked how powerful their role in this study made them feel. Table 1 reports mean statistics of the variables measuring the different emotions, depending on the different power conditions (reward: (1), punishment: (2), neutral: (3)), as well as T-Tests of the differences between the emotions felt in the different power conditions (reward and neutral: (4), punishment and neutral: (5) and reward and punishment: (6)). In the punishment power condition, participants felt more powerful (M = 4.61, SD = 1.69) than participants in the neutral condition (M = 3.19, SD = 1.68), p = .00. In the same way, participants in the reward power condition felt more powerful (M = 4.66, SD = 1.59) than participants in the neutral condition (M = 3.19, SD = 1.68), p = 0.00. Lastly, participants in the reward power condition did not feel more or less powerful (M = 4.66, SD = 1.59) than participants in the punishment power condition (M = 4.61, SD = 1.69), p = 0.38. Moreover, the T-tests reported in column (4) show that participants in the reward power condition report feeling significantly different than participants in the neutral power condition on an overwhelming majority of the emotions we elicited. In the same way, the T-tests shown in column (5) indicate that participants in the punishment

power condition felt significantly different than participants in the neutral power condition on a very large majority of the emotions we measured. These first results show that both our power manipulations affect felt power to the same extent. In addition, there is no significant difference in how participants in the reward and punishment power report feeling on a majority of the emotions we elicited, except that participants in the reward power condition felt more positive (M = 5.31, SD = 1.23) than participants in the punishment power condition (M = 4.63, SD = 1.50), p = 0.00. In the same way, participants in the punishment power condition felt more negative (M = 2.59, SD = 1.55) than participants in the reward power condition (M = 2.14, SD = 1.30), p = 0.04. This is in line with our initial predictions, as we expected that the destructive and negative nature of punishment power would increase the negativity of participants compared to the constructive and positive nature of reward power, which we find evidence for. Hence it seems that our power manipulations succeeded in keeping constant all the emotions that we wished to have no impact on, while only affecting the dimension in which we expected to find a difference depending on power conditions.

Main Experiment

In our main study, participants were randomly allocated to the three different power conditions: reward power, punishment power and neutral power. After going through the power manipulation, participants' propensity to choose between a prosocial, antisocial and default option in an allocation task was measured.

Participants. We recruited 800 individuals from the online platform Amazon Mechanical Turk. Applying our attention and understanding checks[5] resulted in excluding 216 observations, though analyses for the full sample are also reported in Appendix B. Our final sample was made up of 584 individuals. The clean sample had mean age 40 years old (SD 11.39) ; it was 55% male and 71% employed full-time. 15% of participants were unemployed and the rest employed part-time. Out of these 584 participants, 180 were

in the reward power condition, 199 in the punishment power condition and 205 in the neutral power condition. Participants received a fixed participation fee of $3. They could gain an additional amount going from $1.6 to $3.2, depending on their choice in the allocation task measuring pro- and antisocial behavior, described in more details below.

Power Manipulations. We implemented three treatments in a between-subjects design (power manipulation: reward vs punishment vs neutral). It is important to recall the main differences between these power manipulations. Participants randomly allocated to the reward power condition were given the right to reward financially ten randomly selected participants from the pre-study, by giving them any amount between $0 and $1. In contrast, participants randomly allocated to the punishment power condition were given the right to punish financially ten randomly selected participants from the pre-study, by subtracting from their payment any amount between $0 and $1. Lastly, participants randomly allocated to the neutral power condition were passive observers who got to see a randomly chosen decision pattern of another participant in one of the power conditions, called the allocator. The decisions of the allocator in the neutral power condition were neither framed as rewards nor as punishments, but were presented in a neutral way.

Pro- and Antisocial Behavior. We used an allocation task to measure participants' pro- and antisocial behavior. In this allocation task, participants had full autonomy to decide how an endowment was to be split between them and a charity. More precisely, we chose a well-known charity focused on increasing the well-being of children, the UNICEF. We used no deception and fully informed participants of the decisions and payouts available to them. This way we could ensure a direct and salient connection between decisions taken and desired monetary outcome, and therefore "the interpretability and the internal validity of the experiment" (Zizzo, 2010). Participants' propensity to choose a prosocial and efficient option over an inefficient and selfish alternative was measured using this incentivized choice situation. Participants could choose between three options: a default option that splits the initial surplus (100 points) equally between the participant (50 points) and the charity (50 points), an antisocial option that increases

the participant's payoff (80 points) at the expense of the charity (10 points) and reduces overall surplus (from 100 to 90 points), and a prosocial option that decreases the participant's payoff (40 points) to the benefit of the charity (80 points) and increases overall surplus (from 100 to 120 points). In the experiment, 10 points correspond to $0.4. All participants were paid out according to their choice in this allocation task.

Manipulation Check. We devised a manipulation check in order to measure how the punishment power and reward power manipulations affected felt power over the full duration of the experiment. We therefore conducted the manipulation check after eliciting the dependent variable. The manipulation check consisted of several questions about how the participants' role in this study made them feel. Among 17 other emotions, participants were asked how powerful their role in this study made them feel.

Potential for Demand Effects. In previous work we showed that manipulations of actual power, similar to the ones used in this study, create a significantly lower potential for demand effects than manipulating felt power with priming methods (Khademi et al., 2021). To measure how many participants were able to guess correctly the research question behind our study we therefore asked participants to report what they thought this study was about and the reason for their choice in the allocation task.[6]

Results

Pro- & Antisocial Choices

We first discuss how our power manipulations affect participants' choices in the allocation task. Figure 2 shows that both our high power conditions induce a shift towards less prosocial and more antisocial behavior. Participants in both high power conditions choose the prosocial option less often (reward power: 19%, punishment power: 18%) and the antisocial option more often (reward power: 46%, punishment power: 47%) than participants in the neutral power condition (26% and 41%, respectively).

We use the regression analysis report in Table 2 to investigate these changes in the choice distribution in more detail (for the analysis including the whole sample without exclusion criteria please

refer to Appendix B.1). Column (1) explores the effect of our power manipulations on the likelihood of choosing the prosocial allocation. We regress an indicator variable which is one if the participant picked the prosocial choice on indicator variables for our treatments. The neutral power condition is the baseline category (i.e. the constant (0.26) represents the relative frequency of the prosocial choice in the neutral power condition). The regression reveals that the decrease in the frequency of prosocial choices (-0.08) induced by punishment power is statistically significant ($p = 0.04$). The effect of reward power is very similar in magnitude (-0.07), but the statistical significance is marginal ($p = 0.10$). Column (2) uses the same specification but includes controls. The results remain largely unchanged (punishment power: $p = 0.05$, reward power: $p = 0.08$). F-tests show that the effects of the two high power conditions are not statistically different in either column ($p = 0.75$ in Column (1) ; $p = 0.90$ in Column (2)).

Columns (3) and (4) examine the effects of the high power manipulations on the likelihood of choosing the antisocial allocation. The regression specifications are equivalent to those in Columns (1) and (2) except that the dependent variable is now an indicator variable which is one if the participant picked the antisocial choice. We find that the increases in the frequency of antisocial choices in our two high power conditions are not statistically significant (punishment power: (1) $p = 0.21$, (2) $p = 0.26$, reward power: (1) $p = 0.37$, (2) $p = 0.30$). F-tests reveal that the effects of the two high power conditions are not statistically different in either column ($p = 0.74$ in Column (3) ; $p = 0.96$ in Column (4)).

Columns (5) and (6) complete the analysis by using the same specifications to estimate the effects of the high power manipulations on the likelihood of choosing the default allocation. Both effects are small, positive and not statistically significant (punishment power: (1) $p = 0.68$, (2) $p = 0.62$, reward power: (1) $p = 0.63$, (2) $p = 0.67$). F-tests show that the effects of the two high power conditions are not statistically different in either column ($p = 0.95$ in Column (5) ; $p = 0.96$ in Column (6)).

Our analysis reveals a very clear pattern: both high power manipulations induce a shift in the choice distribution away from the prosocial option towards the default option and the antisocial option. In relative

terms, the decrease in the proportion of prosocial choices is actually quite substantial in both conditions (roughly a decrease by 30%). The observed pattern is only partially in line with our hypotheses. Our results provide support to Hypothesis 1, stating that punishment power will increase antisocial behavior. However, our analysis clearly rejects Hypothesis 2, stating that reward power will increase prosocial behavior. In fact, participants endowed with reward power respond in a very similar way to participants in the punishment power condition and select increasingly the antisocial allocation, compared with people in the neutral power condition. From a statistical point of view, the effects of our two manipulations are indistinguishable.

Payoffs & Social Surplus

The shift in the choice distribution discussed in the previous section also has implications for the payoff distribution between the participants and the charity, as well as for social surplus (i.e., the aggregation of participants payoffs and charity payoffs).

As a reminder, social surplus was 100 points if the participant selected the default allocation which split the initial surplus equally between him (50 points) and the charity (50 points). Social surplus was 90 points if the participant selected the antisocial allocation which gave him a larger payoff (80 points) at the expense of the charity (10 points). Lastly, social surplus was 120 points if the participant chose the prosocial allocation decreasing his-her payoff (40 points) to the benefit of the charity (80 points).

Figure 3 shows how our high power conditions affect the payoffs of the participants and the charity, as well as social surplus. The decrease in prosocial choices and the associated increase in default and antisocial choices imply that the charity payoffs (donations) go down from 41.4 points in the neutral condition to 36.6 points in the punishment power condition (minus 12%) and to 37.3 points in the reward power condition (minus 10%). At the same time, participants payoffs increase from 59.7 points to 62.3 points in the punishment power condition (plus 4%) and to 61.9 points in the reward power condition (plus

4%). The high power conditions therefore increase participants payoffs at the expense of the donations to the charity. However, since the increase in participants' payoffs is smaller than the decrease in donations (which is a mechanical effect of our design), social surplus decreases from 101.2 points in the neutral power condition to 98.9 points in the punishment power condition and 99.3 points in the reward power condition.

Table 3 displays OLS estimations in which we regress social surplus (Columns (1) and (2)), charity payoffs (Columns (5) and (6)) and participants payoffs (Columns (9) and (10)) on indicator variables for our two high power conditions. In addition the table also contains specifications in which we regress the same dependent variables on a joint indicator variable for both high power conditions (i.e. the variable is unity if the observation comes from either high power condition ; see Columns (3), (4), (7), (8), (11), and (12)). For a similar table based on the whole dataset without applying exclusion criteria, please refer to Appendix B.2. Table 3 reveals that the negative effects on social surplus are statistically significant in the punishment power condition (see Column (1), $p = 0.05$) and in the joint estimation (see Column (3), $p = 0.04$).[7] Moreover the effects on charity and participants payoffs are not or only marginally significant.[8]

Ex-Post Reported Choice Motives

In light of these results, we next seek to understand why participants chose the allocation they did. In the survey, participants were asked to respond to a free-form question which invited them to report what allocation they selected and why they made this decision. Two coders went through all the responses and coded them using a variety of categories encompassing the most frequent answers, including financial need, fairness concerns, desire to help the charity, lack of knowledge about whether the money would be distributed, concerns that the contribution size was too small and desire to maximize social surplus. Across all the data we collected, a majority of participants reported choosing their allocation because of fairness concerns (29%) and financial need (27%). Another 13% indicated wanting to help the charity and 11% gave no justification. 9% of participants said they disliked the charity, indicating that they had issues with

charities' corruption, administrative costs, their support of certain causes such as abortion and/or they indicated that they would rather donate elsewhere. 6% of participants reported that they wanted to maximize social surplus, and the rest indicated reasons such as a lack of knowledge about whether the money would actually be distributed, that the contribution size was too small hence did not matter, and that they were already donating to UNICEF. Overall, these results show that most participants were sensitive to fairness concerns and wanted to help the charity, with the remaining part choosing the allocation they did because of financial need. For a table detailing motives mentioned across the two batches of data collected, please refer to Table 4.

We next consider the frequency of certain choice and choice motive combinations (the allocation choice and the reported choice motive). We focus on the two most frequently reported motives per allocation choice and pool other motives into an "other" category. Figure 4 reports the frequency of the different choice and choice motive combinations pooling all conditions, highlighting the most frequently cited motives per allocation choice (prosocial, antisocial, default). We find that participants choosing the default allocation most frequently report doing so for fairness reasons (26.5% of all participants), or do not report any reasons at all (5.8% of all participants). Participants choosing the antisocial allocation most frequently explain they did so for financial reasons (25.0% of all participants) and because of their dislike for the charity (8.4% of all participants). Lastly, participants choosing the prosocial allocation most frequently said they did so to help the charity (10.1% of all participants) and to maximize social surplus (5.4% of all participants).

Next we investigate how our treatments affect the frequency of choice - choice motive combinations. This allows a clean analysis avoiding the complicated endogeneity issues encountered with an analysis conditional on choice. Figure 5 indicates the frequency of the different choice and choice motive data, showing the most frequently cited motives per allocation choice (prosocial, antisocial, default) and per power condition. In line with what we expected, we find that our power manipulations seem to decrease

people's social inclinations. Indeed, the share of participants choosing the prosocial allocation and further indicating they did so to help the charity decreases in the punishment (6.9% of participants) and in the reward conditions (8.5% of participants) compared to the neutral condition (14.7% of participants). The power manipulations also have the expected effects on the other choice - choice motive combinations.[9]

Table 5 shows OLS estimations in which we regress an indicator variable for the choice and choice motive combinations on power treatment dummies, as we did in the choice data. More precisely, we regress the two most commonly cited choice motives for participants who chose the default allocation (Columns (1) and (2)), antisocial allocation (Columns (4) and (5)) and prosocial allocation (Columns (7) and (8)) on indicator variables for our two high power conditions. The remaining columns pool together all the other motives stated by participants who chose the default allocation (Column (3)), antisocial allocation (Column (6)) and prosocial allocation (Column (9)). The table reveals a significant negative effect of punishment power on the likelihood that participants reported choosing the prosocial allocation to help (see Column (7), $p = 0.02$)). This provides a further indication that power decreases social inclinations. Moreover the effects of both power conditions on the other choice - choice motive combinations are not significant.[10]

Reported Perceptions

We next analyze the effect of the power manipulations on perceptions after the dependent variable has been elicited.[11] We collected data on 345 participants after eliminating the people who failed our attention and understanding checks. The results ran on the full sample[12] are available in Appendix B.3. The clean sample had mean age 41.5 years old (SD 11.92), it was 53.9% male and 69.9% employed full-time (16.2% were unemployed and the rest employed part-time). Out of these 345 participants, 105 were in the reward condition, 117 in the punishment condition and 123 in the neutral power condition.

We find that participants in the punishment power condition do not feel significantly more powerful ($M = 4.85$, $SD = 1.61$) than participants in the reward power condition ($M = 4.86$, $SD = 1.48$), $p = 0.52$, in

line with results from the pre-test. Moreover, our power manipulations have a significant effect on felt power. In fact the average power felt by people in the punishment power condition (M = 4.85, SD = 1.61) is significantly higher than the average power felt by people in the neutral power condition (M = 2.74, SD = 1.59), p = 0.01. Also, the average power felt by people in the reward condition (M = 4.86, SD =1.48) is significantly higher than the average power felt by people in the neutral power condition (M = 2.74, SD = 1.59), p = 0.01. All these results are consistent with those obtained in the pre-test.

In addition, we collected data on how much participants judged their role in this study made them feel a variety of emotions, as displayed in Figure 6. Analyzing this data in more detail, Table 6 reports mean statistics of the variables measuring the different emotions, depending on power conditions (reward: (1), punishment: (2), neutral: (3)), as well as T-Tests of the differences between the emotions felt in the different power conditions (reward and neutral: (4), punishment and neutral: (5) and reward and punishment: (6)). This table reveals that participants in the reward power condition report feeling significantly more in charge, positive, responsible, dominant, authoritative, in control, determined, influential, respected, concerned and accountable than participants in the neutral power condition. In the same way, participants in the punishment power condition report feeling significantly more in charge, positive, guilty, responsible, dominant, negative, authoritative, in control, determined, influential, respected, self-centered, concerned, accountable and egoistic than participants in the neutral power condition.[13] Overall these results indicate that we succeed in devising strong manipulations of actual power, as the emotions they create survive the elicitation of the dependent variable. Hence our power manipulations do not seem to suffer from spillover issues, which we could not measure in our pre-test. Moreover, our manipulations of actual power have the additional advantage that they do not give rise to demand effects, as we report in Appendix B.4.

Lastly, our results show that there is no significant difference in how participants report feeling in the two high power conditions for a vast majority of the emotions we elicited. There were differences in only

three out of the seventeen emotions we measured. More precisely, participants in the reward power condition feel significantly less guilty (M = 1.87, SD = 1.06) than participants in the punishment power condition (M = 2.26, SD = 1.48), p = 0.01. Participants in the reward power condition also report feeling significantly less negative (M = 2.04, SD = 1.15) than participants in the punishment power condition (M = 2.46, SD = 1.41), p = 0.01. Moreover, participants in the reward power condition feel significantly less self-centered (M = 2.33, SD = 1.30) than participants in the punishment power condition (M = 2.70, SD = 1.42), p = 0.02. It is interesting to observe that results for the full sample reported in Appendix B.3 mirror these results as well as those in the pre-test, insofar as participants in the punishment power condition feel significantly more negative than participants in the reward power condition. Across all the data collected, it seems as though the most robust difference in emotions elicited by our two power manipulations is thus that in perceived negativity. Hence we initially thought that we succeeded in keeping all emotions constant, except those of interest. In fact, we expected the destructive nature of punishment power to increase the negativity of participants, in contrast to the constructive nature of reward power. Yet, we now know this difference in negativity was not enough to yield important effects on behavior, as we find that participants endowed with reward power respond in a very similar way to participants in the punishment power condition.

Discussion & Conclusion

In this paper, we made a first step towards addressing the important concern that using simple manipulations of power to derive general results about power may be problematic. In fact, power has different sources (Raven, 1993), which have different behavioral effects (Lammers et al., 2012; Lammers & Galinsky, 2008; Willis et al., 2010; Lammers et al., 2008, for instance). In turn, simple manipulations of power may vary different types of power, which could then affect behavior in different ways (Wang & Sun, 2016; Caza et al., 2011, for instance). Hence, deriving general conclusions about power from research

using simple power manipulations may be misguided. In an incentivized experimental design, we study the empirical relevance of this concern. More precisely, we develop two novel manipulations of actual power, which both give power holders the same right to allocate resources but differ in whether they are framed as reward or punishment. We investigate the effect of these power manipulations on antisocial and prosocial behavior. Our measure of prosocial behavior indicates the extent to which participants choose an allocation favoring a charity over themselves, compared to a default allocation. Our measure of antisocial behavior shows the extent to which participants choose an allocation favoring themselves over a charity, compared to a default allocation.

In line with our hypothesis, we find that the power manipulation framed as punishment decreases prosocial behavior and has a negative impact on social surplus. This is consistent with previous literature emphasizing the negative behavioral consequences of power. The reward power manipulation, in contrast, does not increase social surplus as we expected, but instead has similar effects as punishment power. In fact, we find that the two power manipulations are not statistically significantly different from each other. Further analysis reveals that participants endowed with power shift away from the prosocial allocation towards the antisocial and default allocations, with this behavior change being statistically significant for people endowed with punishment power. Hence it seems as though the negative behavioral consequences of power prevail in our setting, regardless of the power manipulation used.

We also investigated the extent to which our manipulations of actual power based on resource allocation induced long lasting effects on the perceptions of participants. In particular, we analyzed how much participants felt their role in our study made them feel in charge, positive, guilty, responsible, dominant, powerful, inspired, negative, authoritative, in control, determined, influential, respected, self-centered, concerned about others, opportunistic, accountable and egoistic. We show that participants in both power conditions feel significantly more powerful than people in the neutral power condition, indicating the strength of our power manipulations. In addition, participants in both high power conditions

report feeling equally powerful. The analysis of the other emotions reveals that our two power manipulations were similar in strength, as they triggered many of the emotions we elicited, compared to the neutral power condition. These results constitute evidence that we succeeded in contributing to the power literature by devising strong manipulations of actual power, which could later be used in further experimental investigations of the behavioral effects associated with power.

More generally, our research highlights the negative impact of power on prosocial behavior, which could be influenced by several factors both at the personal and organizational level. At the personal level, social dominance orientation (SDO) has previously been shown to influence the prosocial behavior of high-status group members (Halabi et al., 2008). More precisely, Social Dominance Theory argues that societies form ideologies to minimize intergroup conflict, maintaining group inequality and legitimizing discrimination (Sidanius & Pratto, 2004). Consistent with this, people's support for group dominance varies and can be measured with SDO. Individuals with a high SDO reinforce social hierarchies, whereas individuals low in SDO are more likely to contribute to reducing inequalities (Sidanius & Pratto, 2004). Generosity also influences SDO, as generosity has been found to decrease SDO (Brown, 2011). Relatedly, perceived power distance, defined as the extent to which one accepts that power in institutions and organizations is distributed unequally (Hofstede, 2001), has been linked with prosocial behavior (Winterich & Zhang, 2014).

At the organizational level, culture has been shown to affect how power is experienced. In fact, Western cultures (e.g., US, Europe) tend to be more individualistic, with people construing themselves as separate from others and focusing on their unique attributes (Lee et al., 2000; Markus & Kitayama, 1991), while Eastern cultures (e.g., Southeast Asia) tend to be relatively interdependent. In these cultures, people focus more on their relationships with others and as a result are more interconnected (Heine et al., 1999; Markus & Kitayama, 1991). Power is thus experienced differently in Western cultures, where it is seen as an opportunity to reward oneself, compared to Eastern cultures, where it is seen as a responsibility to others

(Zhong et al., 2006). These differences in the ways in which power is seen has further effects on behavior, demonstrated in empirical studies. For instance, Westerners high in power take more resources for themselves compared to Westerners low in power, though the opposite pattern emerges for Easterners (Kopelman, 2009). Similarly, the percentage of money that was given to others by "allocators" in ultimatum games varied greatly depending on cultures studied, ranging from 26% in a Peruvian sample to 58% in an Indonesian sample (Henrich et al., 2001).

In addition, accountability has been shown to moderate the impact of power on behavior (Pitesa & Thau, 2013; Anderson & Galinsky, 2006). Accountability refers to the implicit or explicit expectation that one may be asked to justify one's beliefs, feelings, and actions to others (Lerner & Tetlock, 1999; Tetlock, 1992). Keltner et al. (2003) suggested that accountability is likely operating as an equalizer, making the powerful behave more like the powerless. In fact, individuals with low power generally consider how their behavior will be perceived by others, which individuals with high power held accountable for their decisions will also be more likely to do. Accountability will indeed increase activity in the behavioral inhibition system, prompting a more careful consideration of behavior (Keltner et al., 2003). Hence the negative behavioral consequences of power are likely to be offset when individuals are made accountable for their deeds.

In future studies, it would be interesting to further build on our current work to investigate in more detail the mechanisms behind the behavioral effects of power we identified, in light of the variables mentioned above. We could also devise manipulations of actual power targeting different types of power, which vary from each other in more fundamental ways. In fact, the only dimension in which the two power manipulations we used vary from each other is the framing with which the different choice options are presented to the power holder. We chose such similar power manipulations as we wished to be able to identify precisely the determinants of a potential behavioral change and argued that to do so, the different forms of power needed to be as similar as possible in all dimensions, except the dimension of interest. In

particular, the compared power forms could not differ in their perceived intensity, nor in the available resources over which the power holder had control, nor in their scope of impact. In fact, if there were variations in these dimensions, then it would have been hard to distinguish between the behavioral effects stemming from the form of power and those coming from these other dimensions. Yet, since our current work did not reveal that power studies were sensitive to subtle changes in the power manipulations, it would now be interesting to relax this assumption and investigate behavioral changes associated with more fundamentally different power manipulations. For instance, while the behavioral consequences of positional power have been explored at length – and those of personal power to a large extent, informational and connection power have yet to be studied to the same degree (Norbom & Lopez, 2016). In fact, there have only been very few studies on the behavioral impact of informational power (Tjosvold et al., 2003, for instance) and on that of connection power (Thomas-Dean, 2022) in the leadership literature. Hence future research should definitely investigate more thoroughly these dimensions of power, which have yet to receive the full attention they deserve.

While the contributions of our paper are positive for the power literature insofar as they indicate a certain robustness of the results revealed by power studies, one should not conclude from our research that power studies are not sensitive at all to differences in power manipulations. This research indeed constitutes only a first step in our research agenda, as future work should now uncover whether these conclusions hold when the power manipulations vary in more fundamental ways. In the meantime, we suggest being prudent about deriving general conclusions on the behavioral effects of power using manipulations of felt power, since the behavioral consequences of power may depend on the way it is construed. Moreover, our paper also allows us to derive interesting practical implications for business environments. In fact, our results to date highlight the negative impact of both reward and punishment power on prosocial behavior. Hence it is crucial for practitioners to consider how best to mitigate the negative impact of power that we identified. For instance, it has been shown that culture seems to play an extremely important role in mediating the

impact of power on behavior (Zhong et al., 2006). Hence, managers might wish to explore further ways in which they could promote a culture emphasizing interdependencies and reliance on others, as opposed to individualism and personal rewards. For instance, this could be done through symbolic management, which clearly communicates desirable behaviors and attitudes (Roberge et al., 2011). Moreover, businesses should consider how to make employees endowed with high power accountable for their decisions, as accountability has been proved to moderate the effect of power on antisocial behavior (Pitesa & Thau, 2013). In practice, individuals could be held accountable through various ways, including required transparency, codes of ethical conducts, performance evaluations, and financial incentives (Schaerer et al., 2018). More generally, it might also be worth thinking about the personalities of individuals who will be granted high power. Are these individuals supportive of group dominance and if so, to what extent? Which inequalities do they believe are acceptable and where would they draw their own ethical boundaries? Assessing such elements before granting high power to an employee would certainly help in carefully selecting the right people into high status roles, overall mitigating certain negative consequences of power.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53*, 63–70.

Anderson, C., & Berdahl, J. L. (2002). The experience of power: examining the effects of power on approach and inhibition tendencies. *Journal of Personality and Social Psychology*, *83*(6), 1362.

Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and risk-taking. *European Journal of Social Psychology*, *36*(4), 511–536.

Anderson, C., John, O. P., & Keltner, D. (2012). The personal sense of power. *Journal of Personality*, *80*(2), 313–344.

Bendahan, S., Zehnder, C., Pralong, F. P., & Antonakis, J. (2015). Leader corruption depends on power and testosterone. *The Leadership Quarterly*, *26*(2), 101–122.

Brinol, P., Petty, R. E., Valle, C., Rucker, D. D., & Becerra, A. (2007). The effects of message recipients' power before and after persuasion: a self-validation analysis. *Journal of Personality and Social Psychology*, *93*(6), 1040.

Brown, M. A. (2011). The power of generosity to change views on social power. *Journal of Experimental Social Psychology*, *47*(6), 1285–1290.

Caza, B. B., Tiedens, L., & Lee, F. (2011). Power becomes you: The effects of implicit and explicit power on the self. *Organizational Behavior and Human Decision Processes*, *114*(1), 15–24.

Chen, S. (2001). Lee–chai, ay, & bargh, ja (2001). relationship orientation as a moderator of the effects of social power. *Journal of Personality and Social Psychology*, *80*, 173–187.

Chen, S., Lee-Chai, A. Y., & Bargh, J. A. (2001). Relationship orientation as a moderator of the effects of social power. *Journal of Personality and Social Psychology*, *80*(2), 173.

Copeland, J. T. (1994). Prophecies of power: Motivational implications of social power for behavioral confirmation. *Journal of Personality and Social Psychology*, *67*(2), 264.

DeCelles, K. A., DeRue, D. S., Margolis, J. D., & Ceranic, T. L. (2012). Does power corrupt or enable? when and why power facilitates self-interested behavior. *Journal of Applied Psychology*, *97*(3), 681.

DeMarree, K. G., Briñol, P., & Petty, R. E. (2014). The effects of power on prosocial outcomes: A self-validation analysis. *Journal of Economic Psychology*, *41*, 20–30.

DeMarree, K. G., Loersch, C., Briñol, P., Petty, R. E., Payne, B. K., & Rucker, D. D. (2012). From primed construct to motivated behavior: Validation processes in goal pursuit. *Personality and Social Psychology Bulletin*, *38*(12), 1659–1670.

Dubois, D., Rucker, D. D., & Galinsky, A. D. (2010). The accentuation bias: Money literally looms larger (and sometimes smaller) to the powerless. *Social Psychological and Personality Science*, *1*(3), 199–205.

Fischbacher, U., & Follmi-Heusi, F. (2013). Lies in disguise,Äîan experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525–547.

Fischer, J., Fischer, P., Englich, B., Aydin, N., & Frey, D. (2011). Empower my decisions: The effects of power gestures on confirmatory information processing. *Journal of Experimental Social Psychology*, *47*(6), 1146–1154.

Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, *48*(6), 621.

Foulk, T. A., Chighizola, N., & Chen, G. (2020). Power corrupts (or does it?): An examination of the boundary conditions of the antisocial effects of experienced power. *Social and Personality Psychology Compass*, *14*(4), e12524.

Foulk, T. A., Lanaj, K., Tu, M.-H., Erez, A., & Archambeau, L. (2018). Heavy is the head that wears the crown: An actor-centric approach to daily psychological power, abusive leader behavior, and perceived incivility. *Academy of Management Journal*, *61*(2), 661–684.

Frieze, I. H., Boneva, B. S., Lee-Chai, A., & Bargh, J. (2001). Power motivation and motivation to help others. *The Use and Abuse of Power*, 75–89.

Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology*, *85*(3), 453.

Galinsky, A. D., Rucker, D. D., & Magee, J. C. (2015). *Power: Past findings, present considerations, and future directions.* American Psychological Association.

Galinsky, A. D., Rus, D., & Lammers, J. (2011). *Power: A central force governing psychological, social, and organizational life.* Routledge/Taylor & Francis Group.

Gruenfeld, D. H., Inesi, M. E., Magee, J. C., & Galinsky, A. D. (2008). Power and the objectification of social targets. *Journal of Personality and Social Psychology*, *95*(1), 111.

Guinote, A., Judd, C. M., & Brauer, M. (2002). Effects of power on perceived and objective group variability: evidence that more powerful groups are more variable. *Journal of Personality and Social Psychology*, *82*(5), 708.

Halabi, S., Dovidio, J. F., & Nadler, A. (2008). When and how do high status group members offer help: Effects of social dominance orientation and status threat. *Political Psychology*, *29*(6), 841–858.

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological review*, *106*(4), 766.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., ... others (2001). Cooperation, reciprocity and punishment in fifteen small-scale societies. *American Economic Review*, *91*(2), 73–78.

Hirsh, J. B., Galinsky, A. D., & Zhong, C.-B. (2011). Drunk, powerful, and in the dark: How general processes of disinhibition produce both prosocial and antisocial behavior. *Perspectives on Psychological Science*, *6*(5), 415–427.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.

Hornsey, M. J., Spears, R., Cremers, I., & Hogg, M. A. (2003). Relations between high and low power groups: The importance of legitimacy. *Personality and Social Psychology Bulletin*, *29*(2), 216–227.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, *110*(2), 265.

Khademi, M., Mast, M. S., Zehnder, C., & De Saint Priest, O. (2021). The problem of demand effects in power studies: Moving beyond power priming. *The Leadership Quarterly*, *32*(4), 101496.

Kopelman, S. (2009). The effect of culture and power on cooperation in commons dilemmas: Implications for global resource management. *Organizational Behavior and Human Decision Processes*, *108*(1), 153–163.

Lammers, J., & Galinsky, A. (2008). How the conceptualization and nature of interdependency moderates the effects of power. *Power and Interdependence in Organizations*, 67–82.

Lammers, J., Galinsky, A. D., Gordijn, E. H., & Otten, S. (2008). Illegitimacy moderates the effects of power on approach. *Psychological Science*, *19*(6), 558–564.

Lammers, J., Galinsky, A. D., Gordijn, E. H., & Otten, S. (2012). Power increases social distance. *Social Psychological and Personality Science*, *3*(3), 282–290.

Lammers, J., Stapel, D. A., & Galinsky, A. D. (2010). Power increases hypocrisy: Moralizing in reasoning, immorality in behavior. *Psychological Science*, *21*(5), 737–744.

Lammers, J., Stoker, J. I., & Stapel, D. A. (2009). Differentiating social and personal power: Opposite effects on stereotyping, but parallel effects on behavioral approach tendencies. *Psychological Science*, *20*(12), 1543–1548.

Landells, E., & Albrecht, S. L. (2013). Organizational political climate: Shared perceptions about the building and use of power bases. *Human Resource Management Review*, *23*(4), 357–365.

Lee, A. Y., Aaker, J. L., & Gardner, W. L. (2000). The pleasures and pains of distinct self-construals: the role of interdependence in regulatory focus. *Journal of personality and social psychology*, *78*(6), 1122.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological bulletin*, *125*(2), 255.

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, *64*, 19–40.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, *98*(2), 224.

Norbom, H. M., & Lopez, P. D. (2016). Leadership and innovation: informal power and its relationship to innovative culture. *Journal of Leadership Studies*, *10*(1), 18–31.

Pitesa, M., & Thau, S. (2013). Compliant sinners, obstinate saints: How power and self-focus determine the effectiveness of social influences in ethical decision making. *Academy of Management Journal*, *56*(3), 635–658.

Raven, B. H. (1993). *The bases of power: Origins and recent developments.* Wiley Online Library.

Roberge, M.-É., Lewicki, R. J., Hietapelto, A., & Abdyldaeva, A. (2011). From theory to practice: Recommending supportive diversity practices. *Journal of Diversity Management (JDM)*, *6*(2), 1–20.

Rucker, D. D., Dubois, D., & Galinsky, A. D. (2011). Generous paupers and stingy princes: Power drives consumer spending on self versus others. *Journal of Consumer Research*, *37*(6), 1015–1029.

Schaerer, M., Lee, A. J., Galinsky, A. D., & Thau, S. (2018). *Contextualizing social power research within organizational behavior.* Routledge/Taylor & Francis Group.

Schaerer, M., Teo, L., Madan, N., & Swaab, R. I. (2020). Power and negotiation: review of current evidence and future directions. *Current Opinion in Psychology*, *33*, 47–51.

Schmid, P. C. (2018). Less power, greater conflict: Low power increases the experience of conflict in multiple goal settings. *Social Psychology*, *49*(1), 47.

Schmid Mast, M., Jonas, K., & Hall, J. A. (2009). Give a person power and he or she will show interpersonal sensitivity: the phenomenon and its why and when. *Journal of Personality and Social Psychology*, *97*(5), 835.

Sidanius, J., & Pratto, F. (2004). Social dominance theory: A new synthesis. In *Political psychology* (pp. 315–332). Psychology Press.

Sturm, R. E., & Antonakis, J. (2015). Interpersonal power: A review, critique, and research agenda. *Journal of Management*, *41*(1), 136–163.

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In *Advances in experimental social psychology* (Vol. 25, pp. 331–376). Elsevier.

Thomas-Dean, F. (2022). *Department chairs,Äô personal power, positional power, and faculty job satisfaction in higher education settings: A moderation of transformational leadership* (Unpublished doctoral dissertation). Louisiana State University in Shreveport.

Tjosvold, D., Coleman, P. T., & Sun, H. F. (2003). Effects of organizational values on leaders' use of informational power to affect performance in china. *Group Dynamics: Theory, Research, and Practice*, *7*(2), 152.

Tost, L. P. (2015). When, why, and how do powerholders ,Äúfeel the power,Äù? examining the links between structural and psychological power and reviving the connection between power and responsibility. *Research in Organizational Behavior*, *35*, 29–56.

Wang, F., & Sun, X. (2016). Absolute power leads to absolute corruption? impact of power on corruption depending on the concepts of power one holds. *European Journal of Social Psychology*, *46*(1), 77–89.

Willis, G. B., Guinote, A., & Rodríguez-Bailón, R. (2010). Illegitimacy improves goal pursuit in powerless individuals. *Journal of Experimental Social Psychology*, *46*(2), 416–419.

Wiltermuth, S. S., & Flynn, F. J. (2013). Power, moral clarity, and punishment in the workplace. *Academy of Management Journal*, *56*(4), 1002–1023.

Winterich, K. P., & Zhang, Y. (2014). Accepting inequality deters responsibility: How power distance decreases charitable behavior. *Journal of Consumer Research*, *41*(2), 274–293.

Zhong, C.-B., Magee, J. C., Maddux, W. W., & Galinsky, A. D. (2006). Power, culture, and action: Considerations in the expression and enactment of power in east asian and western societies. In *National culture and groups.* Emerald Group Publishing Limited.

Zimbardo, P. G. (1973). *On the ethics of intervention in human psychological research: With special reference to the stanford prison experiment.* Elsevier Science.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98.

Appendix A

A.1

In this appendix, we report pre-study results for the full sample, not excluding participants who failed our attention and understanding checks . The resulting sample is made up of 105 participants. It is 61% male and has mean age 38 years old (SD 11.84). It is 71% employed full-time, 18% unemployed and the rest employed part-time.

Results show that 26% of participants chose to misreport their die roll, with the remaining 74% reporting their die roll accurately. The average bonus distributed was $0.29 per participant, while it would have been $0.23 had all participants reported their die roll truthfully.

A.2

In this appendix, we report pre-test results for the full sample, not excluding participants who failed our attention and understanding checks (Table 7). The resulting sample is made up of 393 participants. It is 61% male and has mean age 40 years old (SD 11.71). It is 74% employed full-time, 13% unemployed and the rest employed part-time. Recall that in this pre-test, participants first went through our power manipulations, before being asked to indicate to what extent their role in this study made them feel various emotions. These included how much participants felt in charge, positive, guilty, responsible, dominant, powerful, inspired, negative, authoritative, in control, determined, influential, respected, self-centered, concerned about others, opportunistic, accountable and egoistic. Participants had to report their emotions on a 7-Point Likert scale, from strongly disagree to strongly agree.

In the punishment power condition, participants felt more powerful (M = 4.67, SD = 1.66) than participants in the neutral condition (M = 3.50, SD = 1.75), p = .00. In the same way, participants in the reward power condition felt more powerful (M = 4.89, SD = 1.60) than participants in the neutral condition (M = 3.50, SD = 1.75), p = 0.00. Lastly, participants in the reward power condition did not feel more or

less powerful (M = 4.89, SD = 1.60) than participants in the punishment power condition (M = 4.67, SD = 1.66), p = 0.40.

Appendix B

B.1

In this appendix, we report main study results for the full sample, not excluding participants who failed our attention and understanding checks (Table 8). The resulting sample across both batches of data, for which we elicited the dependent variable, is made up of 753 participants. It is 54% male and has mean age 40 years old (SD 11.43). It is 71% employed full-time, 15% unemployed and the rest employed part-time. We first show how the different power manipulations influence allocation decisions in Figure 7. We report that participants in both high power conditions choose the prosocial allocation less often than participants in the neutral power condition. This is not in line with Hypothesis 2, stating that reward power increases prosocial behavior. In fact, participants endowed with reward power actually select increasingly the antisocial allocation, compared to people in the neutral power condition. However, the results displayed in Figure 7 are in line with Hypothesis 1, stating that punishment power increases antisocial behavior.

Next, we report OLS regressions showing the impact of our power manipulations on choices in the allocation task. Table 8 reports the impact of power on the prosocial (Columns 1 and 2), antisocial (Columns 3 and 4) and default (Columns 5 and 6) allocations. As all the dependent variables are binary measures, the estimated coefficients reported in the table directly correspond to changes in the likelihood of choosing the given allocation. Column (1) of Table 8 provides the effect of the power manipulations on the likelihood of choosing the prosocial allocation, keeping data from the power conditions separate and not including control variables. In line with previous results, we find that punishment power reduces the likelihood of choosing the prosocial allocation (p = 0.01), providing support for Hypothesis 1. In the same way, reward power reduces the likelihood of choosing the prosocial allocation (p = 0.02), in contrast with

Hypothesis 2. Overall the effects of both high power manipulations go in the same direction, as power reduces the likelihood of choosing the prosocial allocation. Column (2) provides the effect of both power manipulations on prosocial behavior, keeping data from the power conditions separate and controlling for age and gender. We report that punishment power reduces the likelihood of choosing the prosocial allocation ($p = 0.01$), again in line with Hypothesis 1. However reward power also decreases the likelihood of choosing the prosocial allocation ($p = 0.01$), not showing support for Hypothesis 2. As for the model without controls, we find that the behavioral effects of both the reward and punishment power manipulations are going in the same direction, since power overall reduces the likelihood of choosing the prosocial allocation.

Column (3) of Table 8 shows the effect of the power manipulations on the likelihood of choosing the antisocial allocation, keeping data from the high power treatments separate and not including control variables. In line with Hypothesis 1, punishment power increases the likelihood of choosing the antisocial allocation ($p = 0.32$) though the effect size is small. In the same way and not showing support for Hypothesis 2, reward power increases the likelihood of choosing the antisocial allocation, though the effect size is small ($p = 0.29$). The effects of both power manipulations go in the same direction, as power overall increases the likelihood of choosing the antisocial allocation although not significantly. Column (4) next provides the effect of both power manipulations on the antisocial allocation, keeping data from the power conditions separate and controlling for age and gender. Again, we find that punishment power increases the likelihood of choosing the antisocial allocation ($p = 0.32$) though the effect size is small, in line with Hypothesis 1. In the same way, reward power increases the likelihood of choosing the antisocial allocation ($p = 0.24$), again in contrast with Hypothesis 2. As for the model without controls, we find that the behavioral effects of both power types go in the same direction, as power overall increases the likelihood of choosing the antisocial allocation, although the effect sizes are small.

Column (5) provides the effect of both power manipulations on the likelihood of choosing the

default allocation, keeping data from the power conditions separate and not including control variables. In this model, punishment power increases to a small extent the likelihood of choosing the default allocation ($p = 0.17$). In the same way, reward power increases to a small extent the likelihood of choosing the default allocation ($p = 0.33$). Moreover both effects go in the same direction, as power overall increases the likelihood of choosing the default allocation, although the effect size is very small. This is consistent with participants in both power conditions shifting away from the prosocial allocation. Column (6) provides the effect of both power manipulations on the default allocation, keeping data from the power conditions separate and controlling for age and gender. Again, we find that punishment power increases to a small extent the likelihood of choosing the default allocation ($p = 0.18$). In the same way, reward power increases to a small extent the likelihood of choosing the default allocation ($p = 0.36$). As for the model without controls, we report that the behavioral effects of both power types are going in the same direction, since power overall increases the likelihood of choosing the default allocation, though the effect sizes are small.

Taken together, these results show that participants endowed with power shift away from the prosocial allocation towards the antisocial and default allocations, regardless of whether they experience the punishment or reward power manipulation, in line with results for the sample excluding participants who failed our attention and understanding checks.

B.2

In this appendix, we analyze the effects of the power manipulations on social surplus, charity and participants' payoffs first looking at the effects of both power manipulations in isolation before pooling them. Table 9 displays OLS estimations in which we first regress social surplus on indicator variables for the power conditions. As a reminder, social surplus is 100 points if the participant selects the default allocation which splits the initial surplus equally between him (50 points) and the charity (50 points). Social surplus is 90 points if the participant selects the antisocial allocation which gives him a larger payoff

(80 points) at the expense of the charity (10 points). Lastly social surplus is 120 points if the participant chooses the prosocial allocation decreasing his payoff (40 points) to the benefit of the charity (80 points). 10 points correspond to $0.4.

Column (1) of Table 9 provides the effect of the power manipulations on social surplus, keeping data from the two different high power conditions separate. We find that punishment power reduces social surplus ($p = 0.01$), in line with Hypothesis 1 stating that punishment power will increase antisocial behavior. In contrast with Hypothesis 2, we find that reward power also reduces social surplus ($p = 0.03$). In fact, there is no significant difference in the two effects stemming from both high power manipulations ($F(1,750) = 0.08$, $p = 0.78$), reducing social surplus. Column (2) displays the impact of both power manipulations on social surplus, considering both power conditions in isolation and adding controls for age and gender. Again we find that reward power marginally reduces social surplus ($p = 0.02$), which is not in line with Hypothesis 2. Yet we find support for Hypothesis 1 as punishment power also marginally reduces social surplus ($p = 0.02$). Next we analyze the effects of both power manipulations on social surplus, pooling their effects.

Column (3) of Table 9 reports the effect of the power manipulations on social surplus, pooling the data of both high power manipulations. We observe that being in a power condition significantly decreases social surplus in comparison to being in the neutral power condition. In fact, we find a statistically significant, negative effect of power on social surplus ($p = 0.01$). Column (4) also reports the effect of the power manipulations on social surplus, pooling the data of both power manipulations and adding controls for age and gender. Again we find a statistically significant, negative effect of power on social surplus ($p = 0.01$).

Overall the estimations reported in Columns (1) to (4) of Table 9 show that both the reward and punishment power manipulations have negative effects on social surplus, increasing antisocial behavior. Next we show the effects of our power manipulations on charity payoffs. Column (5) provides the effect of

the power manipulations on charity payoffs, keeping data from the two different high power conditions separate. We find that punishment power reduces charity payoffs ($p = 0.05$), in line with Hypothesis 1 stating that punishment power will increase antisocial behavior. In contrast with Hypothesis 2, we find that reward power also reduces charity payoffs ($p = 0.06$). In fact, there is no significant difference in the two effects stemming from both high power manipulations ($F(1,750) = 0.02$, $p = 0.88$), reducing charity payoffs. Column (6) displays the impact of both power manipulations on charity payoffs, considering both power conditions in isolation and adding controls for age and gender. Again we find that reward power marginally reduces charity payoffs, $p = 0.05$, which is not in line with Hypothesis 2. Yet we find support for Hypothesis 1 as punishment power also marginally reduces charity payoffs ($p = 0.05$). Next we analyze the effects of both power manipulations on charity payoffs, pooling their effects.

Column (7) of Table 9 reports the effect of the power manipulations on charity payoffs, pooling the data of both high power manipulations. We observe that being in a power condition significantly decreases charity payoffs, in comparison to being in the neutral power condition. In fact, we find a statistically significant, negative effect of power on charity payoffs ($p = 0.03$). Column (8) also reports the effect of the power manipulations on charity payoffs, pooling the data of both power manipulations and adding controls for age and gender. Again we find a statistically significant, negative effect of power on charity payoffs ($p = 0.03$). Overall the estimations reported in Columns (5) to (8) of Table 9 show that both the reward and punishment power manipulations have negative effects on charity payoffs, increasing antisocial behavior. Next we show the effects of our power manipulations on participants payoffs.

Column (9) provides the effect of the power manipulations on participants payoffs, keeping data from the two different high power conditions separate. We find that punishment power increases participants payoffs ($p = 0.12$), in line with Hypothesis 1 stating that punishment power will increase antisocial behavior. In contrast with Hypothesis 2, we find that reward power also increases participants payoffs ($p = 0.14$). In fact, there is no significant difference in the two effects stemming from both high

power manipulations (F(1,750) = 0.00, p = 0.95), increasing participants payoffs. Column (10) displays the impact of both power manipulations on participants payoffs, considering both power conditions in isolation and adding controls for age and gender. Again we find that reward power marginally increases participants payoffs (p = 0.11) which is not in line with Hypothesis 2. Yet we find support for Hypothesis 1 as punishment power also marginally increases participants payoffs (p = 0.13). Next we analyze the effects of both power manipulations on participants payoffs, pooling their effects.

Column (11) of Table 9 reports the effect of the power manipulations on participants payoffs, pooling the data of both high power manipulations. We observe that being in a power condition increases participants payoffs, in comparison to being in the neutral power condition. In fact, we find a marginally significant positive effect of power on participants payoffs (p = 0.08). Column (12) also reports the effect of the power manipulations on participants payoffs, pooling the data of both power manipulations and adding controls for age and gender. Again we find a marginally significant, positive effect of power on participants payoffs (p = 0.07).

B.3

In this appendix, we report the emotions felt by participants in the second batch of the main study (Table 10), without excluding the participants who failed our attention and understanding checks. Recall that in this batch of data, participants first went through our power manipulations, before being asked to indicate to what extent their role in this study made them feel various emotions. These included how much participants felt in charge, positive, guilty, responsible, dominant, powerful, inspired, negative, authoritative, in control, determined, influential, respected, self-centered, concerned about others, opportunistic, accountable and egoistic. Participants had to report their emotions on a 7-Point Likert scale, from strongly disagree to strongly agree.

B.4

We find that including a manipulation check before asking about the study aim increases the likelihood of guessing the study aim (in the form of "this study is about the relation between X and Y"), hence increasing the potential for demand effects.

The correct X was coded as including punishment, reward, power, control, authority and the correct Y was coded as including greed, selfishness, altruism, generosity, fairness, honesty, morality, cheating, unethicality, responsibility and efficiency.

In the first batch of data, where the study design included the power manipulation, followed by the elicitation of the dependent variable, the elicitation of the study aim and the manipulation check, 2% of participants guessed correctly the X (2 out of 119 participants), 10% of participants guessed correctly the Y (12 out of 119 participants) and 1% of participants guessed correctly the relation between X and Y (1 out of 119 participants).

In the second batch of data, where the study design included the power manipulation, followed by the elicitation the dependent variable, the manipulation check and then the elicitation of the study aim, 16% of participants guessed correctly the X (54 out of 345 participants), 28% of participants guessed correctly the Y (95 out of 345 participants) and 7% of participants guessed correctly the relation between X and Y (24 out of 345 participants).

Footnotes

[1]In line with this hypothesis, studies have found that individuals with a communal orientation are more likely to behave in socially responsible ways when primed with power, while individuals with an exchange orientation behave in a more self-serving way when primed with power (Chen, 2001). In addition, increased power can lead to greater interpersonal sensitivity, but only among individuals with an empathic leadership style, rather than an egoistic one (Schmid Mast et al., 2009).

[2]There was one attention check. The item of the attention check was "I sleep less than two hours per night". This type of attention check is categorized as a logical statement (Abbey & Meloy, 2017). Moreover participants had to answer four questions about the experiment's instructions, checking they had understood what the bonus calculation depended on, how many participants were going to be paid directly, how many participants were going to be selected to passively participate in another study and what the bonus structure was.

[3]There was one attention check. The item of the attention check was "I sleep less than two hours per night". This type of attention check is categorized as a logical statement (Abbey & Meloy, 2017). Moreover participants had to answer three questions about the experiment's instructions, checking they had understood what the bonus calculation in the pre-study depended on, how many participants were selected to be shown in this study and what the bonus structure in the pre-study was.

[4]How much participants felt their role in this study made them feel in charge, positive, guilty, responsible, dominant, inspired, negative, authoritative, in control, determined, influential, respected, self-centered, concerned about others, opportunistic, accountable and egoistic

[5]There was one attention check. The item of the attention check was "I sleep less than two hours per night". This type of attention check is categorized as a logical statement (Abbey & Meloy, 2017). Moreover participants had to answer three questions about the experiment's instructions, checking they had understood what the bonus calculation in the pre-study depended on, how many participants were selected

to be shown in this study and what the bonus structure in the pre-study was.

[6]We hypothesized that the presence of a manipulation check would increase the potential for demand effects. In other words, we expected participants to be more likely to guess the study's aim correctly if they were presented with a manipulation check before, as opposed to after, being asked about the study's aim. To test this hypothesis, we collected data in two batches in order to vary the order in which the manipulation check appeared in our study. The participants who were part of the first batch of data collected started the study by going through the power manipulation and the allocation task, then they were asked what they thought this study was about, before taking part in the manipulation check. In contrast, participants in the second batch of data went through the power manipulation and the allocation task, then they went through the manipulation check before being asked what they thought this study was about. Lastly, participants from both batches had to report which allocation they selected in the task and why they made this choice.

[7]The effects of both our power manipulations on social surplus are small and negative. Punishment power: (2) $p = 0.06$, reward power: (1) $p = 0.12$, (2) $p = 0.09$), both high power conditions pooled: (4) $p = 0.04$.

[8]The effects of both our power manipulations on charity payoffs are small and negative. Punishment power: (5) $p = 0.07$, (6) $p = 0.10$, reward power: (5) $p = 0.17$, (6) $p = 0.13$), both high power conditions pooled: (7) $p = 0.06$, (8) $p = 0.06$. The effects of both our power manipulations on participants payoffs are small and positive. Punishment power: (9) $p = 0.12$, (10) $p = 0.15$, reward power: (9) $p = 0.24$, (10) $p = 0.19$), both high power conditions pooled: (11) $p = 0.11$, (12) $p = 0.11$.

[9]The share of participants choosing the default allocation and further reporting doing so for fairness concerns increases in the punishment (26.9% of participants) and reward conditions (28.4% of participants) compared to the neutral condition (24.5% of participants). The share of participants choosing the default allocation and indicating no reasons for their choice also increases in the punishment (5% of participants)

and reward conditions (7.8% of participants) compared to the neutral condition (4.9% of participants). In the same way, the share of participants choosing the antisocial allocation and further reporting doing so for financial reasons increases in the punishment (26.9% of participants) and reward conditions (25.5% of participants) compared to the neutral condition (22.7% of participants). The share of participants choosing the antisocial allocation and reporting disliking the charity also increases in the punishment (10% of participants) and reward conditions (7.8% of participants) compared to the neutral condition (7.4% of participants) . However, the share of participants choosing the prosocial allocation and further reporting doing so to increase social surplus increases in the punishment (5% of participants) and reward conditions (6.4% of participants) compared to the neutral condition (4.9% of participants).

[10]The effects of both our power manipulations on the share of participants choosing the default allocation and further reporting doing so for fairness concerns are small and positive. Punishment power: (1) $p = 0.59$, reward power: (1) $p = 0.43$. The effects of both our power manipulations on the share of participants choosing the default allocation and further indicating no reasons for their choice are small and positive. Punishment power: (2) $p = 0.96$, reward power: (2) $p = 0.31$. The effects of both our power manipulations on the share of participants choosing the default allocation and further indicating other reasons for their choice are small and negative. Punishment power: (3) $p = 0.70$, reward power: (3) $p = 0.25$. The effects of both our power manipulations on the share of participants choosing the antisocial allocation and further saying they did so for financial reasons are small and positive. Punishment power: (4) $p = 0.52$, reward power: (4) $p = 0.53$. The effects of both our power manipulations on the share of participants choosing the antisocial allocation and further reporting doing so for financial reasons are small and positive. Punishment power: (4) $p = 0.52$, reward power: (4) $p = 0.53$. The effects of both our power manipulations on the share of participants choosing the antisocial allocation and saying they did so for their dislike of the charity are small and positive. Punishment power: (5) $p = 0.32$, reward power: (5) $p = 0.90$. The effects of both our power manipulations on the share of participants choosing the antisocial allocation

and further reporting doing so for other reasons are small and positive. Punishment power: (6) p = 0.61, reward power: (6) p = 0.54. The effects of both our power manipulations on the share of participants choosing the prosocial allocation and further indicating doing so to help are small and negative. Reward power: (7) p = 0.54. The effects of both our power manipulations on the share of participants choosing the prosocial allocation and further saying they did so to increase social surplus are small and positive. Punishment power: (8) p = 0.98, reward power: (8) p = 0.52. The effects of both our power manipulations on the share of participants choosing the prosocial allocation and further indicating doing so for other reasons are small and negative. Punishment power: (8) p = 0.39, reward power: (8) p = 0.07.

[11]We focus on the batch of data where the manipulation check was performed before participants were asked what they thought the study aim was. This study design was used solely to collect the second batch of data, as participants reported their emotions after they were asked what they thought the study aim was in the first batch of data collected. Since the scale used to measure felt power in the first batch of data collected was different to that used in the pre-test and in the second batch of data in the main study, we do not report the results for the first batch of data.

[12]Here we refer to the full sample of the second batch of data collected.

[13]Participants in the reward power condition report feeling significantly more in charge (M = 5.66, SD = 1.12) than participants in the neutral power condition (M = 2.97, SD = 1.66), p = 0.00. In the same way, participants in the punishment power condition feel significantly more in charge (M = 5.43, SD = 1.39) than participants in the neutral power condition (M = 2.97, SD = 1.66), p = 0.00. In addition, participants in the reward power condition feel significantly more positive (M = 5.25, SD = 1.24) than participants in the neutral power condition (M = 4.57, SD = 1.21), p = 0.00 and participants in the punishment power condition also feel significantly more positive (M = 4.97, SD = 1.39) than participants in the neutral power condition (M = 4.57, SD = 1.21), p = 0.02. Participants in the reward power condition additionally feel significantly more responsible (M = 5.68, SD = 1.15) than participants in the neutral power condition (M =

3.54, SD = 1.88), p = 0.00 and participants in the punishment power condition also report feeling significantly more responsible (M = 5.68, SD = 1.22) than participants in the neutral power condition (M = 3.54, SD = 1.88), p = 0.00. In the same way, participants in the reward power condition feel significantly more dominant (M = 4.70, SD = 1.53) than participants in the neutral power condition (M = 2.72, SD = 1.56), p = 0.00 and participants in the punishment power condition also report feeling significantly more dominant (M = 4.77, SD = 1.53) than participants in the neutral power condition (M = 2.72, SD = 1.56), p = 0.00. Moreover, participants in the reward power condition feel significantly more authoritative (M = 5.03, SD = 1.51) than participants in the neutral power condition (M = 2.63, SD = 1.53), p = 0.00 and participants in the punishment power condition also report feeling significantly more authoritative (M = 4.87, SD = 1.58) than participants in the neutral power condition (M = 2.63, SD = 1.53), p = 0.00. Participants in the reward power condition additionally feel significantly more in control (M = 5.58, SD = 1.11) than participants in the neutral power condition (M = 3.14, SD = 1.71), p = 0.00 and participants in the punishment power condition also report feeling significantly more in control (M = 5.43, SD = 1.34) than participants in the neutral power condition (M = 3.14, SD = 1.71), p = 0.00. In the same way, participants in the reward power condition feel significantly more determined (M = 5.01, SD = 1.24) than participants in the neutral power condition (M = 4.24, SD = 1.59), p = 0.00 and participants in the punishment power condition also feel significantly more determined (M = 4.90, SD = 1.59) than participants in the neutral power condition (M = 4.24, SD = 1.59), p = 0.00. In addition, participants in the reward power condition report feeling significantly more influential (M = 5.49, SD = 1.19) than participants in the neutral power condition (M = 3.12, SD = 1.66), p = 0.00 and participants in the punishment power condition also feel significantly more influential (M = 5.27, SD = 1.52) than participants in the neutral power condition (M = 3.12, SD = 1.66), p = 0.00. Moreover, participants in the reward power condition report feeling significantly more respected (M = 4.15, SD = 1.49) than participants in the neutral power condition (M = 3.28, SD = 1.55), p = 0.00 and participants in the punishment power condition also report feeling

significantly more respected (M = 4.14, SD = 1.67) than participants in the neutral power condition (M = 3.28, SD = 1.55), p = 0.00. In the same way, participants in the reward power condition feel significantly more concerned (M = 4.71, SD = 1.47) than participants in the neutral power condition (M = 4.01, SD =1.55), p = 0.00 and participants in the punishment power condition also report feeling significantly more concerned (M = 4.71, SD = 1.39) than participants in the neutral power condition (M = 4.01, SD = 1.55), p = 0.00. In addition, participants in the reward power condition feel significantly more accountable (M = 4.91, SD = 1.59) than participants in the neutral power condition (M = 3.71, SD = 1.84), p = 0.00 and participants in the punishment power condition also report feeling significantly more accountable (M = 5.06, SD = 1.42) than participants in the neutral power condition (M = 3.71, SD = 1.84), p = 0.00.

Figure 1: Summary of Results in Pre-Study

	Report						Total
Actual die roll	1	2	3	4	5	6	
1	6	0	0	0	4	0	10
2	0	9	0	0	3	0	12
3	0	0	6	0	3	0	9
4	0	0	0	7	1	0	8
5	0	0	0	0	6	0	6
6	0	0	1	0	3	6	10
Total	6	9	7	7	20	6	55

Figure 2: Allocation Decisions, Per Power Condition

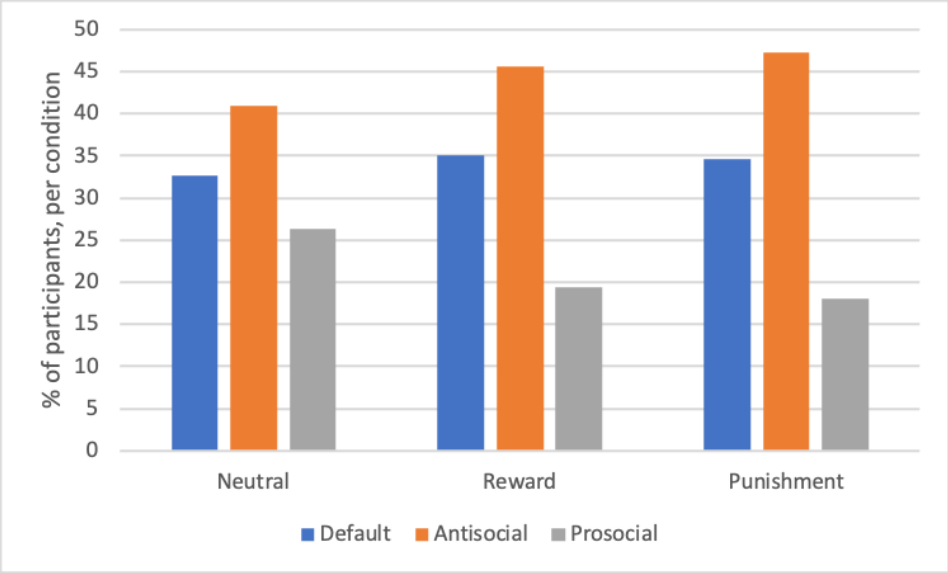


Figure 3: Social Surplus, Charity and Participants Payoffs

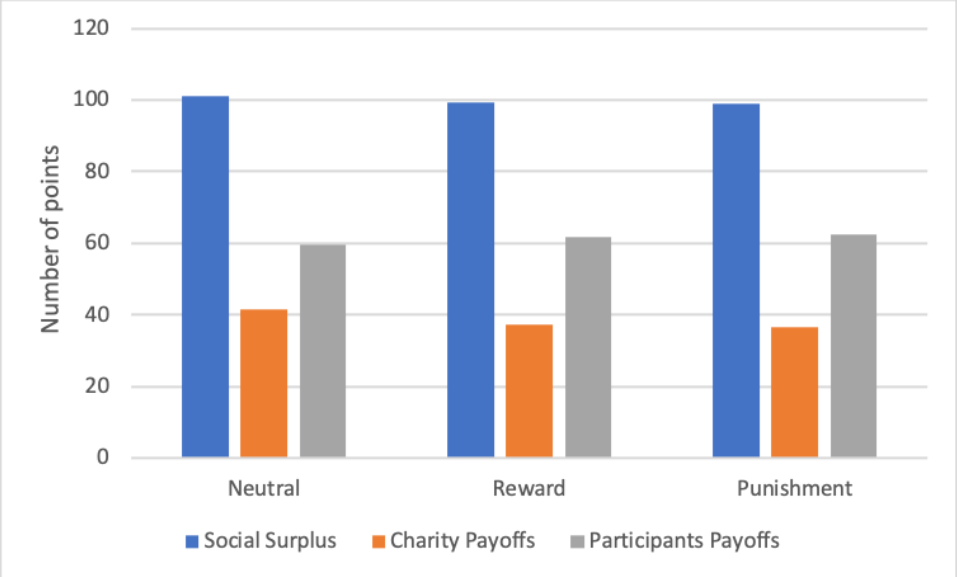
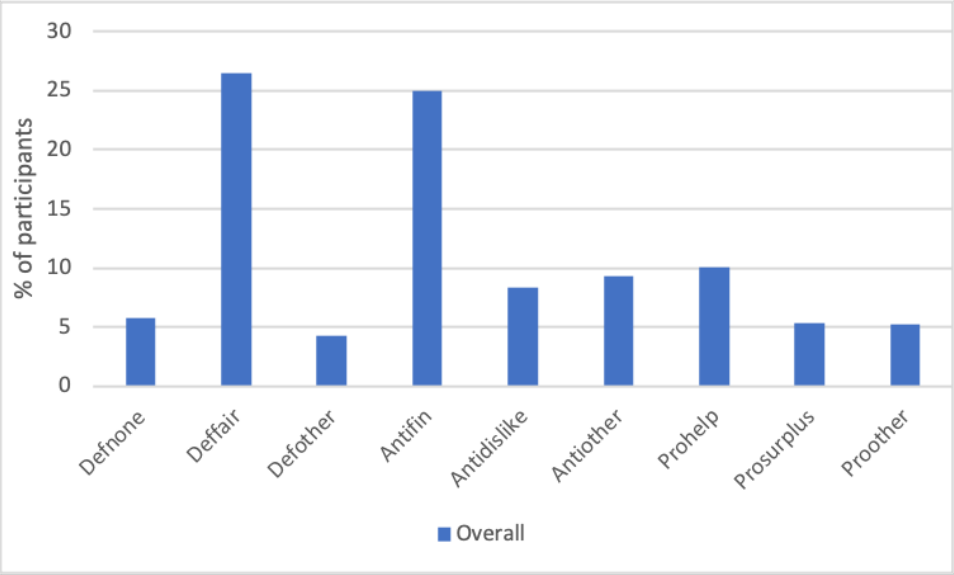
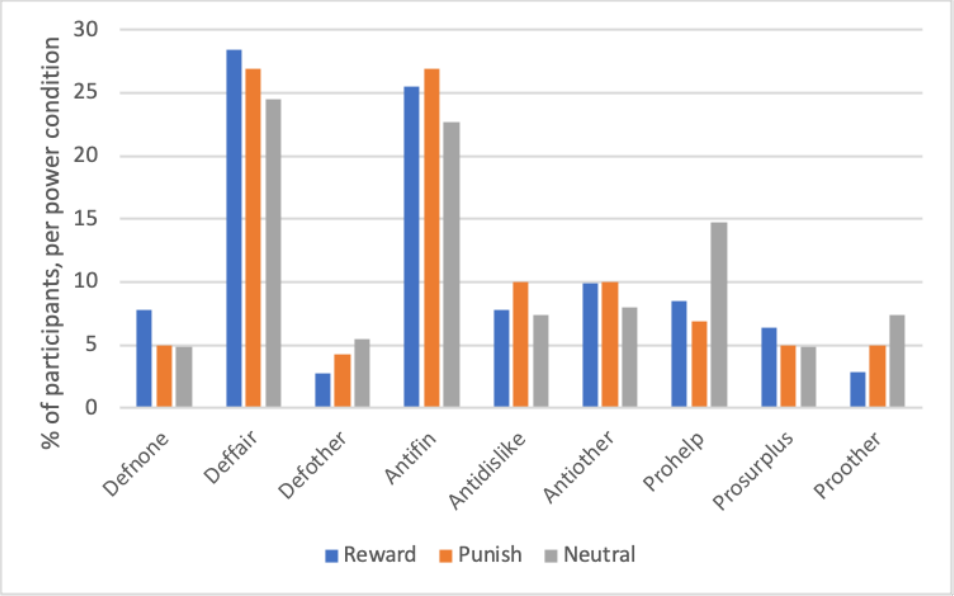


Figure 4: Combined Choice-Choice Motives



Note. Defnone = participants who chose the default allocation and reported no reason for their allocation. Deffair = participants who chose the default allocation and reported doing so out of fairness. Defother = participants who chose the default allocation and reported doing so for other reasons. Antifin = participants who chose the antisocial allocation and reported doing so for financial reasons. Antidislike = participants who chose the antisocial allocation and reported doing so as they disliked the charity. Antiother = participants who chose the antisocial allocation and reported doing so for other reasons. Prohelp = participants who chose the prosocial allocation and reported doing so to help the charity. Prosurplus = participants who chose the prosocial allocation and reported doing so to maximize social surplus. Proother = participants who chose the prosocial allocation and reported doing so for other reasons.

Figure 5: Combined Choice-Choice Motives, Per Power Condition



Note. Defnone = participants who chose the default allocation and reported no reason for their allocation. Deffair = participants who chose the default allocation and reported doing so out of fairness. Defother = participants who chose the default allocation and reported doing so for other reasons. Antifin = participants who chose the antisocial allocation and reported doing so for financial reasons. Antidislike = participants who chose the antisocial allocation and reported doing so as they disliked the charity. Antiother = participants who chose the antisocial allocation and reported doing so for other reasons. Prohelp = participants who chose the prosocial allocation and reported doing so to help the charity. Prosurplus = participants who chose the prosocial allocation and reported doing so to maximize social surplus. Proother = participants who chose the prosocial allocation and reported doing so for other reasons.

Figure 6: Emotions Reported (Main Study 2nd Batch)

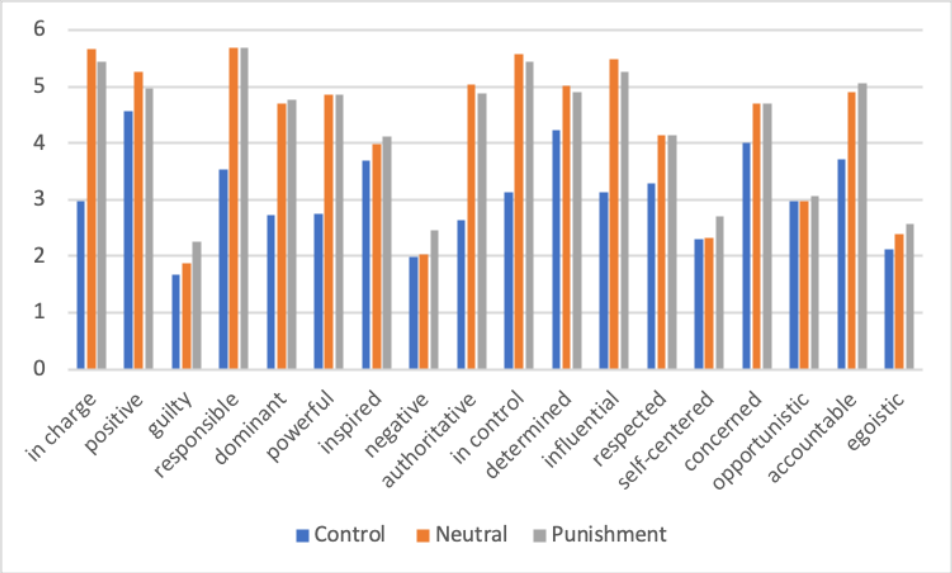


Figure 7: Allocation Decisions, Per Power Condition, Full Sample

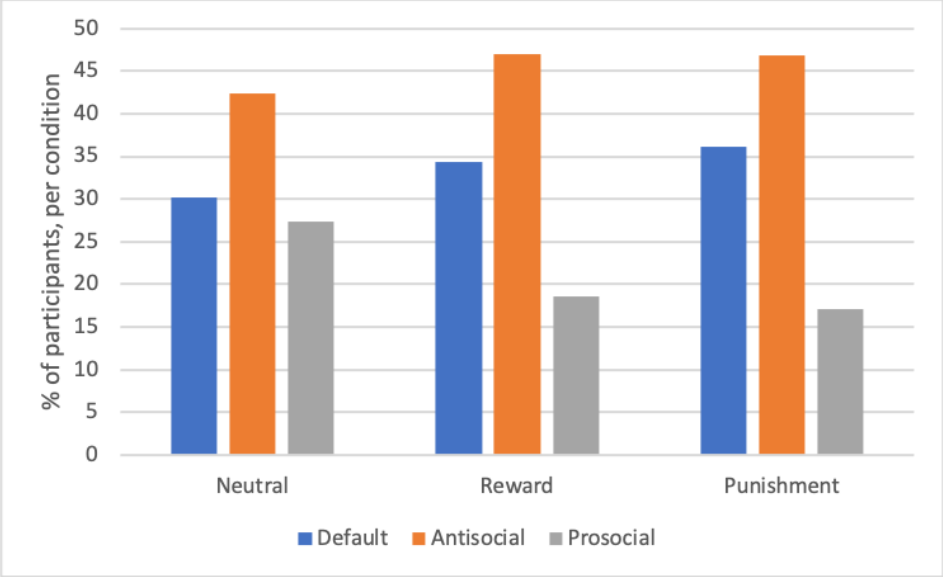


TABLE 1: *Emotions Reported: Means and T-Tests (Pre-Test)*

	RewardMean	PunishmentMean	NeutralMean	TTestRewardNeutral	TTestPunishNeutral	TTestRewardPunish
	(1)	(2)	(3)	(4)	(5)	(6)
InCharge	5.32	5.30	3.34	-2.23***	-2.21***	-0.02
				(0.00)	(0.00)	(0.91)
Positive	5.31	4.63	4.85	-0.81***	-0.13	-0.68**
				(0.00)	(0.57)	(0.00)
Guilty	1.95	2.16	1.82	-0.26	-0.47**	0.21
				(0.14)	(0.01)	(0.33)
Responsible	5.56	5.50	3.733	-2.10***	-2.04***	-0.060
				(0.00)	(0.00)	(0.76)
Dominant	4.44	4.44	3.20	-1.48***	-1.47***	-0.00
				(0.00)	(0.00)	(0.99)
Powerful	4.66	4.61	3.19	-1.70***	-1.65***	-0.05
				(0.00)	(0.00)	(0.85)
Inspired	4.05	3.71	3.63	-0.68*	-0.35	-0.34
				(0.01)	(0.19)	(0.20)
Negative	2.14	2.59	2.04	-0.25	-0.70***	0.44*
				(0.17)	(0.00)	(0.04)
Authoritative	4.94	4.64	3.24	-1.94***	-1.64***	-0.30
				(0.00)	(0.00)	(0.22)
InControl	5.48	5.28	3.49	-2.25***	-2.05***	-0.20
				(0.00)	(0.00)	(0.33)
Determined	5.08	5.03	4.26	-1.14***	-1.09***	-0.051
				(0.00)	(0.00)	(0.83)
Influential	5.30	5.33	3.47	-2.09***	-2.12***	0.03
				(0.00)	(0.00)	(0.88)
Respected	4.00	3.89	3.87	-0.41	-0.31	-0.11
				(0.12)	(0.23)	(0.68)
SelfCentered	2.45	2.52	2.68	0.03	-0.04	0.07
				(0.89)	(0.88)	(0.76)
Concerned	4.75	4.52	4.24	-0.82**	-0.60*	-0.23
				(0.00)	(0.02)	(0.33)
Opportunistic	2.73	3.05	2.88	-0.06	-0.38	0.33
				(0.83)	(0.12)	(0.19)
Accountable	4.93	4.85	3.93	-1.29***	-1.21***	-0.08
				(0.00)	(0.00)	(0.75)
Egoistic	2.25	2.45	2.27	-0.15	-0.35	0.20
				(0.47)	(0.09)	(0.34)

Note. Participants were asked to indicate to what extent their role in this study made them feel the emotions listed on a 7-Point Likert scale, from strongly disagree to strongly agree. * $p < .05$. ** $p < .01$.

TABLE 2: *Linear Regression Analysis: Allocation Choice (Prosocial, Antisocial, Default) on Power*

	Prosocial	Prosocial	Antisocial	Antisocial	Default	Default
	(1)	(2)	(3)	(4)	(5)	(6)
Reward	-0.07	-0.07	0.05	0.05	0.02	0.02
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Punish	-0.08*	-0.08*	0.06	0.06	0.02	0.02
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Constant	0.26**	0.13*	0.41**	0.62**	0.33**	0.25**
	(0.03)	(0.07)	(0.03)	(0.08)	(0.03)	(0.08)
Controls for age and gender	No	Yes	No	Yes	No	Yes
R^2	0.01	0.02	0.00	0.03	0.00	0.01
Observations	584	584	584	584	584	584

Note. Dependent variables are prosocial (Models 1 to 2), antisocial (Models 3 to 4) and default (Models 5 to 6). Prosocial was coded as 1 if the participant chose the prosocial allocation, 0 otherwise. Antisocial was coded as 1 if the participant chose the antisocial allocation, 0 otherwise. Default was coded as 1 if the participant chose the default allocation, 0 otherwise. Reward = whether the participant was allocated to the reward power condition. Reward was equal to 1 if the participant was allocated to the reward power condition, 0 otherwise. Punish = whether the participant was allocated to the punishment power condition. Gender was coded as dummy variables for female and othergender. Unstandardized regression coefficients are shown ; robust standard errors in parentheses.* $p < .05$. ** $p < .01$.

TABLE 3: *Linear Regression Analysis: Social Surplus, Charity and Participants Payoffs on Power*

	Surplus	Surplus	Surplus	Surplus	Charity	Charity	Charity	Charity	Participants	Participants	Participants	Participants
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Reward	-1.84	-1.99			-3.90	-4.31			2.06	2.31		
	(1.16)	(1.16)			(2.84)	(2.82)			(1.76)	(1.74)		
Punish	-2.28*	-2.13			-4.98	-4.57			2.70	2.45		
	(1.13)	(1.13)			(2.77)	(2.75)			(1.71)	(1.70)		
RewardPunish			-2.07*	-2.06*			-4.47	-4.45			2.40	2.38
			(0.99)	(0.98)			(2.41)	(2.39)			(1.49)	(1.48)
Constant	101.17**	96.41**	101.17**	96.39**	41.51**	29.05**	41.51**	29.03**	59.66**	67.35**	59.66**	67.37**
	(0.80)	(1.86)	(0.80)	(1.85)	(1.94)	(4.53)	(1.94)	(4.51)	(1.20)	(2.80)	(1.20)	(2.79)
Controls for age and gender	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.01	0.03	0.01	0.03	0.01	0.03	0.01	0.03	0.00	0.03	0.00	0.03
Observations	584	584	584	584	584	584	584	584	584	584	584	584

Note. Dependent variables are social surplus (Models 1 to 4), charity payoffs (Models 5 to 8) and participants payoffs (Models 9 to 12). Social surplus was coded as 100 if the participant chose the default allocation, 120 if the participant chose the efficient prosocial allocation and 90 if the participant chose the inefficient selfish allocation. Participant payoffs were coded as 50 if the participant chose the default allocation, 40 if the participant chose the efficient prosocial allocation and 80 if the participant chose the inefficient selfish allocation.Charity payoffs were coded as 50 if the participant chose the default allocation, 80 if the participant chose the efficient prosocial allocation and 10 if the participant chose the inefficient selfish allocation. Reward = whether the participant was allocated to the reward power condition, 0 otherwise. Punish = whether the participant was allocated to the punishment power condition. Punishment was equal to 1 if the participant was allocated to the punishment power condition, 0 otherwise. Rewardpunish = whether the participant was allocated to the reward or punishment power condition. Rewardpunish was equal to 1 if the participant was allocated to the reward or punishment power condition, 0 otherwise. Gender was coded as dummy variables for female and othergender. Unstandardized regression coefficients are shown ; robust standard errors in parentheses,* $p < .05$. ** $p < .01$.

TABLE 4: *Reason for Allocation Choice*

Reason	1st batch	2nd batch	Both batches
Financial need	34%	25%	27%
Charity dislike	8%	9%	9%
Fairness	29%	29%	29%
Desire to help	11%	13%	13%
Lack of knowledge about methods	3%	2%	2%
Contribution size too small	5%	2%	2 %
Maximize social surplus	6%	5%	6 %
Already donating to UNICEF	0 %	1%	1%
None given	4 %	14%	11%

TABLE 5: *Linear Regression Analysis: Choice-Choice Motive on Power*

	Defnone	Deffair	Defother	Antifin	Antidislike	Antiother	Prohelp	Prosurplus	Proother
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Reward	0.03	0.04	-0.03	0.03	0.00	0.02	-0.07	0.02	-0.05
	(0.03)	(0.05)	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Punish	0.00	0.03	-0.01	0.03	0.03	0.02	-0.08*	0.00	-0.02
	(0.03)	(0.05)	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Constant	0.02	0.24**	-0.00	0.45**	-0.02	0.16**	0.05	0.07	0.02
	(0.04)	(0.08)	(0.04)	(0.08)	(0.05)	(0.05)	(0.06)	(0.04)	(0.04)
Controls for age and gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.01	0.01	0.01	0.02	0.01	0.01	0.04	0.02	0.01
Observations	464	464	464	464	464	464	464	464	464

Note. Defnone = participants who chose the default allocation and reported no reason for their allocation. Deffair = participants who chose the default allocation and reported doing so out of fairness. Defother = participants who chose the default allocation and reported doing so for other reasons. Antifin = participants who chose the antisocial allocation and reported doing so for financial reasons. Antidislike = participants who chose the antisocial allocation and reported doing so as they disliked the charity. Antiother = participants who chose the antisocial allocation and reported doing so for other reasons. Prohelp = participants who chose the prosocial allocation and reported doing so to help the charity. Prosurplus = participants who chose the prosocial allocation and reported doing so to maximize social surplus. Proother = participants who chose the prosocial allocation and reported doing so for other reasons. Reward= whether the participant was allocated to the reward power condition. Reward was equal to 1 if the participant was allocated to the reward power condition, 0 otherwise. Punish= whether the participant was allocated to the punishment power condition. Gender was coded as dummy variables for female and othergender. Unstandardized regression coefficients are shown ; robust standard errors in parentheses.* $p < .05$. ** $p < .01$.

TABLE 6: *Emotions Reported: Means and T-Tests (Main Study 2nd Batch)*

	RewardMean	PunishmentMean	NeutrallMean	TTestRewardNeutral	TTestPunishNeutral	TTestRewardPunish
	(1)	(2)	(3)	(4)	(5)	(6)
InCharge	5.66	5.43	2.97	-2.69***	-2.46***	-0.23
				(0.00)	(0.00)	(0.18)
Positive	5.25	4.97	4.57	-0.68***	-0.41*	-0.27
				(0.00)	(0.02)	(0.13)
Guilty	1.87	2.26	1.68	-0.18	-0.57***	0.39*
				(0.13)	(0.00)	(0.03)
Responsible	5.68	5.68	3.54	-2.14***	-2.14***	-0.00
				(0.00)	(0.00)	(1.00)
Dominant	4.70	4.77	2.72	-1.98***	-2.05***	0.07
				(0.00)	(0.00)	(0.72)
Powerful	4.86	4.85	2.74	-2.12***	-2.11***	-0.01
				(0.00)	(0.00)	(0.96)
Inspired	3.98	4.12	3.69	-0.29	-0.43	0.14
				(0.17)	(0.05)	(0.52)
Negative	2.04	2.46	1.98	-0.05	-0.48**	0.42*
				(0.71)	(0.00)	(0.02)
Authoritative	5.03	4.87	2.63	-2.39***	-2.24***	-0.16
				(0.00)	(0.00)	(0.45)
InControl	5.58	5.43	3.14	-2.44***	-2.29***	-0.15
				(0.00)	(0.00)	(0.36)
Determined	5.01	4.90	4.24	-0.77***	-0.65**	-0.11
				(0.00)	(0.00)	(0.56)
Influential	5.49	5.27	3.12	-2.36***	-2.15***	-0.21
				(0.00)	(0.00)	(0.25)
Respected	4.15	4.14	3.28	-0.88***	-0.86***	-0.02
				(0.00)	(0.00)	(0.94)
SelfCentered	2.33	2.70	2.29	-0.04	-0.41*	0.37*
				(0.82)	(0.02)	(0.05)
Concerned	4.71	4.71	4.01	-0.71***	-0.70***	-0.01
				(0.00)	(0.00)	(0.98)
Opportunistic	2.97	3.06	2.99	0.01	-0.08	0.09
				(0.96)	(0.72)	(0.69)
Accountable	4.91	5.07	3.71	-1.21***	-1.36***	0.15
				(0.00)	(0.00)	(0.45)
Egoistic	2.40	2.56	2.11	-0.29	-0.44**	0.16
				(0.09)	(0.01)	(0.36)

Note. Participants were asked to indicate to what extent their role in this study made them feel the emotions listed on a 7-Point Likert scale, from strongly disagree to strongly agree. * $p < .05$. ** $p < .01$.

TABLE 7: *Emotions Reported: Means and T-Tests (Pre-Test Full Sample)*

	RewardMean	PunishmentMean	ControlMean	TTestRewardControl	TTestPunishControl	TTestRewardPunish
	(1)	(2)	(3)	(4)	(5)	(6)
InCharge	5.43	5.33	3.64	-2.18***	-2.13***	-0.05
				0.00	0.00	-0.77
Positive	5.40	4.75	4.90	-1.04***	-0.43*	-0.61***
				0.00	-0.05	0.00
Guilty	2.03	2.33	1.97	-0.28	-0.60***	0.32
				-0.10	0.00	-0.10
Responsible	5.65	5.51	4.08	-2.01***	-1.91***	-0.09
				0.00	0.00	-0.58
Dominant	4.56	4.49	3.49	-1.45***	-1.41***	-0.04
				0.00	0.00	-0.86
Powerful	4.89	4.67	3.50	-1.77***	-1.59***	-0.18
				0.00	0.00	-0.40
Inspired	4.35	3.84	3.76	-0.99***	-0.53*	-0.47*
				0.00	-0.03	-0.04
Negative	2.17	2.67	2.17	-0.24	-0.76***	0.52**
				-0.16	0.00	-0.01
Authoritative	5.02	4.64	3.52	-1.88***	-1.53***	-0.35
				0.00	0.00	-0.10
InControl	5.57	5.24	3.83	-2.15***	-1.87***	-0.28
				0.00	0.00	-0.11
Determined	5.09	5.02	4.43	-1.14***	-1.11***	-0.03
				0.00	0.00	-0.90
Influential	5.38	5.19	3.65	-2.12***	-1.97***	-0.15
				0.00	0.00	-0.44
Respected	4.15	4.05	4.00	-0.59*	-0.53*	-0.06
				-0.01	-0.02	-0.79
SelfCentered	2.67	2.78	2.79	-0.19	-0.33	0.14
				-0.38	-0.12	-0.52
Concerned	4.88	4.56	4.19	-1.15***	-0.86***	-0.29
				0.00	0.00	-0.15
Opportunistic	3.02	3.29	3.17	-0.20	-0.50*	0.30
				-0.40	-0.03	-0.19
Accountable	5.04	4.98	4.20	-1.30***	-1.28***	-0.02
				0.00	0.00	-0.94
Egoistic	2.42	2.65	2.46	-0.23	-0.48*	0.25
				-0.24	-0.01	-0.19

Note. Participants were asked to indicate to what extent their role in this study made them feel the emotions listed on a 7-Point Likert scale, from strongly disagree to strongly agree. * $p < .05$. ** $p < .01$.

TABLE 8: *Linear Regression Analysis: Allocation Choice (Prosocial, Antisocial, Default) on Power, Full Sample*

	Prosocial	Prosocial	Antisocial	Antisocial	Default	Default
	(1)	(2)	(3)	(4)	(5)	(6)
Reward	-0.09*	-0.09*	0.05	0.05	0.04	0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Punish	-0.10**	-0.10**	0.04	0.04	0.06	0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Constant	0.27**	0.17**	0.42**	0.62**	0.30**	0.20**
	(0.03)	(0.06)	(0.03)	(0.07)	(0.03)	(0.07)
Controls for age and gender	No	Yes	No	Yes	No	Yes
R^2	0.01	0.02	0.00	0.02	0.00	0.01
Observations	753	753	753	753	753	753

Note. Dependent variable in Models 1 and 2 is prosocial. Prosocial was coded as 1 if the participant chose the prosocial allocation, 0 otherwise. Dependent variable in Models 3 and 4 is antisocial. Antisocial was coded as 1 if the participant chose the antisocial allocation, 0 otherwise. Dependent variable in Models 5 and 6 is default allocation. Default was coded as 1 if the participant chose the default allocation, 0 otherwise. Reward = whether the participant was allocated to the reward power condition. Reward was equal to 1 if the participant was allocated to the reward power condition, 0 otherwise. Punish = whether the participant was allocated to the punishment power condition. Gender was coded as dummy variables for female and othergender. Unstandardized regression coefficients are shown ; robust standard errors in parentheses.* $p < .05$. ** $p < .01$.

TABLE 9: *Linear Regression Analysis: Social Surplus, Charity and Participants Payoffs on Power, Full Sample*

	Surplus	Surplus	Surplus	Surplus	Charity	Charity	Charity	Charity	Participants	Participants	Participants	Participants
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Reward	-2.24*	-2.33*			-4.53	-4.80*			2.29	2.46		
	(1.01)	(1.01)			(2.48)	(2.47)			(1.54)	(1.53)		
Punish	-2.52*	-2.46*			-4.90*	-4.78*			2.38	2.32		
	(1.01)	(1.01)			(2.49)	(2.47)			(1.54)	(1.53)		
RewardPunish			-2.38**	-2.40**			-4.72*	-4.79*			2.34	2.39
			(0.88)	(0.87)			(2.15)	(2.14)			(1.34)	(1.32)
Constant	101.25**	97.25**	101.25**	97.24**	41.29**	30.30**	41.29**	30.30**	59.96**	66.95**	59.96**	66.94**
	(0.72)	(1.63)	(0.72)	(1.62)	(1.77)	(3.98)	(1.76)	(3.97)	(1.09)	(2.46)	(1.09)	(2.46)
Controls for age and gender	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.00	0.02	0.00	0.02
Observations	753	753	753	753	753	753	753	753	753	753	753	753

Note. Dependent variables are social surplus (Models 1 to 4), charity payoffs (Models 5 to 8) and participants payoffs (Models 9 to 12). Social surplus was coded as 100 if the participant chose the default allocation, 120 if the participant chose the efficient prosocial allocation and 90 if the participant chose the inefficient selfish allocation. Participant payoffs were coded as 50 if the participant chose the default allocation, 40 if the participant chose the efficient prosocial allocation and 80 if the participant chose the inefficient selfish allocation.Charity payoffs were coded as 50 if the participant chose the default allocation, 80 if the participant chose the efficient prosocial allocation and 10 if the participant chose the inefficient selfish allocation. Reward = whether the participant was allocated to the reward power condition, 0 otherwise. Punish = whether the participant was allocated to the punishment power condition. Punishment was equal to 1 if the participant was allocated to the punishment power condition, 0 otherwise. Rewardpunish = whether the participant was allocated to the reward or punishment power condition. Rewardpunish was equal to 1 if the participant was allocated to the reward or punishment power condition, 0 otherwise. Gender was coded as dummy variables for female and othergender. Unstandardized regression coefficients are shown : robust standard errors in parentheses. * $p < .05$, ** $p < .01$.

TABLE 10: *Emotions Reported: Means and T-Tests (Second Batch Full Sample)*

	RewardMean	PunishmentMean	NeutralMean	TTestRewardNeutral	TTestPunishNeutral	TTestRewardPunish
	(1)	(2)	(3)	(4)	(5)	(6)
InCharge	5.61	5.41	3.05	-2.61***	-2.42***	-0.19
				0.00	0.00	-0.18
Positive	5.27	4.99	4.55	-0.81***	-0.53**	-0.28
				0.00	0.00	-0.07
Guilty	2.05	2.36	1.80	-0.28*	-0.59***	0.31
				-0.04	0.00	-0.06
Responsible	5.59	5.63	3.57	-2.09***	-2.13***	0.04
				0.00	0.00	-0.77
Dominant	4.75	4.77	2.78	-2.02***	-2.04***	0.02
				0.00	0.00	-0.89
Powerful	4.90	4.86	2.81	-2.15***	-2.10***	-0.05
				0.00	0.00	-0.79
Inspired	4.31	4.17	3.71	-0.67***	-0.54**	-0.14
				0.00	-0.01	-0.46
Negative	2.16	2.49	2.05	-0.16	-0.49***	0.33*
				-0.26	0.00	-0.04
Authoritative	5.02	4.88	2.73	-2.34***	-2.20***	-0.15
				0.00	0.00	-0.42
InControl	5.50	5.38	3.18	-2.38***	-2.27***	-0.12
				0.00	0.00	-0.44
Determined	5.06	4.97	4.22	-0.93***	-0.84***	-0.09
				0.00	0.00	-0.60
Influential	5.36	5.22	3.20	-2.22***	-2.08***	-0.14
				0.00	0.00	-0.37
Respected	4.20	4.21	3.35	-0.92***	-0.93***	0.01
				0.00	0.00	-0.97
SelfCentered	2.67	2.76	2.40	-0.32	-0.41*	0.08
				-0.07	-0.01	-0.64
Concerned	4.69	4.76	4.05	-0.72***	-0.80***	0.08
				0.00	0.00	-0.64
Opportunistic	3.37	3.22	3.07	-0.36	-0.20	-0.16
				-0.08	-0.30	-0.44
Accountable	4.92	5.15	3.78	-1.21***	-1.45***	0.24
				0.00	0.00	-0.18
Egoistic	2.66	2.71	2.21	-0.50**	-0.55***	0.05
				0.00	0.00	-0.76

Note. Participants were asked to indicate to what extent their role in this study made them feel the emotions listed on a 7-Point Likert scale, from strongly disagree to strongly agree. * $p < .05$. ** $p < .01$.

The Impact of Family Choices on Wages: New Evidence from Switzerland

Thesis Paper

Oriana de Saint Priest

Department of Organizational Behavior

University of Lausanne

Switzerland, 2022

Contents

Abstract

Family has far reaching impacts, both on our personal life and on our professional life. Existing research on the impact of family on labor income has focused on two phenomena: the male marital wage premium and the motherhood wage penalty. The current study uses the Swiss Household Panel dataset from 2004 to 2016 as Switzerland is a particularly strong test of the marital wage premium and motherhood wage penalty, being a country with a traditional view of gender roles. Our analyses show that there is no impact of either marriage or parenthood on the wages of men. However, employed women suffer from a robust motherhood wage penalty, persisting across time and caused by responsibility for childcare duties. Interestingly, there is no motherhood wage penalty for self-employed women and female managers, in contrast with other employed women. This phenomenon is consistent with managers and self-employed people benefiting from greater work flexibility, while being less subject to employer discrimination. Taken together, these findings reveal novel evidence on the causes of the gender wage gap with implications for policy.

Introduction

Family impacts our personal life widely and it is likely that this effect extends to our professional life too. Previous studies have uncovered two phenomena: the male marital wage premium, referring to the fact that married men earn more than similar single men, and the motherhood wage penalty, describing the fact that mothers earn less than similar women who do not have children. Recent evidence on the marital wage premium reveals that this phenomenon is likely due to the selection of men with higher wage growth into marriage (Ludwig & Bruderl, 2018; Killewald & Lundberg, 2017), rather than to a causal effect of marriage on wages. Yet most existing research is based on data from the United States and there is little evidence on countries with less progressive views of gender roles. We focus our research on Switzerland, one of the most traditional European countries. In fact, women were only recently given the right to vote in certain parts of Switzerland, such as Appenzell where women were allowed to vote in 1990. Relatedly, Switzerland has one of the biggest gender wage gaps in Europe (Eurostat, 2018), also due to the fact that 45% of women work part-time, compared to 11% of men (Info, 2019). It is hard for women to work full-time in Switzerland as the availability of childcare is limited and expensive compared to other countries, while attitudes to full-time work for women are mixed (Bornatici et al., 2021). The present study seeks to test whether a causal marital wage premium remains in this setting. In addition, recent research suggests that having a child causes a woman to earn substantially less, all else constant (Oesch et al., 2017). The present study wishes to go beyond this observation to understand why women earn less when they become mothers, taking advantage of the extensive information on family and work collected in the Swiss Household Panel.

We use data from years 2004 to 2016 of the Swiss Household Panel to investigate the presence of a marital wage premium and motherhood wage penalty. We contribute to existing literature by adding novel evidence on the size and causes of the marital wage premium and motherhood wage penalty in a country with one of the biggest gender wage gaps in Europe. The main analysis focuses on years 2009 to 2016, as

an exogenous fiscal change which happened in 2008 made it more financially attractive for couples to marry. Data from years 2004 to 2008 is used in a subsequent analysis as a check of effects over time. In the main sample from years 2009 to 2016, there is no causal effect of marriage or parenthood on the wages of men, while there is a robust motherhood wage penalty of about 6% per child. This motherhood wage penalty can further be attributed to unpredictable absences from work occurring because of childcare duties. Interestingly, further analyses reveal that female managers and self-employed women do not experience wage variations due to similar childcare duties. Hence, assuming motherhood impacts the average productivity of working women in the same way, it is likely that female managers and self-employed women do not suffer from a motherhood wage penalty because they benefit from different working conditions. In fact, women in these positions generally experience greater work flexibility. Thus it is likely that work flexibility plays a role in explaining the motherhood wage penalty for employed women. Moreover, we find that men are never penalized by childcare duties causing them to leave work unpredictably. Assuming that childcare duties impact on average the productivity of women and men in the same way, this provides evidence that employer discrimination might play a role in explaining the motherhood wage penalty for employed women. At the end of the paper, we run a check of effects over time and focus our analysis on years 2004 to 2008. This analysis confirms that there is no effect of marriage or parenthood on the wages of men in this earlier sample but a robust motherhood wage penalty, of the same magnitude as that in the later sample.

In the next section, we present a survey of existing evidence on the marital wage premium and on the motherhood wage penalty, before describing the dataset used in this study in the following section.

Literature Review

The Male Marital Wage Premium

The male marital wage premium refers to the fact that married men earn more than single men with the same characteristics. Thus, for a marital wage premium to exist, the event of marriage should have a causal effect on the wages of men. Previous research has found that married men earn more than single men in Europe and in the United States. One of the first works on the topic uses cross-sectional data from the General Household Surveys of 1971 and 1976 to establish the effect of marriage on wages in the United Kingdom (Greenhalgh, 1980). The author finds that married men obtain a 10% wage premium over single men, controlling for potential experience, education level, as well as some job characteristics and demographics of the respondents. Using cross-sectional data as well, a meta-study finds that married men earn as much as 40% more than single men in certain European countries (Schoeni, 1995). Simple descriptive analyses of recent datasets also show that married men earn significantly more than single men of the same age (British Household Panel Survey, Swiss Household Panel).

However, these early analyses have strong limitations as they are unable to reveal whether a true marital wage premium, implying a causal effect of marriage on wages, exists. In fact, it is highly likely that OLS regressions suffer from an omitted variables bias, as unmeasured variables affect both marriage probability and earnings potential. For instance, it is plausible that a stable man is more likely to be married and to have a higher labor income. Similarly, men who like to conform to social expectations may be more likely to be married and to have a higher wage. However, these personal characteristics are hard to measure and hence are completely absent from cross-sectional data. One way to correct for this is to use panel data, where the same individuals are surveyed across different years. In fact, assuming that stability and similar unobservable characteristics are constant over a short period of time at the individual level, panel data models including individual fixed effects are able to correct the bias introduced by

unobservables in cross-sectional estimates. This is why the second generation of works on the marital wage premium has focused on fixed effects models accounting specifically for individual characteristics that are constant throughout time. Moreover, these recent models include extensive demographic, job, company, region and year variables. Using this methodology, most scholars report a largely reduced marital wage premium (Ludwig & Bruderl, 2018; Killewald & Lundberg, 2017; Stratton, 2002; Cornwell & Rupert, 1997; Jacobsen & Rayack, 1996; Korenman & Neumark, 1991; Khazanov, 2020). This indicates that the selection of men into marriage on the basis of unobserved traits positively associated with wages partly explains the fact that married men earn more than single men (Dougherty, 2006), though it does not account for it fully (McDonald, 2020).

In fact there are a number of explanations why marriage might truly cause men to earn more. Firstly, marriage could facilitate the specialization of labor within the household, resulting in men focusing more heavily on market production (Becker, 1991). According to this theory, men and women focus on market or household production, depending on where they have a comparative advantage, defined as a lower opportunity cost of performing the activity. In practice, men specialize in market production as they have a comparative advantage in performing this activity, while women specialize in household production, as this is where they have their own comparative advantage. In turn, married men's increased investment in human capital and job-related skills could make them more productive (Stratton, 2002). To test this hypothesis empirically, many studies have assessed the impact of wives' working hours, showing that the marital wage premium decreases as wives' working hours increase (Chun & Lee, 2001; Gray, 1997; Bellas, 1992). However, it is unclear whether having a working wife necessarily causes a man to specialize less in market production, as richer households could simply decide to outsource home production. To investigate more precisely the impact of specialization on the marital wage premium, some recent studies have added a variable measuring the number of household chores performed by men and by women, in the United Kingdom. These studies have provided some support for the specialization hypothesis as married men seem

to perform, all else constant, fewer household chores than single men (Bardasi & Taylor, 2008). However, other studies using panel data have cast doubt on the specialization hypothesis, as the marital wage premium does not seem to decrease significantly in recent years, although women are increasingly joining the workforce globally (de Linde Leonard & Stanley, 2015). Moreover, the increase in wages associated with marriage sometimes occurs five or more years before the event date, which is therefore unlikely to be the result of specialization, unless the relationship is already stable years before the marriage date (Dougherty, 2006). In addition, some studies find that women also experience a marital wage premium, therefore incompatible with the specialization hypothesis (McConnell & Valladares-Esteban, 2020).

Alternatively, marriage could cause men to experience greater labor force attachment because of the added responsibilities they have to bear (de Linde Leonard & Stanley, 2015). If this is the case, men would be motivated to work harder once they are married, accumulating more human capital and becoming more productive. Consistent with this hypothesis, married men have an increased propensity to work longer hours (Knowles, 2013). Moreover, married men are more likely to receive work-related training and to experience lower turnover as well as shorter periods of unemployment (Ahituv & Lerman, 2011; Loh, 1996). Yet it is not entirely clear whether these results are indeed driven by the greater labor force attachment of married men, or by some other mechanism such as employer discrimination. In fact, employers might favor married men over unmarried men, for instance because they believe that they have greater needs (Killewald & Lundberg, 2017). In addition, employers can also trust that married men are more stable, hence more likely to stay with the firm in the long term. If this is the case, favoring married men by giving them higher salaries saves the firm future hiring and training costs. Hence employer discrimination could also explain the marital wage premium. However, this would not be consistent with the fact that the increase in wages associated with marriage sometimes happens years before the marriage date, unless of course the employee is already in a stable relationship likely to lead to a marriage then. All in all, prior work on the causes of the marital wage premium to date thus reaches contrasting conclusions,

highlighting the need for further research.

The Motherhood Wage Penalty

Another widely researched phenomenon in the literature on the impact of family structure on wages is the motherhood wage penalty. The motherhood wage penalty refers to the fact that mothers earn less than childless women with the same characteristics. Thus, for a motherhood wage penalty to exist, the event of parenthood should have a causal effect on the wages of women. There would be no motherhood wage penalty in the strict sense if mothers differ significantly from other women before they choose to have children. For instance, women who plan to raise a family could have a lower earning potential than other women, which might motivate their decision to become mothers. To control for this possibility, recent studies use individual fixed effects models, accounting for personal characteristics constant throughout time which could influence both wage earning probability and the decision to have children (Cukrowska-Torzewska, 2017; Weeden et al., 2016; Budig et al., 2016).

In fact, there are a number of explanations why motherhood might truly cause women to earn less. According to the productivity hypothesis, women become less attached to the labor market once they become mothers, which lowers their productivity. For instance, there is large empirical support for the fact that mothers take extended time out of the labor force and shift to family-friendly jobs with a part-time schedule when they come back (Weeden et al., 2016; Glass & Camarigg, 1992; Olivetti & Petrongolo, 2016). Mothers end up weighing the financial and non-financial benefits of the job they pick, often preferring a job with a lower pay if it has lower demands (Kahn et al., 2014; Blau & Kahn, 2017). More precisely, studies find that the motherhood wage penalty ranges from 5% to 10% per child in the United States (Killewald & Lundberg, 2017; Budig & England, 2001). This figure decreases to about 3% per child once work experience and shifts to family-friendly jobs are taken into account (Kahn et al., 2014; Glauber, 2007; Budig & England, 2001). However, there seems to be no significant motherhood wage penalty in

Northern countries such as Denmark, Norway and Finland (Petersen et al., 2014; Gash, 2009). Yet, in countries such as Germany and Switzerland, the motherhood wage penalty is significant with estimates ranging from 4% to 10% per child (Oesch et al., 2017). Variations in cultures and institutions certainly explain these cross-country differences (Cukrowska-Torzewska, 2017; Budig et al., 2016; Landais et al., 2019). For instance, Northern countries offer a long parental leave to both parents, promoting gender equality through shared childcare duties (Matysiak & Węziak-Białowolska, 2016). In contrast, countries such as Germany and Switzerland have put in place a shorter maternity leave, consistent with a more traditional view of gender roles.

Overall, taking into account extensive job characteristics and work experience does not always make the motherhood wage penalty disappear entirely (Cukrowska-Torzewska, 2017; Budig & England, 2001). To explain this remaining penalty, researchers have sought to compare the work effort of mothers with that of other women. Kmec (2011) finds no evidence that mothers provide less effort on the job than other women, using self-reported data. However, the study uses self reports, which could be problematic as it is likely that employees all indicate a high level of effort. In fact, other studies find contrasting results. Using a sample of women golfers, Kalist (2008) shows that the average ranking of women golfers decreases after childbirth. However, one might interpret these results cautiously as being a professional golfer implies very frequent travels, which mothers might be less willing to do. It is thus doubtful whether women golfers are representative of the whole population. Using a different approach, another study finds that the motherhood wage penalty decreases with children's age. The authors attribute this decline to the work effort hypothesis, implicitly assuming that older children require less care than younger children (Anderson et al., 2003).

Alternatively, employer discrimination could also explain the motherhood wage penalty. More precisely, two different types of discrimination could occur: statistical (Phelps, 1972) and taste-based discrimination (Becker, 1957). The first type of discrimination could happen if employers expect mothers to deliver a lower performance than other women as a group, although this might not be the case at the

individual level. Since monitoring is costly, mothers might systematically be paid less than other women (Correll et al., 2007). Alternatively, taste-based discrimination could occur if employers dislike working with mothers as opposed to other childless individuals, which seems less likely. Yet this taste-based discrimination could be driven by social norms. Indeed employers might believe that women should not be the primary source of income of their household. Instead they might think they should rather be constantly available for their children (Bornatici et al., 2021). This belief would then drive them to assign mothers lower salaries, as they are meant to represent only a supplementary income (Auspurg et al., 2017). Previous studies have assessed the importance of employer discrimination in hiring decisions (Cuddy et al., 2004; Heilman & Okimoto, 2008). Doing so, an American study finds that prospective employers rate mothers lower than comparable childless women on a variety of performance measures, even though they have similar resumes. These performance measures include competence, promotion prospects, commitment and recommendations for hire. As a result, mothers were called back for an interview half the time compared to other childless women (Correll et al., 2007). Participants also recommended that mothers receive a starting salary on average 7% lower than comparable childless women (Correll et al., 2007). Similarly, a Swiss vignette study shows that recruiters assign mothers wages that are 2 to 3% lower than those they give to comparable women who do not have children (Oesch et al., 2017). Hence mothers seem to be perceived as less competent, which explains their lower likelihood of getting an interview and their lower starting salary.

Data

To uncover the impact of family structure on wages, we take data from years 2009 to 2016 of the Swiss Household Panel. We focus our analysis on these years as there was a major change in family taxation which occurred on January 1st 2008, that made it more financially attractive for couples to marry. One could argue that it would make sense to include the year 2008 in our dataset as the change occurred on January 1st, yet we did not do so for two reasons. Firstly, there are high chances that the whole population

was not aware that the change happened on January 1st. Instead it is more likely that everyone learned about this reform gradually. Secondly, as we use data from all the respondents in every year in which the panel took place, and not only data from the respondents who answered the panel every year from the beginning until the end, there is a significant change in the data in 2008 with regards to participants' work experience. Hence including the year 2008 in our main sample would bias our results. To avoid chances that this exogenous change might bias the data, the year 2008 is excluded from the main analysis. Subsequent robustness checks focus on years 2004 to 2008, as data from 1999 to 2004 did not contain any information on respondents' work experience which is crucial to estimate wage models.

The Swiss Household Panel first started collecting data in 1999 with an initial sample containing 5,074 representative households made up of 12,931 individuals. In 2004, a second sample of 2,358 households and 6,569 individuals was added and finally since 2013 a third sample including 4,093 households and 9,945 individuals is interrogated. Eligible individuals are questioned yearly on a wide range of personal circumstances, including basic demographics, employment history, relationship status, income from all sources and beliefs on a variety of political and social issues. Following existing literature, this paper's analysis is restricted to people aged between 18 and 62 years old, to account for the various retirement ages. We also conducted all our analyses using the official Swiss retirement ages of 64 years old for women and 65 years old for men, yielding no significant changes in the main findings of the paper. As a result, these analyses are not reported here but are available from the author on request. Initial analysis is also restricted to people who are not self-employed. The dependent variable, which is the logarithm of gross hourly wage, is obtained by dividing gross yearly work income by annualized weekly hours of work. All prices are deflated to December 2015 level using the "Indice des prix a la consommation" from the Swiss National Statistics Office (Office Federal de la Statistique). In the Swiss Household Panel, all individuals aged 14 and above are asked about their income from various sources at the moment of the interview. The question regarding income from employment, both for employees and for managers, is

worded in the following way: “Since (month-year of the last interview) have you received a professional income as an employee?”. The question regarding income from self-employment is phrased in a similar fashion: “Since (month-year of the last interview) have you received a professional income from being self-employed?”. If respondents reply positively to these questions, they are then asked to report how much they made per month, or per year if more convenient.[1] In addition, respondents have to indicate how many hours they usually work each week in the Labor Force Survey, including usual paid and unpaid overtime.[2]

Focusing on years 2009 to 2016, Table 1 highlights that married men enjoy the highest gross wage, at 63.77 CHF per hour on average, followed by single men who earn 39.52 CHF per hour. Similarly, married women earn 42.03 CHF per hour while single women earn only 35.73 CHF per hour. Thus married men enjoy a premium of 61% compared to single never married men and married women of 18% compared to single never married women. This confirms that married people earn more than single people in Swiss data, although in isolation this raw data does not allow us to draw interesting causal conclusions. In fact, descriptive statistics cannot help us determine whether there is a causal marital wage premium, or whether married people earn more than single people for instance because they tend to be older on average. Moreover, Table 1 shows the impact of family size, given by number of children, on the wages of men and women. Men’s gross hourly wages increase with number of children. In fact, men who do not have children earn 50.95 CHF on average per hour, while men with one child earn 56.99 CHF and those with two or more earn 64.74 CHF. On the opposite, the wages of women do not seem to increase with number of children. Women who do not have children earn on average 40.84 CHF per hour, while those with one child earn 39.81 CHF per hour and those with two children or more 40.62 CHF per hour. Hence men with two children or more earn a premium of 27% compared to childless men. On the opposite, women with two children or more experience a raw wage penalty of 1% compared to childless women. These summary statistics give us a first indication that women who have children might earn less than women who do not have children. However, once again, these raw data do not give us interesting insights on whether there is a

causal motherhood wage penalty, which is investigated in more detail below.

Econometric And Empirical Specification

Pooled Ordinary Least Squares

We assume throughout that wages are set by the following equation, where *i* represents the individual and *t* represents the year from which observations are taken:

$$\ln w_{it} = x_{it}\beta + m_{it}\rho + c_{it}\gamma + u_{it} \text{ where } u_{it} = a_i + v_{it}$$

In this equation, *w* is the gross hourly wage deflated to December 2015 price level, *x* represents a range of demographic, job, region and year controls, *m* captures the relationship status of the individual (married or divorced-separated-widowed), *c* captures the number of children the respondent has and *u* is unobservable individual error. The vector *x* is composed of extensive controls, that are categorized into demographic, job, region and year controls. Demographic controls include a person's age, qualification level, experience, experience squared (to account for non-linearity in experience), whether the individual is an immigrant and whether he is registered disabled. Job controls include company size, company sector, industry and occupation. We use panel-corrected standard errors, clustered at household level, to correct for the clustering of observations across years. As our sample includes individuals who switched households during the time period of the survey, we created new household clusters encompassing all the households in which the switchers have ever lived, as well as all the individuals in those households. For instance, say Individual A goes from Household 1, where he lived with his parents Individuals B and C to Household 2, where he lives with his spouse Individual D. To account for this switch, we created a new household cluster variable, including Households 1 and 2 as well as Individuals A, B, C and D. This procedure also implies that all people with whom D ever lived are in this same cluster, as the procedure used for A is applied to D in the same way. Performing this on the entire sample resulted in 9745 newly created household clusters. Moreover, we added a variable *years married* to the model in case the marital wage premium became

apparent only after several years. However, this yielded no significant change and we dropped this variable from reported specifications. Similarly, a sensitivity analysis dropping the 5th and 95th percentile of the income distribution yielded no significant change. Hence we decided to report models of the full income distribution.

By Gauss-Markov, for OLS to be valid, regressors must be exogenous and hence satisfy $Cov(\mathbf{u,x}) = 0$, $Cov(\mathbf{u,m}) = 0$ and $Cov(\mathbf{u,c}) = 0$. Here, these equations are unlikely to hold as the unobservable error is composed of a random component as well as a time invariant individual effect, which affects the probability of being married, having children, wages and other controls. In fact, there is $Cov(\mathbf{u,x}) \neq 0$, $Cov(\mathbf{u,m}) \neq 0$, and $Cov(\mathbf{u,c}) \neq 0$. For instance, one would expect a gifted and hardworking individual to have higher educational qualifications, earn a higher wage, work in a more competitive industry and have more chances of finding a partner with reproductive success. To address this issue and assuming these characteristics are constant throughout a short period of time at the level of the individual, fixed effects regressions are used.

Fixed Effects

Fixed effects regressions take into account deviations of variables from their means. Applied to the analysis of the marital wage premium and motherhood wage penalty, these regressions take the form:

$$\ln w_{it} - \ln \bar{w_i} = (x_{it} - \bar{x_i})\zeta + (m_{it} - \bar{m_i})\eta + (c_{it} - \bar{c_i})\mu + (u_{it} - \bar{u_i}) =$$

$$(x_{it} - \bar{x_i})\zeta + (m_{it} - \bar{m_i})\eta + (c_{it} - \bar{c_i})\mu + (a_i - \bar{a_i}) + (v_{it} - \bar{v_i}) =$$

$$(x_{it} - \bar{x_i})\zeta + (m_{it} - \bar{m_i})\eta + (c_{it} - \bar{c_i})\mu + (v_{it} - \bar{v_i})$$

In this specification, time invariant unobservable variables are accounted for, since their mean is constant over time. Thus, this model corrects for the bias introduced by individual characteristics that are constant throughout time and impossible to measure in surveys, such as attractiveness or desire to conform to social expectations. As for OLS regressions, we use panel-corrected standard errors clustered at

household level, using the household clusters which group together all the individuals from households in which the respondents have ever lived.

Results

Main Sample: Years 2009-2016

The male marital wage premium. Ignoring issues of unobserved heterogeneity, selection and endogeneity, the first three models we report use pooled ordinary least squares (Table 2). Model (1) includes no controls, Model (2) includes demographic controls and Model (3) includes demographic, job, region and year controls. Model (3), which has the most complete specification, shows a marital wage premium for men of 5%, much smaller than the one observed in raw data which was as high as 61% (Table 1). The fact that the premium decreases in OLS specifications controlling for observable variables compared to raw data indicates that men who choose to marry have certain observable characteristics associated with a higher earning probability. In other words, men are selected into marriage based on observable characteristics, which also affect their wage earning probability. One obvious example of such an observable characteristic is age, as married men tend to be older and older men tend to earn more than younger men. In addition, the OLS models highlight that there is a premium for men whose marriages have dissolved. If this result is robust in fixed effects specifications, it could indicate that men accumulate human capital when they are married, hence increasing their lifetime earnings. Lastly, OLS specifications reveal that a man earns 4% more for every child he has. While we should not take this result for granted unless it holds in fixed effects specifications, we speculate that children could have a beneficial effect on wages as they could indicate higher financial need. In turn, this might convince employers to assign higher wages to fathers as opposed to other childless individuals. All in all, OLS models reveal that a large part of the marital wage premium is due to selection on observables. Yet we reiterate that OLS results should be taken with caution as it is likely that unobservable variables bias coefficients of interest. To address this

issue, fixed effects models will be analyzed next.

Fixed effects models report an insignificant marital wage premium in the specifications including demographic, job, region and year controls (models (5) and (6)). As the significant effect of marital status on wages present in OLS models disappears, these results highlight that married men are selected into marriage based on unobservable characteristics, as well as on observable ones. In fact, fixed effects models take into account time constant unobservable variables, thus isolating the effect of marital status on wages in a cleaner way. On a different note, fixed effects models highlight that there is no premium for men whose relationships have dissolved, in contrast with OLS results. This difference certainly comes from the fact that there was an omitted variable in the OLS regressions affecting both divorce likelihood and productivity. Hence it is after all unlikely that significant human capital is accumulated during marriage as opposed to other marital statuses. Also in contrast with OLS models, children no longer have a significant effect on the wages of men. This highlights again that unobservable time constant characteristics were affecting both men's probability of having children and of having a higher wage. For instance, it is likely that good looking and healthy men have a higher probability of being fathers and of being selected into important roles. In conclusion, OLS and fixed effects models report that men appear to be selected both on observable and unobservable characteristics into marriage and parenthood. There is no causal effect of marital status and parenthood on the wages of men.

The motherhood wage penalty. Ignoring issues of heterogeneity, selection and endogeneity, the first three models we report use OLS. These models highlight a significant marital wage penalty for women ranging between 4 and 11% when demographic controls are included. While one should be cautious about overly interpreting OLS results, a variety of factors could explain this marital wage penalty for women. For a start, women who choose to marry could have a lower earnings potential than other women or they could be less motivated to work hard than other women. Alternatively, they could be seen as more likely to quit the labor force, since they are more likely to have children. This could encourage employers to avoid

promoting married women as opposed to other individuals. In addition, OLS specifications highlight that there is no effect of marriage dissolution on wages, once demographic controls are taken into account. More importantly, the motherhood wage penalty of between 4% and 7% disappears when job, region and year controls are taken into account. This is a first indication that mothers compensate differentials, choosing different occupations and companies when they become mothers. As OLS models surely suffer from endogeneity though, further analysis will focus on fixed effects models.

Fixed effects models report that marital status does not impact significantly women's wages, once demographic controls are taken into account. This indicates that women are selected into marriage based on unobservable characteristics, as OLS models highlighted a significant marital wage penalty. Hence women who choose to marry have unobservable characteristics correlated with a lower earning potential. In addition, we find additional support for the fact that marriage dissolution does not impact significantly the wages of women once demographic controls are added. More importantly, fixed effects regressions show that women suffer from a gross motherhood wage penalty of around 12%, as seen in Model (4). Models (5) and (6) give us a first indication of what might explain this motherhood wage penalty. Firstly, Model (5) reveals that women who choose to have children have observable characteristics that are associated with a lower earnings probability. In fact, when demographic controls are added to the model, the motherhood wage penalty decreases from 12% to 5%. Hence selection into motherhood happens to an extent. Women who choose to have children have a lower earnings potential than other women. Secondly, Model (6) gives us an indication that mothers compensate differentials, preferring to work in occupations and companies associated with a lower pay. This is clear as the motherhood wage penalty drops to 3% when job, region and year controls are taken into account in the model. Overall, these results indicate that selection and compensating differentials play a role in explaining the motherhood wage penalty. Yet we are left with an unexplained difference between the wages of mothers and those of other women. We will now look further into the causes of this gap.

Firstly, we will check whether it is motherhood *per se* that penalizes women, or what motherhood entails for women's work productivity. To estimate this in our models, we will use a variable indicating whether the respondent takes care of her children when they are ill.[3] As illnesses are very unpredictable, it is likely that these specific childcare duties will entail a temporary decrease in productivity since child carers would most likely need to leave their workplace. Moreover, since children require full attention, it is not probable that a person taking care of her ill children would be able to work productively at the same time. This is true especially if the children need to go to the doctor or require specific care. Hence emergency childcare duties are a good way to approximate the effect of motherhood on work productivity.

Including a dummy for emergency childcare duties in Model (1) of Table 4 shows clearly that the motherhood wage penalty disappears upon the inclusion of this variable, as the coefficient on number of children becomes insignificant. Overall, becoming a mother *per se* does not harm the wages of women, but being in charge of emergency childcare duties and hence being liable to unpredictable absences from work does. Yet this result might be biased if women in charge of emergency childcare duties have a significantly different earning potential than other women. In fact, women with a lower earning capability might deliberately choose to take care of their children when they are ill, as they know they should not focus too much on their work. If this is the case, emergency childcare duties would just be a proxy for lower earning capability, and hence would not really explain the motherhood wage penalty. To test whether this is the case or whether childcare duties are really key in explaining the motherhood wage penalty, we will consider an alternative specification. In this alternative specification, we will include the second lead of the variable indicating emergency childcare duties. This variable represents whether a woman will take care of her ill children in two years from now, regardless of whether she has childcare duties now. If women who will have childcare duties in two years receive a wage penalty compared to other women now, this means that they have a lower earning potential compared to other women, indicating selection into childcare duties. On the opposite, if women who will have childcare duties in two years do not receive a wage

penalty compared to other women now, this is an indication that there is no selection into childcare duties. Model (2) in Table 4 shows that this is the case, as the coefficient on the emergency childcare variable's second lead is not significant. Women who do not have childcare duties now but will do in two years are not penalized now compared to other women. Hence these women do not necessarily have a lower earning potential, meaning women do not self-select into emergency childcare duties. In conclusion, the negative effect of emergency childcare duties on women's wages does not seem to be driven by selection. Our results indicate that emergency childcare duties really affect the wages of women.

To investigate further why childcare duties impact negatively the wages of women, the previous fixed effects models will first be applied to the sub-sample of self-employed women. As these women are less likely to suffer from employer discrimination and as they benefit from greater work flexibility, it is interesting to find out whether they suffer from a similar motherhood wage penalty to that of other employed women. These fixed effects models reveal that self-employed women do not experience a motherhood wage penalty at all, as opposed to employed women (Table 5). Now, is work flexibility solely responsible for the motherhood wage penalty or does employer discrimination play a role too? To answer this question, subsequent analyses will focus on comparing the effect of parenthood for men and women at similar hierarchical levels. More precisely, the models will be applied separately to managers and to other professionals, of both genders. Assuming that all professionals benefit from the same average level of work flexibility, female professionals, but not male professionals, should experience a parenthood wage penalty if they are the victims of employer discrimination. In addition, assuming managers are more flexible than other professionals, managers of both genders should not suffer from a large parenthood wage penalty if work flexibility truly improves the parenthood wage penalty.

Models (1) and (3) in Table 6 reveal that managers of both genders do not experience a parenthood wage penalty, as the coefficient on the number of children is insignificant. Moreover, models (2) and (4) report that female professionals experience a wage penalty of 5% on average, while male professionals are

not affected. This provides evidence that employer discrimination and lack of workplace flexibility both likely play a role in explaining the motherhood wage penalty. As a robustness check for our results, we will now consider specifications including both genders. This will allow us to compare women with men, instead of comparing men with men and women with women. This is important as the difference between a statistically significant coefficient and a coefficient that is not statistically significant is not necessarily statistically significant (Gelman & Stern, 2006).

Model (1) in Table 7 reports that women are indeed penalized by parenthood and more specifically by emergency childcare duties more than men are, as indicated by the significant negative interaction effect of emergency childcare duties with the dummy indicating female gender. Interestingly, Model (2) is in line with previous findings and shows that self-employed women are not punished at all compared to men, whether for motherhood or for childcare duties. Also in line with previous results, Model (3) highlights that managers of both genders do not suffer from parenthood or childcare duties, certainly due to their enhanced flexibility. On the opposite, Model (4) reports that female professionals who are not managers experience a significant penalty due to childcare duties. This provides evidence that employer discrimination is likely explaining part of the motherhood wage penalty.

Overall, our results emphasize that there are a few factors explaining the motherhood wage penalty. While selection and compensating differentials play an important role, they are by no means the full explanation. In fact, a motherhood wage penalty of about 6% remains in the most complete fixed effects model. This difference is robust in specifications comparing women with men, providing evidence for the gender wage gap. Further analyses reveal that employer discrimination might play a role in explaining the motherhood wage penalty. Moreover, we find evidence that women in more flexible positions, such as self-employed women and managers, do not suffer from a motherhood wage penalty.

Robustness Check: Years 2004-2008

The male marital wage premium. Ignoring issues of unobserved heterogeneity, selection and endogeneity, the first results reported in models (1), (2) and (3) use pooled ordinary least squares. The most complete specification reports a wage premium of 5% for married men (Table 8), in line with the one observed in later years in similar models, which was of 5% too. Including demographic, job, region and year controls decreases the marital wage premium significantly. Hence this result confirms that men are likely selected into marriage based on observable characteristics which also affect their wage earning probability. Overall, OLS models on the earlier sample report similar results than in the later sample, indicating that a large part of the marital wage premium is due to selection on observables. Fixed effects models report an insignificant marital wage premium in the specifications including demographic controls. As the significant positive effect of marital status on wages present in OLS models disappears, these results highlight that married men are selected on unobservable characteristics, as well as observable ones. Moreover, there is no significant effect of relationship dissolution or parenthood on wages, in fully specified models. Overall, these results confirm that there is no causal effect of marital status and parenthood on the wages of men, in line with conclusions for the later sample.

The motherhood wage penalty. Leaving aside issues of heterogeneity, selection and endogeneity, the first results reported in models (1), (2) and (3) use OLS models to estimate the motherhood wage penalty. These models reveal the existence of a motherhood wage penalty ranging between 7% and 11% in models which do not control for job, region and year variables. Note that these effect sizes are slightly larger than those observed in the later sample, where the motherhood wage penalty ranged between 4% and 7% in similar models. Moreover, there is no robust effect of relationship dissolution or marital status on the wages of women, once demographic, job, region and year controls are taken into account. Yet these OLS regressions should be analyzed with caution, as they are likely to be biased by unobservable variables.

Fixed effects models reveal added evidence that marital status and relationship dissolution do not impact the wages of women, while there is a robust motherhood wage penalty. In fact, the motherhood wage penalty is of 9% on average in models accounting for demographic, job, region and year controls. As this penalty was smaller in the later sample, we next wished to see whether this difference between the earlier and the later sample was significant. To test whether there is a significant decrease in the motherhood wage penalty in Switzerland over time, we conducted two fixed effects regressions. The first one included an interaction term between the number of children and a continuous time variable. The second one included an interaction term between the number of children and a dummy variable indicating the years from 2009 onwards. These models indicate that there is no significant improvement in the motherhood wage penalty in Switzerland over time (Table 10). In fact, the coefficient on the interaction term between the continuous time variable and the number of children is not significant. In the same way, the coefficient on the interaction term between the dummy variable indicating the years after 2008 and the number of children is not significant.

Discussion & Conclusion

This paper provides novel and unique evidence on the impact of family structure on wages in Switzerland. It reveals that marital status and parenthood do not causally affect the wages of men while women suffer from a robust motherhood wage penalty of 6% on average, in line with previous estimates (Oesch et al., 2017). Going further than previous studies, this paper provides unique evidence that childcare responsibilities causing unpredictable absences from work are key in explaining the motherhood wage penalty. Furthermore, we report that female managers and self-employed women do not suffer from a motherhood wage penalty. Assuming childcare duties impact the productivity of all working women in a similar way, these results show that female managers and self-employed women are able to mitigate the negative impact of childcare duties on their work thanks to their more supportive work environment. Hence

work flexibility seems key to improve the negative impact of motherhood on wages. Moreover, by comparing men and women in similar positions, we find that men do not suffer from a wage penalty, even when they are in charge of the same childcare duties as women. Assuming fathers and mothers are as productive, this constitutes evidence that employer discrimination also plays a role in explaining the motherhood wage penalty. In particular, our data indicates that there might be taste-based discrimination happening against women. In fact, it is only logical to assume that emergency childcare duties should impact the productivity of men and women in the same way, as they imply stopping to work productively for a few hours to focus on child care. As a result, emergency childcare duties should have the same effect on the work productivity of men and women, as well as the same impact on their wages in the absence of taste-based discrimination. Since this is not the case, our data seems to indicate that women's productivity, which could result in further statistical discrimination, is not the key factor explaining the gender wage gap in Switzerland. It rather seems as if there is evidence for taste-based discrimination, which resonates with findings by Combet & Oesch (2019) who compared the evolution of wages between men and women matched on educational attainment and intellectual ability, before the onset of family formation and gendered household specialization. Doing so, they find that young women earn lower wages than young men with the same productive characteristics long before they have children, providing further evidence for taste-based discrimination in Switzerland. More generally, a recent meta study has also highlighted the widespread presence of taste-based discrimination against ethnic minorities in the labor market (Lippens et al., 2022), further corroborating our hypothesis.

We acknowledge that our analysis is not without limitations. The varying coefficients in reported models suggest that there is considerable uncertainty as to the precise extent of the motherhood wage penalty. Displaying results with and without controls aims to make this uncertainty transparent. Yet, it is likely that the estimates still suffer from unobserved determinants of work productivity, which are impossible to measure for. For instance, a recent study argues that current panel datasets do not include

information on relevant confound variables which change over time, such as mental health and substance abuse (Sobel, 2012). Moreover, there is recent evidence showing that women might misreport their income in Swiss survey data to comply with the male breadwinner norm (Roth & Slotwinski, 2019). Lastly, it is impossible to establish causal effects with complete certainty in the absence of random assignment.

While the Swiss Household Panel contains extensive data, it would indeed be important to have access to other variables that would allow us to identify more precisely the causes of the gender wage gap. First and foremost, it would be ideal to obtain an objective performance measure, which would enable us to pinpoint whether there are any important productivity differences between men and women that could explain wage differentials. In the same way, having access to more extensive data on the noncognitive or soft skills of respondents, including psychological attributes, preferences and personality, would be extremely interesting for our analysis. For instance, it would be an important addition to obtain data on the propensity to negotiate of our respondents, as well as on their competitiveness and risk aversion. The Swiss Household Panel already collects some data about the Big Five personality traits, but only in a partial way and solely in 2015. To the extent that labor market outcomes may affect personality, there would be a reverse causality issue in using this data in our regression analysis based mostly on earlier data. In addition, it would be very interesting to explore further the norms to which our respondents adhere with respect to housework and marriage. While some of these norms (such as whether a child suffers if his-her mother works) are already elicited in the panel, collecting more detailed information about these would be a great addition to the dataset. In fact, some existing studies on the impact of psychological factors on the gender pay gap use information on respondents' answers to attitudinal questions to construct indexes of psychological traits, which then become explanatory variables in wage regressions (Blau & Kahn, 2017). This way, one can then assess the quantitative importance of such elements in explaining the level or change in the gender pay gap.

Although having access to more extensive data would be important, this should not distract from the

main finding that motherhood in our data is associated with a decrease in wages. These results have interesting implications for research on the gender wage gap. Firstly, this study highlights that workplace flexibility seems important to best combine the demands of childcare and work. This is in line with previous work showing that the policies with the strongest impact on reducing gender disparities are those making it easier to be a working mother (Olivetti & Petrongolo, 2017), enhancing temporal flexibility (Goldin, 2014). Yet these policies could also be problematic as they could contribute to a slower career progression of women, since it has been shown that people on telework are less likely to be promoted than their peers (Bloom et al., 2015). Moreover, some research suggests that flexible work increases the likelihood of work-family conflicts (Allen et al., 2013), hence further research could seek to understand which aspects of work flexibility are key in easing the burden of childcare duties while still allowing a reasonable career progression. One could then explore which type of flexible work scheme would be best to decrease work-family conflict. Secondly, our study reveals the likely existence of employer discrimination against mothers. To help raise awareness about this issue, future research could examine the effectiveness of business led diversity initiatives. In fact, businesses commonly implement policies, guidelines and trainings to promote diversity. While previous research has found that trainings have little impact on the share of women in the workplace (Kalev et al., 2006), it would be interesting to investigate whether the same is true for diversity policies. In the same way, future research could analyze whether common diversity initiatives, such as policies and trainings, help reduce the gender pay gap. Moreover, policies encouraging universal child care and establishing longer paternity leaves have been shown to be particularly positive for the wages of women, across 38 countries (Ciminelli et al., 2021). In fact, early childhood spending raises both female labour supply and their wages relative to men (Olivetti & Petrongolo, 2017) and maternity leaves are positive for the wages of women as long as they are not longer than a few months, since if they are too long they tend to decrease the wages of women once they come back to the labor market (Akgunduz & Plantenga, 2013). Future research could simulate the gender wage

gap in Switzerland should these policies be put in place, with a view to initiating positive change. Taken together, the present findings highlight initial actionable recommendations to organizations and policy-makers interested in reducing the gender pay gap, while uncovering interesting new alleys for future research.

References

Ahituv, A., & Lerman, R. I. (2011). Job turnover, wage rates, and marital stability: How are they related? *Review of Economics of the Household*, *9*(2), 221–249.

Akgunduz, Y. E., & Plantenga, J. (2013). Labour market effects of parental leave in europe. *Cambridge Journal of Economics*, *37*(4), 845–862.

Allen, T. D., Johnson, R. C., Kiburz, K. M., & Shockley, K. M. (2013). Work–family conflict and flexible work arrangements: Deconstructing flexibility. *Personnel Psychology*, *66*(2), 345–376.

Anderson, D. J., Binder, M., & Krause, K. (2003). The motherhood wage penalty revisited: Experience, heterogeneity, work effort, and work-schedule flexibility. *ILR Review*, *56*(2), 273–294.

Auspurg, K., Hinz, T., & Sauer, C. (2017). Why should women get less? evidence on the gender pay gap from multifactorial survey experiments. *American Sociological Review*, *82*(1), 179–210.

Bardasi, E., & Taylor, M. (2008). Marriage and wages: A test of the specialization hypothesis. *Economica*, *75*(299), 569–591.

Becker, G. S. (1957). *The economics of discrimination: an economic view of racial discrimination*. University of Chicago.

Becker, G. S. (1991). *A treatise on the family: Enlarged edition*. Harvard university press.

Bellas, M. L. (1992). The effects of marital status and wives'employment on the salaries of faculty men: The (house) wife bonus. *Gender & Society*, *6*(4), 609–622.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, *55*(3), 789–865.

Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics*, *130*(1), 165–218.

Bornatici, C., Gauthier, J. A., & Le Goff, J. M. (2021). *Attitudes towards gender equality in switzerland.* (https://www.socialchangeswitzerland.ch/?p=2227)

Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American Sociological Review*, 204–225.

Budig, M. J., Misra, J., & Boeckmann, I. (2016). Work–family policy trade-offs for mothers? unpacking the cross-national variation in motherhood earnings penalties. *Work and Occupations*, *43*(2), 119–177.

Chun, H., & Lee, I. (2001). Why do married men earn more: Productivity or marriage selection? *Economic Inquiry*, *39*(2), 307–319.

Ciminelli, G., Schwellnus, C., & Stadler, B. (2021). Sticky floors or glass ceilings? the role of human capital, working time flexibility and discrimination in the gender wage gap. *OECD Library*.

Combet, B., & Oesch, D. (2019). The gender wage gap opens long before motherhood. panel evidence on early careers in switzerland. *European Sociological Review*, *35*(3), 332–345.

Cornwell, C., & Rupert, P. (1997). Unobservable individual effects, marriage and the earnings of young men. *Economic Inquiry*, *35*(2), 285–294.

Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology*, *112*(5), 1297–1338.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *Journal of Social Issues*, *60*(4), 701–718.

Cukrowska-Torzewska, E. (2017). Cross-country evidence on motherhood employment and wage gaps: The role of work–family policies and their interaction. *Social Politics: International Studies in Gender, State & Society*, *24*(2), 178–220.

de Linde Leonard, M., & Stanley, T. D. (2015). Married with children: What remains when observable biases are removed from the reported male marriage wage premium. *Labour Economics*, *33*, 72–80.

Dougherty, C. (2006). The marriage earnings premium as a distributed fixed effect. *Journal of Human Resources*, *41*(2), 433–443.

Eurostat. (2018). *Gender pay gap statistics.* (https://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics)

Gash, V. (2009). Sacrificing their careers for their families? an analysis of the penalty to motherhood in europe. *Social Indicators Research*, *93*(3), 569–586.

Gelman, A., & Stern, H. (2006). The difference between significant and not significant is not itself statistically significant. *The American Statistician*, *60*(4), 328–331.

Glass, J., & Camarigg, V. (1992). Gender, parenthood, and job-family compatibility. *American Journal of Sociology*, *98*(1), 131–151.

Glauber, R. (2007). Marriage and the motherhood wage penalty among african americans, hispanics, and whites. *Journal of Marriage and Family*, *69*(4), 951–961.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091–1119.

Gray, J. S. (1997). The fall in men's return to marriage: Declining productivity effects or changing selection? *Journal of Human Resources*, 481–504.

Greenhalgh, C. (1980). Male-female wage differentials in great britain: Is marriage an equal opportunity? *The Economic Journal*, *90*(360), 751–775.

Heilman, M. E., & Okimoto, T. G. (2008). Motherhood: A potential source of bias in employment decisions. *Journal of Applied Psychology*, *93*(1), 189.

Info, S. (2019). *Swiss workplace inequality, by the numbers.* (https://www.swissinfo.ch/eng/mind-the-gap_inequality-in-the-swiss-workplace--by-the-numbers/44809274)

Jacobsen, J. P., & Rayack, W. L. (1996). Do men whose wives work really earn less? *The American Economic Review*, *86*(2), 268–273.

Kahn, J. R., García-Manglano, J., & Bianchi, S. M. (2014). The motherhood penalty at midlife: Long-term effects of children on women's careers. *Journal of Marriage and Family*, *76*(1), 56–72.

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, *71*(4), 589–617.

Kalist, D. E. (2008). Does motherhood affect productivity, relative performance, and earnings? *Journal of Labor Research*, *29*(3), 219–235.

Khazanov, M. (2020). *Love-love: Marital status and the effect on professional tennis player performance* (Unpublished doctoral dissertation). Haverford College.

Killewald, A., & Lundberg, I. (2017). New evidence against a causal marriage wage premium. *Demography*, *54*(3), 1007–1028.

Kmec, J. A. (2011). Are motherhood penalties and fatherhood bonuses warranted? comparing pro-work behaviors and conditions of mothers, fathers, and non-parents. *Social Science Research*, *40*(2), 444–459.

Knowles, J. A. (2013). Why are married men working so much? an aggregate analysis of intra-household bargaining and labour supply. *Review of Economic Studies*, *80*(3), 1055–1085.

Korenman, S., & Neumark, D. (1991). Does marriage really make men more productive? *Journal of Human Resources*, 282–307.

Landais, C., Kleven, H., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings*, *109*, 122–126.

Lippens, L., Baert, S., Ghekiere, A., Verhaeghe, P.-P., & Derous, E. (2022). Is labour market discrimination against ethnic minorities better explained by taste or statistics? a systematic review of the empirical evidence. *Journal of Ethnic and Migration Studies*, 1–34.

Loh, E. S. (1996). Productivity differences and the marriage wage premium for white males. *Journal of Human Resources*, 566–589.

Ludwig, V., & Bruderl, J. (2018). Is there a male marital wage premium? new evidence from the united states. *American Sociological Review*, *83*(4), 744–770.

Matysiak, A., & Węziak-Białowolska, D. (2016). Country-specific conditions for work and family reconciliation: An attempt at quantification. *European Journal of Population*, *32*(4), 475–510.

McConnell, B., & Valladares-Esteban, A. (2020). *On the marriage wage premium.* (unpublished)

McDonald, P. (2020). The male marriage premium: selection, productivity, or employer preferences? *Journal of Marriage and Family*, *82*(5), 1553–1570.

Oesch, D., Lipps, O., & McDonald, P. (2017). The wage penalty for motherhood: Evidence on discrimination from panel data and a survey experiment for switzerland. *Demographic Research*, *37*, 1793–1824.

Olivetti, C., & Petrongolo, B. (2016). The evolution of gender gaps in industrialized countries. *Annual Review of Economics*, *8*, 405–434.

Olivetti, C., & Petrongolo, B. (2017). The economic consequences of family policies: lessons from a century of legislation in high-income countries. *Journal of Economic Perspectives*, *31*(1), 205–30.

Petersen, T., Penner, A. M., & Høgsnes, G. (2014). From motherhood penalties to husband premia: The new challenge for gender equality and family policy, lessons from norway. *American Journal of Sociology*, *119*(5), 1434–1472.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, *62*(4), 659–661.

Roth, A., & Slotwinski, M. (2019). Gender norms and income misreporting within households. *CESifo Working Paper*.

Schoeni, R. F. (1995). Marital status and earnings in developed countries. *Journal of Population Economics*, *8*(4), 351–359.

Sobel, M. E. (2012). Does marriage boost mens wages: Identification of treatment effects in fixed effects regression models for panel data. *Journal of the American Statistical Association*, *107*(498), 521–529.

Stratton, L. S. (2002). Examining the wage differential for married and cohabiting men. *Economic Inquiry*, *40*(2), 199–212.

Weeden, K. A., Cha, Y., & Bucca, M. (2016). Long work hours, part-time work, and trends in the gender gap in pay, the motherhood wage penalty, and the fatherhood wage premium. *RSF*.

Footnotes

[1]If respondents report their monthly income, they are asked about the number of months in which they received their income. If income has been declared on a monthly basis, the monthly income is multiplied by 12. If income has only been received during parts of the year, yearly income is constructed based on the number of months in which income has been received and the activity calendar. The coefficient applied is the ratio of the number of months in which an income has been received to the interval between the two interviews. If respondents declare a change in their working status, they are asked to give their working status month by month since the last interview date. An hourly wage is then computed from actual occupation rate and work income.

[2]Teachers are advised to count both the weekly hours spent teaching and the hours spent preparing and correcting. Measuring work hours this way has several advantages. Firstly, it ensures that all hours effectively worked on a usual week are taken into account, as opposed to the number of contractual hours which could be greatly different to the number of hours usually worked. Secondly, this calculation method makes sure that participants do not report the number of hours worked on an unusually busy or calm week, which could happen if for instance participants had been asked to report how long they had worked in the week preceding the interview date.

[3]In the Swiss Household Panel questionnaire, respondents are asked *Regarding the children in your household, can you tell me who usually takes care of the following tasks? 1. Looking after them in case of illness*. To this question, respondents can answer 1 = Mainly myself, 2 = Mainly my partner, 3 = Both equally, 4 = Another child in our household, 5 = Mainly another person in our household, 6 = Mainly a domestic help, 7 = Mainly the children's father/mother (not living in the household), 8 = Myself and the children's father/mother not in the household, 9 = Mainly another person who is not part of our household, 10 = Other solution, 11 = Doesn't apply. We recoded this variable as 0 = the respondent does not take care of her children when they are ill, 1 = the respondent takes care of her children when they are ill. Notice we

considered an alternative recoding, with 0 = the respondent does not take care of her children when they are ill, 1 = the respondent takes care of her children when they are ill, including the responses originally coded as 1, and half of responses coded as 3 and 8 in the original questionnaire. As this alternative recoding yielded no significant change, we focused on the cleaner original coding.

TABLE 1: *Summary Statistics of Gross Hourly Wage, by Civil Status and Number of Children*

	Men		Women	
	Mean	S.D.	Mean	S.D.
Panel A: Civil Status				
Married	63.77	(58.55)	42.03	(33.96)
Single	39.52	(42.66)	35.73	(29.63)
Panel B: Number of Children				
None	50.95	(49.42)	40.84	(33.15)
One	56.99	(53.13)	39.81	(32.72)
Two or more	64.74	(74.77)	40.62	(30.15)

Note. SHP 2009-2016. Dependent variable is gross hourly wages (in swiss francs) deflated to December 2015 prices.

TABLE 2: *OLS and Fixed Effects Models for Men*

	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)
	(1)	(2)	(3)	(4)	(5)	(6)
Married	0.69**	0.09**	0.05**	0.25**	0.04	0.03
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Divorced/Separated/Widowed	0.65**	0.07*	0.05	0.21**	0.01	-0.01
	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Number of Children	0.01	0.03**	0.04**	-0.04**	-0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	3.33**	2.24**	2.84**	3.65**	2.17**	2.45**
	(0.02)	(0.05)	(0.06)	(0.02)	(0.52)	(0.63)
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Job Region and Year Controls	No	No	Yes	No	No	Yes
Fixed effects	No	No	No	Yes	Yes	Yes
R^2	0.19	0.32	0.42	0.01	0.05	0.07
Observations	16410	14502	13129	16410	14502	13129

Note. SHP 2009-2016. Dependent variable is log gross hourly wages (in swiss francs) deflated to December 2015 prices. Models (1), (2) and (3) use OLS, (4), (5) and (6) fixed effects. Married = 0 if the respondent is not married, 1 otherwise. Divorced/separated/widowed = 0 if the respondent is not divorced/separated/widowed, 1 otherwise. Demographic controls = age, qualification level, experience, experience squared, immigrant and disabled statuses. Job controls = company size, company sector, industry and occupation. Unstandardized regression coefficients and standard errors are shown. Errors are clustered at household level. * p < .05. ** p < .01.

TABLE 3: *OLS and Fixed Effects Models for Women*

	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)
	(1)	(2)	(3)	(4)	(5)	(6)
Married	0.31**	-0.11**	-0.04	0.19**	-0.06	-0.03
	(0.03)	(0.03)	(0.02)	(0.04)	(0.05)	(0.04)
Divorced/Separated/Widowed	0.45**	0.03	0.03	0.29**	0.03	0.03
	(0.03)	(0.03)	(0.02)	(0.06)	(0.06)	(0.06)
Number of Children	-0.07**	-0.04**	0.01	-0.12**	-0.05**	-0.03*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	3.26**	2.42**	2.96**	3.38**	3.40**	4.54**
	(0.02)	(0.05)	(0.06)	(0.03)	(0.56)	(0.78)
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Job Region and Year Controls	No	No	Yes	No	No	Yes
Fixed effects	No	No	No	Yes	Yes	Yes
R^2	0.04	0.19	0.33	0.01	0.03	0.07
Observations	18184	15914	14055	18184	15914	14055

Note. SHP 2009-2016. Dependent variable is log gross hourly wages (in swiss francs) deflated to December 2015 prices. Models (1), (2) and (3) use OLS, (4), (5) and (6) fixed effects. Married = 0 if the respondent is not married, 1 otherwise. Divorced/separated/widowed = 0 if the respondent is not divorced/separated/widowed, 1 otherwise. Demographic controls = age, qualification level, experience, experience squared, immigrant and disabled statuses. Job controls = company size, company sector, industry and occupation. Unstandardized regression coefficients and standard errors are shown. Errors are clustered at household level. * p < .05. ** p < .01.

TABLE 4: *Fixed Effects Models including Childcare Responsibility for Women*

	Log(Wage)	Log(Wage)
	(1)	(2)
Married	-0.02	-0.01
	(0.04)	(0.05)
Divorced/Separated/Widowed	0.04	0.04
	(0.06)	(0.08)
Number of Children	-0.02	-0.04*
	(0.01)	(0.02)
Cares for Ill Child	-0.06**	
	(0.02)	
Cares for Ill Child, t+2		-0.01
		(0.02)
Constant	4.52**	4.25**
	(0.78)	(1.12)
Demographic Controls	Yes	Yes
Job Region and Year Controls	Yes	Yes
Fixed effects	Yes	Yes
R^2	0.07	0.06
Observations	14055	8615

Note. SHP 2009-2016. Dependent variable is log gross hourly wages (in swiss francs) deflated to December 2015 prices. Married = 0 if the respondent is not married, 1 otherwise. Divorced/separated/widowed = 0 if the respondent is not divorced/separated/widowed, 1 otherwise. Cares for Ill Child = 0 if the respondent does not care for her ill children alone, 1 otherwise. Demographic controls = age, qualification level, experience, experience squared, immigrant and disabled statuses. Job controls = company size, company sector, industry and occupation. Unstandardized regression coefficients and standard errors are shown. Errors are clustered at household level. * p < .05. ** p < .01.

TABLE 5: *Fixed Effects Models for Self-Employed Women*

	Log(Wage)	Log(Wage)	Log(Wage)
	(1)	(2)	(3)
Married	-0.23	-0.24	-0.18
	(0.24)	(0.24)	(0.26)
Divorced/Separated/Widowed	-0.20	-0.23	-0.15
	(0.36)	(0.35)	(0.37)
Number of Children	0.07	0.10	0.11
	(0.06)	(0.07)	(0.07)
Constant	3.27**	7.62	5.87
	(0.21)	(4.36)	(5.60)
Demographic Controls	No	Yes	Yes
Region and Year Controls	No	No	Yes
Fixed effects	Yes	Yes	Yes
R^2	0.00	0.03	0.04
Observations	1257	1200	1200

Note. SHP 2009-2016. Dependent variable is log gross hourly wages (in swiss francs) deflated to December 2015 prices. Married = 0 if the respondent is not married, 1 otherwise. Divorced/separated/widowed = 0 if the respondent is not divorced/separated/widowed, 1 otherwise. Demographic controls = age, qualification level, experience, experience squared, immigrant and disabled statuses. Job controls = company size, company sector, industry and occupation. Unstandardized regression coefficients and standard errors are shown. Errors are clustered at household level. * p < .05. ** p < .01.

TABLE 6: *Fixed Effects Models for Managers vs Others*

	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)
	(1)	(2)	(3)	(4)
Married	-0.13	-0.06	0.02	0.03
	(0.07)	(0.05)	(0.06)	(0.03)
Divorced/Separated/Widowed	-0.29*	0.05	0.01	-0.01
	(0.11)	(0.07)	(0.10)	(0.04)
Number of Children	-0.01	-0.05**	0.00	-0.01
	(0.04)	(0.02)	(0.02)	(0.01)
Constant	6.34**	4.09**	-0.07	2.60**
	(2.11)	(0.87)	(2.18)	(0.72)
Demographic Controls	Yes	Yes	Yes	Yes
Region and Year Controls	Yes	Yes	Yes	Yes
Gender	Women	Women	Men	Men
Managers	Yes	No	Yes	No
Fixed effects	Yes	Yes	Yes	Yes
R^2	0.07	0.03	0.04	0.05
Observations	1019	14895	2269	12231

Note. SHP 2009-2016. Dependent variable is log gross hourly wages (in swiss francs) deflated to December 2015 prices. Married = 0 if the respondent is not married, 1 otherwise. Divorced/separated/widowed = 0 if the respondent is not divorced/separated/widowed, 1 otherwise. Demographic controls = age, qualification level, experience, experience squared, immigrant and disabled statuses. Job controls = company size, company sector, industry and occupation. Unstandardized regression coefficients and standard errors are shown. Errors are clustered at household level. * p < .05. ** p < .01.

TABLE 7: *Fixed Effects Models for Men vs Women*

	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)
	(1)	(2)	(3)	(4)
Married × Women	-0.09**	0.01	-0.04	-0.10**
	(0.03)	(0.23)	(0.07)	(0.04)
Married	0.05*	0.06	0.00	0.05*
	(0.02)	(0.17)	(0.07)	(0.03)
Divorced/Separated/Widowed	0.02	0.13	-0.12	0.02
	(0.04)	(0.19)	(0.07)	(0.04)
Number of Children × Women	-0.02	0.13*	-0.01	-0.02
	(0.02)	(0.07)	(0.04)	(0.02)
Number of Children	-0.00	-0.03	-0.00	-0.00
	(0.01)	(0.03)	(0.02)	(0.01)
Cares for Ill Child × Women	-0.07**	0.06	0.01	-0.07**
	(0.02)	(0.11)	(0.06)	(0.02)
Cares for Ill Child	0.00	0.08	-0.03	0.01
	(0.01)	(0.05)	(0.02)	(0.01)
Constant	3.58**	6.54	3.41**	3.64**
	(0.52)	(3.40)	(1.09)	(0.56)
Demographic Controls	Yes	Yes	Yes	Yes
Job Controls	Yes	No	Yes	Yes
Region and Year Controls	Yes	Yes	Yes	Yes
Self-employed	No	Yes	No	No
Job Function	All	All	Managers	Non Managers
Fixed effects	Yes	Yes	Yes	Yes
R^2	0.06	0.04	0.07	0.06
Observations	27184	2585	2963	24221

Note. SHP 2009-2016. Dependent variable is log gross hourly wages (in swiss francs) deflated to December 2015 prices. Married = 0 if the respondent is not married, 1 otherwise. Divorced/separated/widowed = 0 if the respondent is not divorced/separated/widowed, 1 otherwise. Cares for Ill Child = 0 if the respondent does not care for her ill children alone, 1 otherwise. Women = 0 if the respondent is a man, 1 otherwise. Demographic controls = age, qualification level, experience, experience squared, immigrant and disabled statuses. Job controls = company size, company sector, industry and occupation. Unstandardized regression coefficients and standard errors are shown. Errors are clustered at household level. * p < .05. ** p < .01.

TABLE 8: *OLS and Fixed Effects Models for Men, Earlier Sample*

	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)
	(1)	(2)	(3)	(4)	(5)	(6)
Married	0.72**	0.09**	0.05*	0.17**	-0.02	-0.03
	(0.03)	(0.03)	(0.02)	(0.04)	(0.04)	(0.04)
Divorced/Separated/Widowed	0.60**	-0.03	-0.03	0.24**	0.05	0.06
	(0.04)	(0.04)	(0.03)	(0.06)	(0.06)	(0.06)
Number of Children	-0.02*	0.01	0.02*	-0.03*	-0.02	-0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Constant	3.23**	2.03**	2.77**	3.59**	1.06	1.27
	(0.03)	(0.06)	(0.07)	(0.03)	(0.57)	(0.87)
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Job Region and Year Controls	No	No	Yes	No	No	Yes
Fixed effects	No	No	No	Yes	Yes	Yes
R^2	0.18	0.38	0.47	0.00	0.08	0.11
Observations	8363	7999	6138	8363	7999	6138

Note. SHP 2004-2008. Dependent variable is log gross hourly wages (in swiss francs) deflated to December 2015 prices. Models (1), (2) and (3) use OLS, (4), (5) and (6) fixed effects. Married = 0 if the respondent is not married, 1 otherwise. Divorced/separated/widowed = 0 if the respondent is not divorced/separated/widowed, 1 otherwise. Demographic controls = age, qualification level, experience, experience squared, immigrant and disabled statuses. Job controls = company size, company sector, industry and occupation. Unstandardized regression coefficients and standard errors are shown. Errors are clustered at household level. * p < .05. ** p < .01.

TABLE 9: *OLS and Fixed Effects Models for Women, Earlier Sample*

	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)	Log(Wage)
	(1)	(2)	(3)	(4)	(5)	(6)
Married	0.25**	-0.10**	-0.01	0.26**	-0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.07)	(0.07)	(0.07)
Divorced/Separated/Widowed	0.39**	0.06	0.06	0.43**	0.11	0.09
	(0.04)	(0.04)	(0.03)	(0.10)	(0.10)	(0.12)
Number of Children	-0.11**	-0.07**	-0.01	-0.14**	-0.08**	-0.09**
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Constant	3.23**	2.30**	2.86**	3.24**	-0.15	-2.33*
	(0.03)	(0.06)	(0.08)	(0.05)	(0.71)	(0.93)
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Job Region and Year Controls	No	No	Yes	No	No	Yes
Fixed effects	No	No	No	Yes	Yes	Yes
R^2	0.03	0.22	0.36	0.01	0.06	0.11
Observations	9228	8748	6877	9228	8748	6877

Note. SHP 2004-2008. Dependent variable is log gross hourly wages (in swiss francs) deflated to December 2015 prices. Models (1), (2) and (3) use OLS, (4), (5) and (6) fixed effects. Married = 0 if the respondent is not married, 1 otherwise. Divorced/separated/widowed = 0 if the respondent is not divorced/separated/widowed, 1 otherwise. Demographic controls = age, qualification level, experience, experience squared, immigrant and disabled statuses. Job controls = company size, company sector, industry and occupation. Unstandardized regression coefficients and standard errors are shown. Errors are clustered at household level. * p < .05. ** p < .01.

TABLE 10: *Fixed Effects Models with Time Variables for Women, Earlier Sample*

	Log(Wage)	Log(Wage)
	(1)	(2)
Married	0.01	-0.00
	(0.03)	(0.03)
Divorced/Separated/Widowed	0.05	0.05
	(0.05)	(0.05)
Years	0.03**	
	(0.01)	
Number of Children	-0.03	-0.04**
	(0.02)	(0.01)
Years × Number of Children	-0.00	
	(0.00)	
After 2008		0.33*
		(0.15)
After 2008 × Number of Children		-0.00
		(0.01)
Constant	2.96**	3.16**
	(0.22)	(0.55)
Demographic Controls	Yes	Yes
Job Controls	Yes	Yes
Region and Year Controls	Yes	Yes
Fixed effects	Yes	Yes
R^2	0.14	0.13
Observations	19026	20932

Note. SHP 2004-2016. Dependent variable is log gross hourly wages (in swiss francs) deflated to December 2015 prices. Married = 0 if the respondent is not married, 1 otherwise. Divorced/separated/widowed = 0 if the respondent is not divorced/separated/widowed, 1 otherwise. Years= 1 for 2004, 2 for 2005... 12 for 2015. After 2008 = 1 for years 2009... 2015, 0 otherwise. Demographic controls = age, qualification level, experience, experience squared, immigrant and disabled statuses. Job controls = company size, company sector, industry and occupation. Unstandardized regression coefficients and standard errors are shown. Errors are clustered at household level. * p < .05. ** p < .01.

More Than Just Window Dressing: Fostering Age Diversity In Teams Through Pro-Diversity Culture Communication

Thesis Paper

Oriana de Saint Priest, Franciska Krings

Department of Organizational Behavior

University of Lausanne

Switzerland, 2022

Contents

Abstract

To promote diversity and inclusion, organizations frequently use diversity communications, for example diversity statements that highlight the value of diversity and work in diverse teams. Considering the current demographic ageing and continued discrimination against older workers, many organizations explicitly include age or generation in their communications, for example underlining the benefits of working in age-mixed teams. We suggest that diversity communication is more than just window dressing but may effectively foster diversity by signaling clearly what is valued and desirable in the organization. In five experimental studies, we investigate the impact of pro-age diversity statements on age diversity in newly constituted teams, its underlying processes, and potential unintended consequences, i.e., decrease of team diversity on other dimensions than age. We found consistent evidence that pro-age diversity statements increase age diversity in teams. Moreover, in the absence of a diversity statement, older teammates were hardly ever chosen. Our research also shows that the positive impact of diversity statements is primarily due to an increased salience of the desirability of work in diverse teams. Importantly, pro-age diversity communication did not reduce the diversity of the final team in terms of personality or gender. Results of this research are encouraging for both research and practice struggling to identify diversity initiatives that effectively influence behavior and increase the representation of under-represented groups in organizations.

Introduction

Efforts to reduce discrimination and increase inclusion in the workplace date back to the race relations workshops offered by federal agencies in the late nineteen sixties. These diversity initiatives have kept flourishing since then. In fact, an estimated three quarters of firms implement diversity practices, defined as organizational measures that aim to increase the fairness of the work environment by promoting the inclusion, hiring, retention, and promotion of underrepresented groups (Dover et al., 2020; Bartels et al., 2013). Moreover, organizations and businesses are increasingly allocating resources to these initiatives, up to an average $16 billion a year for Fortune 500 companies (Staley, 2017), as workforce diversity has been growing in recent years and will certainly continue to do so in the future. Globalization and an increasing number of women joining the labor force have meant that the workforce has become even more diverse in terms of ethnicity, race, skin color and gender (Staley, 2017). At the same time, demographic aging has become a global phenomenon, marked by decreasing fertility rates and increased life expectancies around the globe. Western countries in particular experience historically low fertility rates and high life expectancies, with individuals spending a larger number of years in good health after official retirement age. These factors have led to an important increase in the age diversity of the workforce (Dychtwald et al., 2004; Tempest et al., 2002).

To respond to these challenges and ensure that diverse workers are given an equal chance to thrive in the workplace, a wide variety of diversity initiatives has been developed, including diversity trainings, pro-diversity value statements and communication, targeted recruitment efforts and mentorship programs (Berrey, 2021; Kalev et al., 2006; Leslie, 2019). Though these initiatives are widespread, they often do not show the desired results. Studies on the effects of diversity trainings reveal contradictory results, with some studies showing that these diversity initiatives have negative effects, others positive, and yet others no effects (Bezrukova et al., 2012, 2016; Kalev et al., 2006; Roberson & Kulik, 2007). Moreover, most studies have focused on the impact of diversity initiatives on individual attitudes and intentions, and much less on

behaviors (Bezrukova et al., 2016; Chang et al., 2019; Dobbin & Kalev, 2016; Kalinoski et al., 2013). Thus, there is a pressing need for research and practice investigating the behavioral outcomes of diversity initiatives (Devine & Ash, 2022).

In this article, we focus on pro-diversity value or culture communication, one of the most frequently used diversity initiative looking to increase fairness in the workplace, most often put in place through diversity policies. We examine the impact of organizational pro-age diversity communication on team age diversity in five studies. It is particularly important to investigate how to increase age diversity as discrimination against older workers is widespread and still understudied (Suh, 2021). Drawing on symbolic management, defined as the way management frames and portrays the organization to its members (Roberge et al., 2011), we propose that pro-age diversity culture communication increases the representation of older workers in teams. We focus on teams specifically as teams are ubiquitous in organizations, with the benefits of work in diverse teams being manifold (Bell et al., 2011).

In particular, we propose that pro-diversity culture communication highlights the desirability of work in diverse teams, and that in turn the perceived desirability of work in diverse teams increases team diversity. We hypothesize that this mechanism is particularly important, as organizational values reflected in pro-diversity culture communication indicate desirable behaviors to be adopted by employees. Finally, we investigate to what extent pro-diversity culture communication may also produce unintended consequences. Theoretical frameworks on the conceptualization and effects of diversity initiatives have highlighted that many initiatives create unintended consequences that may hinder the attainment of the very goals that they aim to achieve (Dover et al., 2020; Leslie, 2019). We focus on negative spill-over, i.e., effects of an initiative in an undesirable direction (Leslie, 2019), suggesting that pro-age diversity communications may increase team age diversity but at the same time reduce diversity in terms of personality and values, due to similarity attraction dynamics. If this mechanism is at play and pro-diversity culture communication decreases deep-level diversity in teams, then this is particularly problematic as

deep-level diversity has been associated with positive effects on team performance (Stahl et al., 2010).

Taken together, our research contributes in many important ways to the diversity literature. First, it promotes a better understanding of diversity initiatives and of their impact on a yet under-studied dimension of diversity. By revealing that diversity initiatives which are easy to implement can have effects on team diversity, this research shows that many businesses can start implementing effective diversity measures without huge upfront costs. Moreover, this research sheds light on the mechanisms explaining the behavioral impact of diversity initiatives, suggesting why they produce the desired effect. This understanding is crucial as it provides guidance to practitioners and researchers alike, both interested in developing diversity measures with proven consequences on behavior. In addition, this work helps us understand under which conditions individuals choose to work with diverse others, investigating the potential limits of diversity initiatives. In fact, it is crucial to assess whether increasing diversity in a certain dimension might yield unintended side effects, in particular decreasing diversity in another important dimension.

Literature Review

Diversity Initiatives

Diversity initiatives are organizational practices "aimed at improving the workplace experiences and outcomes of groups that face disadvantages in society" (Leslie, 2019). More specifically, they seek to increase the inclusion and representation of members of disadvantaged or underrepresented groups in all units and at all levels of the hierarchy of the organization.

Organizations implement diversity initiatives for three main reasons: to create fairer, more efficient and effective workplaces, and to communicate their values (Dover et al., 2020). These reasons are not mutually exclusive, and many organizations pursue two or all three of them simultaneously. Firstly, by adopting diversity initiatives, organizations seek to create a work environment in which discrimination

against protected groups is eliminated. Indeed, many initiatives originated as anti-discrimination efforts, motivated by the fundamental rationale to end discrimination. To reach this goal, diversity initiatives sought to ensure that discrimination cases could be reported and remediated, without fear of retaliation, thereby ensuring compliance with anti-discrimination laws. Second, organizations may implement diversity initiatives to create a more efficient and effective work environment. This rationale is related to the so-called business case argument for diversity, which emerged in the nineteen nineties. Its primary argument to promote diversity is that diversity is beneficial for team performance, and thus helps create a more efficient and effective workplace (Cox & Blake, 1991). Indeed, diverse teams have been shown to be more creative and innovative (Rock & Grant, 2016), better at problem solving (Stevens et al., 2008) and faster at information processing (Kearney et al., 2009). Third, organizations may implement diversity initiatives in order to communicate their values, in particular to current and potential employees, but also to clients, and to the general public. By outlining their commitment to diversity and the emphasis they put on diversity and inclusion, for example on their website or in an internal newsletter, organizations communicate their pro-diversity values and thus signal, among other things, what type of behavior they expect from their current employees and what type of workplace potential employees can expect to find. Such signaling may then contribute to strengthening the diversity climate or culture, i.e., the "aggregate perceptions about the organization's diversity-related formal structure characteristics and informal values" (Gonzalez & DeNisi, 2009) and hence indirectly promote a fairer and more productive work environment (Marques, 2010). Indeed, signaling that an organization cherishes diversity and inclusion can increase the number of applications from diverse candidates, especially if the organization signals that employee diversity is unrestricted and extends beyond merely entry-level positions (Avery, 2003). Moreover, such signaling provides guidelines to existing employees on how to treat and interact with individuals from diverse backgrounds in specific situations, aligned with the organization's diversity values (Hicks-Clarke & Iles, 2000; Avery et al., 2007; McKay et al., 2007).

In order to create fairer and more efficient workplaces while communicating their values, organizations have developed a variety of diversity initiatives. Most commonly, these include diversity trainings, pro-diversity culture communication, traditional affirmative action policies, targeted recruitment efforts, diversity committees and personnel, mentorship programs and affinity groups (Berrey, 2021; Kalev et al., 2006; Leslie, 2019). Compared to the other initiatives, pro-diversity culture communication has received considerably less attention in the literature. This is somewhat surprising given that pro-diversity culture communication is very widespread and may indeed be a powerful tool, as outlined in the following section.

Pro-Diversity Culture Communication

All organizations have a specific organizational culture, i.e., organizational values and beliefs that provide norms of desirable behaviors for employees (Schein, 2010). Values act as social principles or philosophies which inform behavior and set the framework for organizational practices (O'Reilly III et al., 1991; Hatch, 1993). For instance, by promoting certain values and associated behavioral norms, managers can endorse an organizational culture with far-reaching effects on employee behavior (Tellis et al., 2009; Mumford et al., 2002). Previous research has shown that organizational culture significantly impacts employee attitudes (Alas & Vadi, 2006; Farrell, 2005; Webber et al., 2010) and behaviors (Boye & Jones, 1997). More specifically, diversity culture or climate, i.e., the perceived extent to which a company advocates fair human resource policies and socially integrates underrepresented employees, has positive effects on employee satisfaction, commitment and performance (Holmes IV et al., 2020; Avery et al., 2007; McKay et al., 2007; Hicks-Clarke & Iles, 2000). However, effects tend to be different depending on whether the employee is part of a minority or majority group. For instance, White viewers are not increasingly attracted to an organization explicitly endorsing diversity in its job advertisements, while Black viewers are increasingly attracted to such an organization, but only when the job ad specifies that

diversity support extends to supervisory level positions (Avery, 2003). Moreover, the effect of race on reactions to ad diversity is contingent on the viewer's openness to racial diversity (Avery, 2003). In the same way, it appears that pro-diversity climate impressions are related to favorable worker attitudes and reduced turnover intentions, though all members of a minority group do not respond identically to a firm's diversity climate (McKay et al., 2007). Hence organizational culture is a powerful means to attain desired outcomes, though there are differences in the way in which members from minorities react to these.

Pro-diversity culture communication, explicitly transcribing a firm's commitment to diversity, frequently highlights that diversity is particularly valuable in teams, as benefits of diversity are reaped in teams with diverse individuals (Rau & Hyland, 2003). For instance, UBS states that "a diverse workforce and inclusive culture are crucial to (their) long-term success. That's because, in (their) experience, teams with diversity in gender, race, age, ethnicity, education, background, sexual orientation, gender identity, etc. better understand and relate to (their) equally diverse clients' needs" (UBS, 2017). This signal is largely correct, as diversity has a positive impact on several aspects of team performance processes and outcomes (Horwitz & Horwitz, 2007). In particular, it has been shown that encouraging work in age diverse teams is particularly beneficial for performance (Wegge et al., 2012). This is especially true when tasks require complex decision-making without time pressure, consistent with information processing models. In fact, these models propose that age-related deficits may be reduced by high task variety and be counteracted by selection, optimization and compensation strategies under complex task requirements without too much time pressure (Wegge et al., 2012).

Overall, pro-age diversity culture communication may therefore be a powerful yet largely unexplored tool in the attempt to create a more diverse and inclusive organization. In fact, most research on the effectiveness of diversity initiatives has focused on training and produced mixed results, as we outline in the following paragraph.

Effects of Diversity Initiatives

While organizations implement diversity initiatives to create a fairer and more diverse workplace, evidence shows that this is not necessarily the case. Research on the impact of diversity initiatives has primarily focused on diversity trainings, and their effects on individuals. These studies reveal that diversity trainings can have positive effects on participants' knowledge and attitudes towards diversity and underrepresented groups (Kalinoski et al., 2013). Yet, these positive effects only occur in the short-term (Kulik & Roberson, 2008; Roberson et al., 2013; Bezrukova et al., 2016). In addition, it is important to consider whether changes in knowledge or attitudes translate into behavior, for example, at hiring, or when creating teams. Arguably, because diversity initiatives aim at increasing the diversity of the workforce, an increased representation of minority employees is the most crucial outcome of such initiatives. However, few studies have looked at the effects of diversity trainings on individual behavior, for example toward minority applicants, and workplace outcomes of members from underrepresented groups. The few that do find contradictory results. For example, they find null or negative effects of diversity trainings on the careers of minorities (Kellough & Naff, 2004) and marginal, null or even negative effects on the representation of minority individuals in managerial positions (Dobbin & Kelly, 2007; Kalev et al., 2006). Likewise, evidence for the effectiveness of other diversity initiatives is mixed. For example, networking groups are associated with an increased representation of women but a decreased representation of Black men (Kalev et al., 2006). On the other hand, managerial feedback through performance evaluations could reduce bias in recruitment, as laboratory experiments show that when subjects are aware that their decisions will be reviewed by experimenters, they show lower levels of bias when assigning jobs (Salancik & Pfeffer, 1978; Tetlock, 1985).

The effects of pro-diversity culture communication have not yet been widely explored, though some recent studies start to investigate the way in which diversity is defined on corporate websites (Point & Singh, 2003; Singh & Point, 2006) as well as its impact on individual performance (Wilton et al., 2015;

Windscheid et al., 2016). We propose that pro-diversity culture communication may be particularly effective in positively influencing behavior and thus workplace outcomes for underrepresented groups. In fact, pro-diversity culture communication explicitly promotes certain desirable values and associated behavioral norms through symbolic management, which employees then internalize and make their own (O'Reilly III et al., 1991). Hence pro-diversity culture communication may be particularly effective in influencing behavior, the most crucial outcome of diversity initiatives.

In light of the theoretical arguments and empirical evidence above, we propose that pro-age diversity culture communication increases age diversity. More specifically, we expect pro-age diversity culture communication to increase age diversity in teams, by increasing the selection rates of older employees into teams and thus their representation.

H1: Pro-age diversity culture communication will increase age diversity in teams by increasing the selection of older adults into work teams.

Moreover, we propose that pro-diversity culture communication affects behavior by signaling which behavior is desirable in the given situation. More precisely, we hypothesize that pro-diversity culture communication increases perceptions of desirability of work in age diverse teams, which in turn increases the selection rate of older adults into teams and hence team age diversity.

H2: Pro-age diversity culture communication will be positively related to the perceived desirability of work in age diverse teams.

H3: Perceived desirability of work in age diverse teams will be positively related to selecting older adults into work teams.

Finally, we investigate whether in addition to yielding intended consequences such as increasing the representation of minorities into teams, pro-diversity culture communication might also have unintended consequences. Indeed, diversity initiatives could have unintended adverse consequences, due to what they

signal. For instance, there is empirical evidence that diversity initiatives can prompt members to overlook diversity issues (Kaiser et al., 2013; Dover et al., 2014; Kirby et al., 2015) as they signal to majority groups that an organization treats minorities fairly, even when there is evidence to the contrary (Kellough & Naff, 2004). Moreover, the mere presence of diversity initiatives can lead advantaged groups to believe that an organization is unduly favoring minorities, making them feel more threatened and less included (Dover et al., 2016). Diversity initiatives also act indirectly as signals for competence, suggesting that members of minority groups might need to be helped in order to succeed in the workplace. In fact, members of advantaged groups often presume that a diverse hire is less competent than an otherwise similar diverse hire, in the presence of a diversity initiative (Gündemir et al., 2017). Moreover, some members of disadvantaged groups report feeling less competent if the organization they interview for explicitly promotes diversity, and regardless of their actual interview outcome (Dover et al., 2020). These feelings leading to cognitive depletion can then hinder actual competence on intellectual tasks (Carter et al., 2015; Schmader & Johns, 2003).

More generally, it is likely that work in diverse teams yields general feelings of discomfort and uneasiness, as it runs counter to similarity attraction. Similarity attraction refers to the fact that people are attracted to others who are similar to themselves, as similar attitudes serve as reinforcers (Clore & Byrne, 1974). Since individuals need and seek a consistent view of the world, they are more likely to shun the people who disagree with them. This way individuals reduce their own anxiety, uncertainty and confusion thanks to the unified version of the world they create (Clore & Byrne, 1974). In the work place, individuals also thrive to reduce uncertainty by controlling their work networks (Kanter, 1977), in order to regain control over the fate of the organization (Thompson, 1967). Hence, it is likely that individuals will choose to work with similar others if they are given this opportunity. Previous research has indeed established that people prefer work group members who are similar to themselves (Hinds et al., 2000). This is likely to be particularly problematic, as team deep-level diversity, including personality, values and attitudes, has an

effect on team performance. Indeed different studies have assessed how the personality traits of team members affect team functioning and outcomes, highlighting the positive effects of personality diversity on performance (Barrick et al., 1998; Barry & Stewart, 1997; Heslin, 1964; Neuman & Wright, 1999; Thoms et al., 1996; Van Vianen & De Dreu, 2001). Seeing the positive impact of deep-level diversity on performance, it is particularly important to assess whether pro-diversity culture communication backfires by resulting in decreased deep-level diversity, in line with similarity attraction. In fact, this could be the case if individuals mitigate the perceived uncertainty of work with demographically diverse others by working with colleagues similar to them on the deep-level. Hence we investigate whether pro-diversity culture communication decreases team diversity in terms of deep-level characteristics, including personality and world views.

R1: Does pro-age diversity culture communication decrease team diversity in terms of personal characteristics and values?

Overview of the Studies

We conduct five studies to investigate the impact of pro-diversity culture communication, instantiated by a diversity policy, on team age diversity (see Table 1 for an overview). All our studies are conducted online on Mechanical Turk with participants aged below 30 years old. We purposefully focused on a younger sample of participants, as ageism is more pronounced in this age group compared to other age groups (Rupp et al., 2005). Moreover, in-group out-group dynamics are particularly likely to be at play in such settings. Study 1 is a scenario-based study in which participants are asked to choose teammates for an upcoming business project crucial for their career. Studies 2 to 4 use a more realistic setting in which participants have to choose teammates for a quiz that they also take, where team performance is rewarded financially by a bonus. Study 5 is a follow-up study where we investigate how the settings of studies 1 to 4 were perceived by participants. In our first four studies, we examine the impact of pro-diversity culture

communication on team age diversity, also testing whether the behavioral impact of the diversity policy can be explained by its effect on the perceived desirability of work in age diverse teams. In addition, we investigate whether pro-diversity culture communication has unintended side effects by promoting team personality homogeneity. In studies 1, 3 and 4, we study whether pro-diversity culture communication decreases perceived team personality diversity. In study 2, we investigate whether pro-diversity culture communication decreases diversity in world views.

Study 1

Method

Design and procedure. The experiment had 2 conditions (company diversity policy: yes or no), with diversity policy as a between-subjects factor. Data collection was done on Amazon Mechanical Turk (MTurk). Participants were randomly assigned to one of the two experimental conditions.[1] In all the conditions, participants read a business scenario in which they were described as solely responsible for a challenging project. The importance of this business project was emphasized, as it was made clear that the participant's upcoming promotion depended on the project's outcome. Moreover, participants were described as very happy with their job and with their work environment. Participants were further informed that they had to create a project team of four to complete this important project. There were already two people on the team: the participant and another member similar in age, described as having started college at the same time. We avoided using direct references to age so that the study's aim did not become immediately clear to participants, reducing the likelihood of demand effects (Khademi et al., 2021). In this first study, participants got to select two teammates out of a list of four, all described as equally competent. The four teammates available for selection varied across two dimensions: personality similarity with the participant and age. More precisely, the candidates were described in the following way: "The four collaborators currently available to work on your project are Robert (56 years old), David (24 years old),

Michael (52 years old), and Paul (26 years old). You have met all four of them and got an idea of what they are like. You noticed that Robert and David are pretty similar to you in terms of their personalities, while Michael and Paul seem to be less similar to you."

In addition, half of participants were randomly assigned to the company diversity policy condition. If they were in the diversity policy condition, participants were told that the company they worked for cherished age diversity in teams and encouraged people to work in teams that are mixed in terms of age. This statement appeared on a separate screen. If participants were not in the diversity policy condition, they were not given any specific information about the company's diversity guidelines. We preferred a control condition with no diversity policy, as opposed to a control condition with a placebo policy, for ecological validity reasons. In fact, prior to displaying their diversity policies, businesses did not explicitly mention related guidelines, rather than advocating placebo guidelines.

Participants. Participants were 105 people aged under 30 (55% men; mean age 25.51, SD 2.81). 53% were employed full-time, 24% were employed part-time and the remaining 23% were unemployed. Out of the people currently employed, 71% indicated that they had a supervisory function in their job.

Dependent measures. Participants had to choose two teammates from the list of four. The four potential male team members differed in terms of age and personality similarity with the participant. They were either younger (23-26 years old) or older (51-56 years old) and with a different or similar personality to the participant. To create an age diverse team, participants had to select two older candidates, as the two candidates already present on the team were both younger candidates.

Process variables. After making their teammate choices, participants were asked to report how much they thought about some common features of diversity when they made their teammate selection. More precisely, participants were asked "Please think carefully now about *how* you selected the two teammates you chose. How much did you consider the following points when making your decision", on a

scale from 0 to 100. The items on the list were "choosing these people will increase the team's performance", "choosing these people will make sure the team is happy working together", "choosing these people will bring more talent to the team", "choosing these people will not slow down the team's work" and "choosing these people is the morally right thing to do". In addition, participants had to report whether they considered the following aspects when making their decision about who to choose for their team: "creating high performing age diverse teams is very desirable" and "creating high performing age diverse teams is easily feasible".

Control variables. Because participants' own gender, experience and supervisor status may influence their teammate selection decisions, gender, experience and supervisory duties were included as control variables. Gender was coded as a female dummy, equal to 0 if the participant was a man and 1 otherwise. Supervisory duties values were coded as 0, if the participant indicated having no supervisory duties, 1 otherwise. Experience was coded on a 5-point Likert scale from 0 if the participant had no experience working in teams mixed in terms of age, up until 4 if the participant had a lot of experience working in teams mixed in terms of age.

Additional exploratory outcome variables. Participants were asked to rate how nice they thought the two people they chose were, on a 7-Point Likert scale from "not nice at all" to "very nice", and how enjoyable they thought it would be to work with them, on a similar scale from "not enjoyable at all" to "very enjoyable". They were additionally asked to rate how diverse they thought the team they created was, both in terms of age and personality, also on a 7-Point Likert scale from "not diverse at all" to "very diverse".

Attention checks. To ensure data quality, participants had to answer "No" among various responses (including "Yes", "Maybe" and "No"). At the end of the survey, they were also asked whether in their honest opinion, we should use their data in our analyses for this study. They were reassured that they would be remunerated regardless of their answer to this question.

Manipulation checks. The effectiveness of the company diversity policy manipulation was tested with a question asking whether the company particularly encouraged age diversity in teams, where the correct answer varied across conditions. Moreover, participants had to answer a close-ended question about the characteristics of the teammates they had to choose from when they created their team. They could choose one answer out of three, where the correct answer was that all the teammates had different ages. The other answers they could pick from were that the teammates were all female, and that the teammates were all above 50 years old. Lastly, participants had to identify who they created the team for, where the correct answer was for themselves.

Results

Attention checks. No one replied incorrectly to the attention check asking to answer "No". However one participant was excluded from the analysis as he indicated that he would not use his data in our study.

Manipulation checks. Fifteen participants failed to recognize correctly the presence of the age diversity policy in their fictitious business scenario. Hence they were excluded from further analyses. Another participant did not understand that the teammates he could pick from all had different ages and was thus also excluded. Moreover, two participants failed to indicate that they had to create the team for themselves. A further four participants chose not to disclose information about their current supervisory duties. The final sample consisted of 82 participants, with 39 participants in the no diversity policy condition and 43 participants in the diversity policy condition.[2]

Age diversity. Correlations and descriptive statistics are presented in Table 2. To test our hypotheses, we conducted linear regressions with robust standard errors. Team age diversity was measured by the number of older teammates selected to join the participant's team. We coded *older* as 0 for no selected older candidate, 1 for one selected older candidate and 2 for two selected older candidates. We coded *diversity policy* as 0 if the participant was in the no diversity policy condition, 1 if he-she was in the

diversity policy condition. We coded *same* as 0 if the participant chose only teammates with different personalities to him-her, 1 if the participant chose a teammate with a similar personality to him-her and 2 if the participant chose two teammates with a similar personality to him-her. We first predicted team age diversity with the company diversity policy, including control variables in Table 3.

In line with Hypothesis 1, participants were more likely to select an older individual into their team when there was an age diversity policy in place (M = 0.19, SD = 0.08, t = 2.33, p = 0.02). In fact, the diversity policy increased the number of older workers such that there was one more older worker per about five teams. Hence, pro-age diversity culture communication increases age diversity in teams by increasing the selection of older adults into work teams. Moreover, there was no significant effect on the selection of older adults into teams of the participants' gender (M = 0.00, SD = 0.09, t = 0.00, p = 0.99), supervisory duties (M = 0.05, SD = 0.12, t = 0.38, p = 0.71) and experience (M = 0.04, SD = 0.04, t = 0.96, p = 0.34). Next, we investigated whether there was a mediation effect of the process variables on the relationship between the company diversity policy and team age diversity, controlling for the supervisor status, gender and experience of participants. To do so, we firstly checked whether the diversity policy had an effect on the features of diversity participants thought about when making their teammate selection decisions (Table 4).

Consistent with Hypothesis 2, we find that the diversity policy has a significant positive effect on the perceived desirability of work in age diverse teams (M = 20.87, SD = 5.64, t = 3.70, p = 0.00). Thus, pro-age diversity culture communication is positively related to the perceived desirability of work in age diverse teams. Moreover, there was no significant effect on the perceived desirability of work in age diverse teams of the participants' gender (M = -2.64, SD = 5.51, t = -0.48, p = 0.63) and supervisory duties (M = -3.10, SD = 7.16, t = -0.43, p = 0.67), but a significant positive effect of experience (M = 6.62, SD = 2.64, t = 2.50, p = 0.01). Additionally there is no significant effect of the diversity policy on the perceived feasibility of high performing age diverse teams. In the same way, the diversity policy has no effect on an

index made by the sum score of the advantages of diversity. These included information about whether the participant felt that choosing these particular teammates will increase the team's performance, make sure the team is happy working together, bring more talent to the team, not slow down the team's work and is the morally right thing to do. In addition, the diversity policy does not affect the perceived characteristics of teammates chosen, neither in terms of how nice they are perceived to be nor in terms of how enjoyable it would be to work with them. However the diversity policy has a significant positive effect on the perceived team age diversity (M = 0.64, SD = 0.03, t = 2.29, p = 0.03), which can be expected to correlate with team age diversity.

In line with Hypothesis 3, we further report that the perceived desirability of work in age diverse teams is positively related to selecting older adults into work teams (M = 0.00, SD = 0.00, t = 2.56, p = 0.01), as shown in Table 5. Moreover, there was no significant effect on the selection of older adults into teams of the participants' gender (M = 0.01, SD = 0.09, t = 0.11, p = 0.92), supervisory duties (M = 0.06, SD = 0.13, t = 0.46, p = 0.65) and experience (M = 0.13, SD = 0.04, t = 0.34, p = 0.73). We next investigated further whether there was a mediation effect of diversity's perceived desirability on the relationship between company diversity policy and team age diversity, controlling for supervisor status, gender and experience of participants. To do so, we used the Stata SEM function (Hayes, 2012) for mediated moderation bootstrapping analyses with the relative desirability of high performing teams entered as a covariate, creating 500 bootstrap samples by randomly sampling observations with replacement from the original dataset (Preacher & Hayes, 2004). Ninety-five percent confidence intervals show a significant indirect effect of company diversity policy on age diversity through the relative salience of the desirability of high performing age diverse teams (95% CI = [.00, .10]). Hence we find support for the fact that the effect of the diversity policy seems to be mediated by the perceived desirability of high performing age diverse teams (Figure 1).

Yet these results should be interpreted with caution as beliefs were not exogenous, since they were

not manipulated in our study. The Durbin-Wu-Haussman test suggests beliefs are indeed endogenous, as the residuals in the augmented regression are significantly different from 0 ($F(1,76) = 5.11$; $p = 0.03$). We therefore follow the approach suggested in the literature, using our exogenous treatment variation as an instrument for the desirability of work in age diverse teams (Antonakis et al., 2010; Sajons, 2020). Specifically, we perform a two-stage least squares estimation in which the first stage regresses the desirability of work in age diverse teams on a variable indicating the company diversity policy, and controlling for participants' gender, supervisor status and experience. The second stage then uses the predicted values from the first stage as an exogenous regressor to estimate the effects of the diversity policy on age diversity. Yet the first-stage F-statistic shows the strength of our instrument does not pass the required critical value of 10 (Stock & Yogo, 2005), $F(4,77) = 7.38$. According to previous literature, reduced form estimates should thus be used for interpretation purposes (Sajons, 2020). At this point, we investigated whether the age diversity policy had unintended side effects, driving participants to select increasingly teammates more similar to them in terms of personality.

Our results further reveal that pro-age diversity culture communication does not significantly decrease team diversity in terms of personality ($M = 0.03$, $SD = 0.13$, $t = 0.21$, $p = 0.84$), as seen in Table 3. In addition, there was no significant effect on the selection of personally similar adults into teams of the participants' gender ($M = 0.12$, $SD = 0.14$, $t = 0.91$, $p = 0.37$), supervisory duties ($M = -0.08$, $SD = 0.17$, $t = -0.47$, $p = 0.64$) and experience ($M = -0.03$, $SD = 0.06$, $t = -0.49$, $p = 0.63$). However we find that participants are more inclined to select teammates for their team when they perceive them as similar in personality, independently of the policy. In fact, only 5% of participants chose two teammates different in personality to themselves for their team, in line with the similarity attraction principle.[3]

Discussion. This study shows that diversity policies are successful at favoring collaboration with more diverse teammates, supporting Hypothesis 1. In fact, participants are more likely to choose older teammates when there is a company age diversity policy, hence constituting evidence that pro-diversity

culture communication is successful at increasing team age diversity. We also report that pro-age diversity culture communication is positively related to the perceived desirability of work in age diverse teams. In turn, the perceived desirability of work in age diverse teams is positively related to selecting older adults into work teams. Thus there is evidence that the diversity policy's success can be at least partly explained by its positive impact on the perceived desirability of work in diverse teams. Moreover, this study finds no support for the fact that diversity policies might specifically increase personality homogeneity in teams. Hence pro-age diversity culture communication does not appear to decrease team diversity in terms of personal characteristics and values, which could have been problematic for performance. Study 2 aims to check the robustness of these results by making the teammate choices incentive compatible. In fact, participants in study 2 are rewarded financially depending on the performance of the teammates they choose. Moreover, study 2 provides a more realistic description of the teammates available for selection, listing their age, gender and value orientation. In addition, study 2 investigates what drives the effect of the diversity policy on teammate choices using a variety of different likely process variables. In fact, we hypothesize that the process variables in the present study were worded in a way that might be too generic for participants. In the next study, we change this wording to make it more relevant to the study's setting. Hence study 2 hopes to test the robustness of the positive effect of policy found in study 1 and shed further light on the mechanisms explaining this result.

Study 2

Method

Design and procedure. The experiment had 2 conditions (company diversity policy: yes or no), with diversity policy as a between-subjects factor. In contrast with the previous study, real incentives were used. Participants were given a similar scenario to that of study 1, except in this study participants were told that their teammate choices would have a financial impact on them. In fact, bonus payments depended on the

performance on a food quiz of the teammates that participants chose, as well as on the participants' own performance. To determine the performance of the teammates chosen, a pre-study was conducted on MTurk, in which participants with the same gender, age and world view combinations as the teammates available for selection in the main study took the food quiz. From this data, an average performance per teammate with a certain gender, age and world view combination was computed and used to determine payment in the main study. In the main study, participants had to choose two teammates out of a list of eight. The eight teammates available for selection varied across three dimensions: age (younger from 25 to 30 years old or older from 50 to 57 years old), gender and world views. The age ranges of the teammates available for selection were similar to those used in the previous study. For instance, teammate D was described in the following way: "David is 25 years old. He believes that it is a dog-eat-dog world where one has to be ruthless at times, in particular in the business world". The same eight teammate profiles were shown across the two experimental conditions. In contrast with the previous study, there was no one already on the team, so it was possible to create an entirely age homogeneous team.

The wording of the business scenario was consistent across studies, with the diversity policy appearing on a separate screen too. However in this study, the diversity policy specified that in view of the current demographic aging, it was very important to encourage work with older employees. More precisely, the diversity policy read "Our team of researchers promotes age diversity in teams. In view of the current demographic ageing, it is very important to encourage work with older employees. We therefore encourage people to work in teams with older colleagues." This diversity policy was thus worded slightly differently than that in the first study.

Pre-study. The goal of the pre-study was to determine the performance on a food quiz of teammates with certain world view, gender and age combinations, in order to determine an average performance for teammates with these characteristics. In turn, the average performance of teammates with certain world view, gender and age combinations was used to determine winning teams in the main study. To this end,

participants in the pre-study took a food quiz, exactly similar to that used in the main study. Specifically, participants successively saw fifteen different pictures of food items, which they had to name accurately in corresponding free form boxes. The foods pictured included, for instance, tiramisu, meringues, cinnamon sticks, beef wellington and salmon roe. After participants had been shown the fifteen food items, they were then asked to report their demographics, including their age, gender and which world view they most closely identified with. In addition, participants had to answer "strongly opposed" in an attention check question, on a 5-point scale from "strongly opposed" to "strongly in favour". Participants were 60 U.S. residents (63% men; mean age 39.02, SD 19.74). 57% were employed full-time, 13% were employed part-time and the remaining 30% were unemployed. One participant was eliminated from our pre-study as he failed to reply correctly to the attention check. We coded all participants' answers to the food quiz, attributing 1 point per correct answer and 0 otherwise. As there were fifteen different food items, the resulting performance score was an integer variable ranging from 0 to 15. We then categorized participants per age group, world view and gender, randomly selecting five participants per category and computing an average test score for all the categories. Average test scores varied from 5.4 to 8.2 depending on categories.

Participants. Participants in the main study were 156 U.S. residents aged under 30 (51% men; mean age 24.75, SD 3.03). 53% were employed full-time, 25% were employed part-time and the remaining 22% were unemployed.

Dependent measures. Participants had to choose two teammates from a list of eight. To create an age diverse team and comply with the age diversity policy, participants had to select two older candidates.

Process variables. After making their teammate choices, participants were asked to report how much they thought about some features of diversity when they made their teammate selection. In contrast with study 1, the features of diversity that people could choose from targeted specifically work in diverse teams. Precisely, participants were asked "Thinking back on the teammates you chose, how much did you think about the following points", on a scale from 0 to 100. The items on the list were "working with the

people I chose will increase the team's performance", "the people I chose will bring more talent to the team", "selecting the people I chose was the morally right thing to do", "the people I chose will not slow down the team's work", "the people I chose will make sure that the team's work satisfaction will not decrease" and "the people I chose will make the team's outcome less uncertain". In addition, participants were asked whether they thought "creating high performing age diverse teams is very desirable" and "creating high performing age diverse teams is easily feasible".

Control variables. Because participants' gender and experience might influence their teammate selection decisions, they were included as control variables in the regressions. Moreover, we added a control for employment status instead of supervisory duties. In fact, as controls for supervisory duties were not significant in the previous study, we chose to investigate whether employment status influenced more significantly diversity choices. Employment status was coded as 0, if the participant was unemployed, 1 otherwise. Gender was coded as a female dummy, equal to 0 if the participant was a man and 1 otherwise. Experience was coded on a 5-point Likert scale from 0 if the participant had no experience working in teams mixed in terms of age, up until 4 if the participant had a lot of experience working in teams mixed in terms of age. Lastly, we took into account the personality of participants by measuring their world views.

Additional exploratory outcome variables. After choosing their teammates, participants had to report how close they felt to the team members they picked, on a 7-Point Likert scale from "not close at all" to "very close". Also, participants were asked to rate how diverse they thought the team created was, in terms of age, gender and personality, all also on a 7-Point Likert scale from "not diverse at all" to "very diverse".

Pre-test. To measure world views with an item that allowed enough variability, we first ran a pre-test. In this pre-test, which was part of a larger study on attitudes towards job perceptions, the participants were asked to report which world view they most closely identified with, out of five world view pairs. Specifically, they were told: "Below you find five pairs of statements. For each pair, please choose the one that you find most applicable to you, personally." For the first pair, participants could

choose between "I believe that in uncertain times, one should always expect the best" or "I believe that in uncertain times, one should always plan for the worst". For the second pair, participants had to pick either "I believe that if you put out acceptance and warmth, you will attract the same, even in the business world" or "I believe that if you put out acceptance and warmth, you cannot expect the same in return, in particular in the business world". For the third pair, participants could choose "I am a believer in the idea that *every cloud has a silver lining*" or "I don't believe in the idea that *every cloud has a silver lining* because after all, we all need to face reality". For the fourth pair, participants could select "I firmly believe that in various areas of life, injustices are the exception rather than the rule" or "I firmly believe that in various areas of life, injustices are the rule rather than the exception". Finally the last pair offered participants the choice between "I believe that it's a dog-eat-dog world where you have to be ruthless at times, in particular in the business world" and "I believe that the world, including the business world, is a mostly cooperative and harmonious place in which people help and share with each other". The order of these pairs was randomized to participants. In addition, information about the gender, age and employment status of participants was collected. At the end of the survey, we included an attention check asking participants to select "slightly oppose" on a scale from "strongly oppose" to "strongly favor". Participants were 103 U.S. residents (55% men; mean age 36.24, SD 5.48). 74% were employed full-time, 5% were employed part-time, 9% were unemployed and the remaining 12% were self-employed. Eight participants were eliminated from our analyses as they failed to reply correctly to the attention check. We investigated the variance of answers for each of the pairs measuring world views. For the first pair, 67% of participants indicated "I believe that in uncertain times, one should always expect the best" and 33% "I believe that in uncertain times, one should always plan for the worst". For the second pair, 62% of participants picked "I believe that if you put out acceptance and warmth, you will attract the same, even in the business world" and 38% "I believe that if you put out acceptance and warmth, you cannot expect the same in return, in particular in the business world". For the third pair, 65% of participants chose "I am a believer in the idea

that *every cloud has a silver lining*" and 35% "I don't believe in the idea that *every cloud has a silver lining* because after all, we all need to face reality". For the fourth pair, 57% of participants selected "I firmly believe that in various areas of life, injustices are the exception rather than the rule" and 43% "I firmly believe that in various areas of life, injustices are the rule rather than the exception". For the last pair 59% of participants chose "I believe that it's a dog-eat-dog world where you have to be ruthless at times, in particular in the business world" and 41% "I believe that the world, including the business world, is a mostly cooperative and harmonious place in which people help and share with each other". Hence the pair exhibiting most variance was the fourth pair, followed by the fifth pair. Before selecting a pair for inclusion in our main study, we tested whether there was a significant difference in world views by gender. As there was a significant difference in world views by gender described in the fourth pair (Cohen's Kappa = -0.19, SE = 0.1, t = - 1.92, p = 0.05), this pair was eliminated. However, as choices between the world views in the fifth pair were not sensitive to gender (Cohen's Kappa = -0.14, SE = 0.10, t = - 1.44, p = 0.15), this fifth pair was selected for inclusion in our main study, since this pair also displayed the most variability. Hence the world view pair we chose to measure participants' world views in our main study included the two following items: "I believe that it's a dog-eat-dog world where you have to be ruthless at times, in particular in the business world" and "I believe that the world, including the business world, is a mostly cooperative and harmonious place in which people help and share with each other".

Attention checks. To ensure data quality in our main study, participants had to answer whether they slept less than two hours per night every day. This type of attention check is categorized as a logical statement (Abbey & Meloy, 2017). At the end of the survey, they were also asked whether in their honest opinion, we should use their data in our analyses for this study. They were reassured that they would be remunerated regardless of their answer to this question.

Manipulation checks. The effectiveness of the diversity policy was tested with a question asking whether the company particularly encouraged age diversity in teams. Moreover, participants had to reply to

a close-ended question about the characteristics of the teammates they had to choose when they created their team. They could again pick one answer out of three, where the correct answer was that all the teammates had different ages. In addition, participants were asked to identify who they created the team for, where the correct answer was for themselves. These manipulation checks were exactly the same as those used in the previous study.

Results

Attention checks. Three participants were eliminated from the study as they indicated wrongly that they slept less than two hours per night every day. Another four participants were excluded from the analysis as they indicated that they would not use their responses in our study.

Manipulation checks. Twenty five participants failed to recognize correctly the presence of a diversity policy in their survey. Hence they were excluded from further analyses. Another four participants did not identify that the teammates they could pick from all had different ages and were thus also excluded from the analyses. A further two people failed to correctly understand who they had to create the team for. The final sample consisted of 118 participants. Out of the 118 participants, 52 participants were in the no diversity policy condition and 66 in the diversity policy condition.[4]

Age diversity. Correlations and descriptive statistics are presented in Table 6. To test our hypotheses, we conducted linear regressions with robust standard errors. Team age diversity was measured by the number of older teammates selected to join the participant's team. We coded *older* as 0 for no selected older candidate, 1 for one selected older candidate and 2 for two selected older candidates. We coded *diversity policy* as 0 if the participant was in the no diversity policy condition, 1 if he-she was in the diversity policy condition. We coded *same* as 0 if the participant chose only teammates with a different world view to him-her, 1 if the participant chose one teammate with a similar world view to him-her and 2 if the participant chose two teammates with a similar world view to him-her. We first predicted team age

diversity with the diversity policy, including control variables (Table 3).

We find that pro-age diversity culture communication increases age diversity in teams by increasing the selection of older adults into work teams (M = 0.90, SD = 0.12, t = 7.65, p = 0.00), in line with Hypothesis 1. In fact, the diversity policy increased the number of older workers such that there was almost one more older worker per team. Moreover, there was no significant effect on the selection of older adults into teams of the participants' gender (M = 0.13, SD = 0.12, t = 1.11, p = 0.27), employment status (M = 0.07, SD = 0.15, t = 0.43, p = 0.67) and experience (M = 0.11, SD = 0.05, t = 2.14, p = 0.34). We next checked which process explained the effect of the diversity policy on team age diversity. We investigated whether there was a mediation effect of the likely process variables we identified on the relationship between the diversity policy and team age diversity, controlling for the employment status, gender and experience of participants. In a first step, we investigated whether the diversity policy had an effect on the features of diversity participants reported thinking about when making their teammate selection decisions (Table 4).

In line with Hypothesis 2, we find that pro-age diversity culture communication is positively related to the perceived desirability of work in age diverse teams (M = 19.22, SD = 5.23, t = 3.68, p = 0.00). Moreover, there was no significant effect on the perceived desirability of work in age diverse teams of the participants' gender (M = 8.02, SD = 5.29, t = 1.52, p = 0.13) and employment status (M = -9.22, SD = 5.72, t = -1.61, p = 0.11), but a significant positive effect of experience (M = 7.56, SD = 2.37, t = 3.19, p = 0.00). Additionally we find that there is a significant effect of the diversity policy on the perceived feasibility of high performing age diverse teams (M = 11.68, SD = 5.33, t = 2.19, p = 0.03). However, the diversity policy has no effect on an index made by the sum score of the advantages of diversity. These included information about whether the participant felt that working with the people chosen will increase the team's performance, whether the people chosen will bring more talent to the team, not slow down the team's work, make sure the team's work satisfaction will not decrease, make the team's outcome less

uncertain and whether selecting these people was the morally right thing to do. In addition, the diversity policy does not affect the perceived closeness of the participant to the members of the team. However the diversity policy has a significant positive effect on the perceived team age diversity (M = 1.93, SD = 0.35, t = 5.50, p = 0.00), as expected. To test further the behavioral impact of the desirability of diversity, we then investigate whether the perceived desirability of work in teams affects team age diversity (Table 5).

Supporting Hypothesis 3, this study shows that the perceived desirability of work in age diverse teams is positively related to the selection of older adults into work teams (M = 0.01, SD = 0.00, t = 5.36, p = 0.00). In addition, there was no significant effect on the selection of older adults into teams of the participants' gender (M = -0.03, SD = 0.13, t = -0.25, p = 0.80), employment status (M = -0.27, SD = 0.17, t = 1.57, p = 0.12) and experience (M = -0.01, SD = 0.05, t = -0.15, p = 0.88). Next we check whether there is a mediation effect of diversity's perceived desirability on the relationship between diversity policy and team age diversity, controlling for the employment status, gender and experience of participants. To do so, we use the same method as in the previous study, with the perceived desirability of high performing teams entered as a covariate, creating 500 bootstrap samples by randomly sampling observations with replacement from the original data set (Preacher & Hayes, 2004). Ninety-five percent confidence intervals show a significant indirect effect of diversity policy on diversity choices through the relative salience of the desirability of high performing age diverse teams (95% CI = [.02, .23]). Hence we find support for the fact that the effect of the diversity policy seems to be mediated by the desirability of high performing age diverse teams (Figure 2).

Yet one should again interpret these last results cautiously, as beliefs were not exogenous since they were not manipulated in our study. The Durbin-Wu-Haussman test confirms that beliefs are indeed endogenous, as the residuals in the augmented regression of the older teammate choice on the residuals from the first stage regression of the perceived desirability of diversity on diversity policy with controls for age, employment status and experience, are significantly different from 0 ($F(1,112) = 12.17$; $p = 0.00$).

Hence we follow the same approach as in the previous study, using our exogenous treatment variation as an instrument for the desirability of work in age diverse teams (Antonakis et al., 2010; Sajons, 2020). Yet again the first-stage F-statistic shows the strength of our instrument does not pass the required critical value of 10 (Stock & Yogo, 2005), F(4,113) = 6.52. So we need to focus our interpretation on the reduced form estimates. At this point, we investigated whether the age diversity policy could have adverse consequences, for instance by driving participants to select increasingly teammates similar to them in terms of world views, or gender (Table 3).

This study shows that pro-age diversity culture communication does not decrease team diversity in terms of worldviews (M = -0.09, SD = 0.12, t = -0.77, p = 0.44). In addition, there was no significant effect on the selection of personally similar adults into teams of the participants' employment status (M = -0.02, SD = 0.13, t = -0.12, p = 0.90) and experience (M = 0.03, SD = 0.05, t = 0.65, p = 0.52), but a significant effect of gender (M = -0.25, SD = 0.12, t = -2.11, p = 0.04) meaning women were less likely to select personally similar adults into their teams. However, in line with the similarity attraction principle, we find evidence that participants prefer to select teammates similar in world views to them, independent of the policy. In fact, only 8% of participants chose two teammates different in world views to themselves for their team. In addition, the diversity policy has no significant impact on the team's gender diversity (M = -0.07, SD = 0.10, t = -0.73, p = 0.47). Thus we can conclude that it is unlikely that the diversity policy has unintended side effects by increasing personality or gender homogeneity.

Discussion. This study provides additional support for the fact that pro-age diversity culture communication increases age diversity in teams by increasing the selection of older adults into work teams, supporting Hypothesis 1. In line with study 1, this study also provides evidence for the fact that the diversity policy has a positive effect on team age diversity, mainly thanks to its impact on the perceived desirability of diversity. This is in line with Hypotheses 2 and 3, confirming that the perceived desirability of diversity is an important process underlying the behavioral impact of pro-diversity culture

communication. Moreover, we replicate the results of study 1 showing that the diversity policy has no unintended side effects, insofar as it does not promote additional team homogeneity in terms of worldviews or gender. We next investigate further the robustness of these results in an experimental setting with real incentives, just as in this study, but this time varying the perceived personality similarity of teammates available for selection, as opposed to their actual world views. We also reduce the choice set of teammates available to work on the project, which could be argued is more ecologically valid.

Study 3

Method

Design and procedure. The experiment had 2 conditions (company diversity policy: yes or no), with diversity policy as a between-subjects factor. As in study 2, real incentives were used and participants knew that their teammate selection decisions would have a financial impact on them. In fact, participants' payment depended on their performance on a food quiz, as well as on that of the teammates they selected. To determine the performance of the teammates chosen, we used the same pre-study as the one conducted before study 2. In this pre-study which took place on MTurk, participants with the same gender, age and world view combinations as the teammates available for selection took the food quiz. From this data, an average performance was computed and used to determine payment for participants in study 3 alongside their own performance on the quiz. In the main study, participants had to choose two teammates out of a list of four, before taking the food quiz themselves. The four teammates available for selection only varied across two dimensions and as in the previous study, there was no one already on the team, so it was possible to create an entirely age homogeneous team. More precisely, the teammates varied in terms of age (younger from 24 to 26 years old or older from 52 to 56 years old) and personalities, insofar as they were similar or not to the participant. The age ranges of the teammates available for selection were comparable to those used in our previous studies. So that participants would not get cues about the teammates' genders

and in contrast with the previous studies, the teammates were named by their initials only. For instance, candidate R (56 years old) and candidate D (24 years old) were described as pretty similar to the participant in terms of their personalities, based on information collected before the study. The same four teammate profiles were shown across experimental conditions.

In addition, participants were randomly assigned to the diversity or to the no diversity policy condition. If they were assigned to the diversity policy condition, the diversity policy appeared on a separate screen as in previous studies. Like in study 2, the diversity policy specified that in view of the current demographic aging, it was very important to encourage work with older colleagues. Participants in the no diversity policy condition were not shown any specific information about the company's diversity guidelines.

Participants. Participants were 164 U.S. residents aged under 30 (46% men; mean age 24.98, SD 2.69). 52% were employed full-time, 28% were employed part-time and the remaining 20% were unemployed.

Dependent measures. Participants had to choose two teammates from a list of four. To create an age diverse team and comply with the age diversity policy, participants needed to select two older candidates.

Process variables. As in previous studies, participants were asked to report how much they thought about some common features of diversity when they made their teammate selection. Precisely, participants were asked "Thinking back on the teammates you chose, how much did you think about the following points". The items on the list that participants could rate from 0 to 100 were "working with the people I chose will increase my team's performance", "the people I chose will bring more talent to my team", "selecting the people I chose was the morally right thing to do", "the people I chose will not slow down my team's work", "the people I chose will make sure that my team's work satisfaction will not decrease" and "the people I chose will make my team's outcome less uncertain". In addition, participants were asked

whether they thought "creating high performing age diverse teams is very desirable" and "creating high performing age diverse teams is easily feasible".

Control variables. Because participants' gender and experience might influence their teammate selection decisions, they were included as control variables in the regressions. Moreover, we added a control for employment status. Employment status was coded as 0, if the participant was unemployed, 1 otherwise. Gender was coded as a female dummy, equal to 0 if the participant was a man and 1 otherwise. Experience was coded on a 5-point Likert scale from 0 if the participant had no experience working in teams mixed in terms of age, up until 4 if the participant had a lot of experience working in teams mixed in terms of age. Lastly, we took into account the personality of participants by measuring their world views.

Additional exploratory outcome variables. After choosing their teammates, participants had to report how close they felt to the team members they chose, on a 7-Point Likert scale from "not close at all" to "very close". Also, participants were asked to rate how diverse they thought the team created was, in terms of age and personality, on a 7-Point Likert scale from "not diverse at all" to "very diverse".

Attention checks. To ensure data quality, participants had to choose "hardly ever" from a list of answers including "always", "often", "hardly ever" and "never". At the end of the survey, they also had to say whether in their honest opinion, we should use their data in our analyses for this study. They were reassured that they would be remunerated regardless of their answer to this question.

Manipulation checks. The effectiveness of the company diversity policy was tested with a question asking whether the company particularly encouraged age diversity in teams, as in other studies. Moreover, participants had to answer a close-ended question about the characteristics of the teammates they had to choose when they created their team. They could again select one answer out of three, where the correct answer was that all the teammates had different ages.

Results

Attention checks. Two participants were eliminated from the study as they failed to indicate correctly "hardly ever" in the attention check. Another three participants were excluded from the analyses as they specified that they would not use their responses in our study.

Manipulation checks. Seventeen participants did not recognize correctly the presence of the diversity policy, hence they were excluded from further analyses. Another six participants did not understand that the teammates they could select all had different ages and were thus also excluded from the analyses. Three participants had duplicate MTurk codes and were successively eliminated. The final sample consisted of 133 participants, with 60 participants in the no diversity policy condition and 73 in the diversity policy condition.[5]

Age diversity. Correlations and descriptive statistics are presented in Table 7. To test our hypotheses, we conducted linear regressions with robust standard errors. Team age diversity was measured by the number of older teammates selected to join the participant's team. We coded *older* as 0 for no selected older candidate, 1 for one selected older candidate and 2 for two selected older candidates. We coded *diversity policy* as 0 if the participant was in the no diversity policy condition, 1 if he-she was in the diversity policy condition. We coded *same* as 0 if the participant chose only teammates with different personalities to him-her, 1 if the participant chose a teammate with a similar personality to him-her and 2 if the participant chose two teammates with a similar personality to him-her. We first predicted team age diversity with the company diversity policy, including control variables (Table 3).

Consistent with Hypothesis 1, this study finds that pro-age diversity culture communication increases age diversity in teams by increasing the selection of older adults into work teams ($M = 0.43$, $SD = 0.09$, $t = 4.85$, $p = 0.00$). In fact, the diversity policy increased the number of older workers such that there was almost one more older worker per two teams. Moreover, there was no significant effect on the

selection of older adults into teams of the participants' employment status (M = 0.15, SD = 0.10, t = 1.53, p = 0.13) and experience (M = -0.05, SD = 0.05, t = -1.05, p = 0.30) but a significant effect of gender (M = 0.16, SD = 0.08, t = 2.00, p = 0.05). Next, we checked whether the diversity policy had a positive effect on the perceived desirability of work in diverse teams (Table 4).

Consistent with previous studies, we find that pro-age diversity culture communication is positively related to the perceived desirability of work in age diverse teams (M = 5.44, SD = 4.23, t = 1.29, p = 0.20), but no longer significantly so, unlike in previous studies. Moreover, there was no significant effect on the perceived desirability of work in age diverse teams of the participants' gender (M = 4.87, SD = 3.97, t = 1.23, p = 0.22), employment status (M = 5.03, SD = 6.25, t = 0.80, p = 0.42) and experience (M = 1.69, SD = 2.43, t = 0.69, p = 0.49). Hence overall this study provides evidence for Hypothesis 2 only to a certain extent. Additionally the policy does not have a significant effect on the perceived feasibility of high performing age diverse teams. In the same way, it has no effect on an index made by the sum score of the advantages of diversity. These advantages of diversity included whether the participant felt that working with the people chosen will increase his-her team's performance, whether the people chosen will bring more talent to his-her team, not slow down his-her team's work, make sure his-her team's work satisfaction will not decrease, make his-her team's outcome less uncertain and whether selecting these people was the morally right thing to do. In addition, the diversity policy does not affect the perceived closeness of the participant to the members of the team, nor the perceived personality diversity of the team created. In line with previous studies, the policy has a small positive effect on the perceived age diversity of the team created (M = 0.54, SD = 0.31, t = 1.78, p = 0.08), albeit not significant. There is also no significant effect of the policy on the perceived personality diversity of the team created. In a next step, we investigate whether the perceived desirability of diversity affects significantly behavior, checking whether the perceived desirability of diversity increases the number of older teammates chosen into the team (Table 5).

In this study, we find that the perceived desirability of work in age diverse teams is positively related

to selecting older adults into work teams, although not significantly so (M = 0.00, SD = 0.00, t = 1.36, p = 0.18). Moreover, there was no significant effect on the selection of older adults into teams of the participants' gender (M = 0.15, SD = 0.09, t = 1.68, p = 0.09), employment status (M = 0.11, SD = 0.10, t = 1.02, p = 0.31) and experience (M = -0.03, SD = 0.05, t = -0.51, p = 0.61). Hence, overall we find limited support for Hypothesis 3 only and we do not investigate mediation models further. We next assess whether the age diversity policy yields unintended side effects, checking its impact on perceived team personality similarity (Table 3).

This study shows that pro-age diversity culture communication does not decrease team diversity in terms of personal characteristics and values (M = -0.12, SD = 0.11, t = -1.09, p = 0.28). Moreover, there was no significant effect on the selection of personally similar adults into teams of the participants' gender (M = 0.04, SD = 0.11, t = 0.31, p = 0.75), employment status (M = 0.13, SD = 0.15, t = 0.89, p = 0.38) and experience (M = 0.03, SD = 0.05, t = 0.62, p = 0.54). Yet we find evidence that participants prefer to select teammates similar in personalities to them, in line with the similarity attraction principle. In fact, just 11% of participants chose two teammates different in personalities to them for their team.

Discussion. Study 3 confirms the findings of previous studies, as we find that pro-age diversity culture communication increases age diversity in teams by increasing the selection of older adults into work teams. This result has been robust across all our studies so far. In addition, we do not find evidence that there is a backlash effect of the age diversity policy, to the extent that it does not increase significantly the personality homogeneity of the team created. This study also finds partial evidence for the fact that the perceived desirability of diversity might explain the behavioral impact of the diversity policy, in line with the first two studies and with our second and third hypotheses. The next study will seek to test further the robustness of these results, in a setting where teammate incentives are also considered. In fact, it could be argued that the three studies presented so far could be made even more ecologically valid, insofar as the teammates selected for inclusion on the team are not rewarded financially. In the business world, including

an employee in an important project is likely to have positive consequences on his-her career, while excluding him-her is likely to make him-her relinquish these positive consequences and thereby punish him-her. To take this into account, the next study will use the same basic setting as the previous studies, but including experimental conditions in which the teammates selected onto the teams are rewarded financially for their performance. Moreover, the next study will start with a basic team that is already diverse, as it includes an older person. This way it will provide information on whether the diversity policy is also effective when the initial team diversity is higher.

Study 4

Method

Design and procedure. The experiment had a 2 (diversity policy: yes or no) x 2 (teammate incentives: yes or no) between-subjects design, with company diversity policy and teammate incentives as between-subjects factors. The basic setting of this study was similar to that of studies 2 and 3, except with regards to teammate incentives. In fact, as in studies 2 and 3, real incentives were used and participants knew that their teammate selection decisions would have a financial impact on them. In addition, participants' payment depended on their performance on a food quiz, as well as on that of the teammates they selected. Again, teammate performance was determined by the pre-study. However, in contrast to previous studies, participants were randomly allocated to one of four conditions: in two of these four conditions, the four teammates on the winning team with the highest total performance on the food quiz were each paid the study bonus (team incentives). We argued this could approximate a situation which is more ecologically valid, as the teammates who are not selected into teams are effectively punished in financial terms by not having the possibility to win the bonus. In the other two conditions, only the participant who created the winning team got the study bonus (no team incentives), as in previous studies. Furthermore, half of participants were assigned to the company diversity policy condition, and the other

half to the no diversity policy condition. If participants were assigned to the diversity policy condition, they saw the diversity policy appearing on a separate screen as in previous studies. Consistent with study 1, the diversity policy stated that our team of researchers promoted age diversity in teams and therefore encouraged people to work in teams that are mixed in terms of age.

In contrast to previous studies, the participant was already on the team with two other pre-selected teammates described as teammate A (55 years old) and teammate B (25 years old) and should choose only one additional teammate. This new design allowed for a higher initial team diversity, while ensuring that there was no interdependency in the participants' choices, since only one choice was made. This is in contrast to our previous studies in which one teammate being selected meant he-she could not be selected again. In this study, there were four candidates available for selection, who all varied in terms of age (younger from 24 to 26 years old or older from 52 to 56 years old) and personality similarity to the participant. The age ranges of the teammates available for selection were comparable to those used in our previous studies. So that participants would not get cues about the teammates' genders, the teammates were named by their initials only, as in study 3. For instance, candidate R (56 years old) and candidate D (24 years old) were described as pretty similar to the participant in terms of their personalities and how they viewed the world, based on information collected before the study. The same four teammate profiles were shown across experimental conditions.

Participants. Participants were 237 U.S. residents aged under 30 (56% men; mean age 26.67, SD 2.77). 62% were employed full-time, 20% were employed part-time and the remaining 18% were unemployed.

Dependent measures. Participants had to select one teammate from a list of four. The diversity policy encouraged work in age mixed teams, which should induce participants to select the older candidate.

Process variables. Participants had to report how much they thought about some common features of diversity when they made their teammate selection, as in previous studies. In this study, participants

were asked "Thinking back on the teammate you chose, how much did you think about the following points". The items on the list that participants could rate from 0 to 100 were "working with the teammate I chose will increase my team's performance", "the teammate I chose will bring more talent to my team", "selecting the teammate I chose was the morally right thing to do", "the teammate I chose will not slow down my team's work", "the teammate I chose will make sure that my team's work satisfaction will not decrease" and "the teammate I chose will make my team's outcome less uncertain". Moreover, participants had to indicate whether they thought "creating high performing age diverse teams is very desirable" and "creating high performing age diverse teams is easily feasible".

Control variables. Because participants' employment status, gender and experience might influence their teammate selection decisions, these were included as control variables in the regressions. Employment status was coded as 0, if the participant was unemployed, 1 otherwise. Gender was coded as a female dummy, equal to 0 if the participant was a man and 1 otherwise. Experience was coded on a 5-point Likert scale from 0 if the participant had no experience working in teams mixed in terms of age, up until 4 if the participant had a lot of experience working in teams mixed in terms of age. Lastly, we took into account the personality of participants by measuring their world views.

Additional exploratory outcome variables. After choosing their teammates, participants had to report how close they felt to the team members they chose, on a 7-Point Likert scale from "not close at all" to "very close". Also, participants were asked to rate how diverse they thought the team created was, in terms of age and personality, both on a 7-Point Likert scale from "not diverse at all" to "very diverse".

Attention checks. To ensure data quality, participants had to indicate whether they slept less than two hours per night every day. At the end of the survey, they were also asked whether in their honest opinion, we should use their data in our analyses for this study. They were reassured that they would be paid no matter their answer to this question.

Manipulation checks. The effectiveness of the diversity policy was tested with a question on whether we particularly encouraged age diversity in teams, as in other studies. Moreover, participants were asked about the characteristics of the teammates they had to choose when they created the team. They could select one answer out of three, where the correct answer was that all the teammates available for selection had different ages.

Results

Attention checks. Two participants were eliminated from the study as they failed to answer correctly the attention check. Another four participants were excluded from the analyses as they indicated that they would not use their responses in our study.

Manipulation checks. Forty six participants failed to recognize correctly the presence of a diversity policy in their questionnaire. Hence they were excluded from further analyses. Another participant did not understand that the teammates available for selection all had different ages and was thus not considered either in our analyses. The final sample consisted of 184 participants, including 46 participants in the condition with team incentives and the diversity policy, 51 in the condition with no team incentives and the diversity policy, 40 participants in the condition with team incentives and no diversity policy and 47 in the condition with no team incentives and no diversity policy.[6]

Age diversity. Correlations and descriptive statistics are presented in Table 8. To test our hypotheses, we conducted linear regressions with robust standard errors. Team age diversity was measured by the number of older teammates selected to join the participant's team. We coded *older* as 0 for no selected older candidate and 1 for one selected older candidate. We coded *diversity policy* as 0 if the participant was in the no diversity policy condition, 1 if he-she was in the diversity policy condition. We coded *teamincentives* as 0 if only the participant on the winning team was paid the study bonus, 1 if all the teammates on the winning team were paid the study bonus. We coded *same* as 0 if the participant did not

choose a teammate with a similar personality to him-her and 1 if the participant chose a teammate with a similar personality to him-her. We first predicted team age diversity with the diversity policy, including control variables (Table 3). We then predicted team age diversity with team incentives, including control variables. We next considered a model with both the diversity policy and team incentives, including control variables.

In this study, we find that pro-age diversity culture communication increases age diversity in teams by increasing the selection of older adults into work teams, but not significantly (M = 0.10, SD = 0.07, t = 1.38, p = 0.17). In qualitative terms, this effect can be interpreted as the selection of one more older worker per ten teams. We suspect this result to be driven by the fact that the initial team diversity is higher than in previous studies, which could lessen the perceived need for further inclusion of diverse candidates. Moreover, there was no significant effect on the selection of older adults into teams of the participants' gender (M = -0.02, SD = 0.07, t = -0.35, p = 0.73) and employment status (M = -0.10, SD = 0.10, t = -1.00, p = 0.32), but a significant positive effect of experience (M = 0.10, SD = 0.04, t = 2.85, p = 0.01). In addition, there was no significant impact of team incentives on the selection of older workers, nor on the impact of the diversity policy in increasing older teammate selection. This is reassuring for us as it provides evidence for the robustness of our results across different incentive schemes. Next, we checked whether the diversity policy had an effect on the features of diversity that participants reported thinking about when making their teammate selection decisions (Table 4).

In line with Hypothesis 2, we find that pro-age diversity culture communication is positively related to the perceived desirability of work in age diverse teams (M = 17.72, SD = 3.89, t = 4.55, p = 0.00). In addition, there was no significant effect on the perceived desirability of work in age diverse teams of the participants' gender (M = 0.81, SD = 3.70, t = 0.22, p = 0.83), employment status (M = 3.75, SD = 5.97, t = 0.63, p = 0.53) and experience (M = 1.12, SD = 1.99, t = 0.56, p = 0.57). Moreover the policy has a significant positive effect on the perceived feasibility of high performing age diverse teams. In the same

way, it has a positive effect on an index made by the sum score of the advantages of diversity. This index included information about whether the participant felt that working with the teammate chosen will increase his-her team's performance, whether the teammate chosen will bring more talent to his-her team, not slow down his-her team's work, make sure his-her team's work satisfaction will not decrease, make his-her team's outcome less uncertain and whether selecting this teammate was the morally right thing to do. In addition, the diversity policy affects positively the perceived closeness of the participant to the members of the team, and the perceived personality diversity of the team created. Moreover, in line with previous studies, the policy has a positive effect on the perceived age diversity of the team created (M = 0.47, SD = 0.23, t = 2.04, p = 0.04).

To test whether the perceived desirability of diversity has a significant effect on behavior, we then investigate whether the perceived desirability of diversity increases the number of older teammates chosen into the team (Table 5). In support of Hypothesis 3, we find that the perceived desirability of work in age diverse teams is positively related to selecting older adults into work teams (M = 0.01, SD = 0.00, t = 4.09, p = 0.00). Moreover, there was no significant effect on the selection of older adults into teams of the participants' gender (M = -0.03, SD = 0.07, t = -0.41, p = 0.69) and employment status (M = -0.12, SD = 0.09, t = -1.25, p = 0.21), but a significant positive effect of experience (M = 0.10, SD = 0.03, t = 2.85, p = 0.01). We then seek further whether there is a mediation effect of diversity's perceived desirability on the relationship between company diversity policy and team age diversity, controlling for employment status, gender and experience of participants. Again we used the Stata SEM function (Hayes, 2012) for mediated moderation bootstrapping analyses with the relative desirability of high performing teams entered as a covariate, creating 500 bootstrap samples by randomly sampling observations with replacement from the original data set (Preacher & Hayes, 2004). Ninety-five percent confidence intervals show a significant indirect effect of company diversity policy on diversity choices through the relative salience of the desirability of high performing age diverse teams (95% CI = [.03, .15]). Hence we find support for the fact

that the effect of the diversity policy seems to be mediated by the desirability of high performing age diverse teams (Figure 3).

However it is likely that beliefs were not exogenous, since we did not manipulate them in our study. This is confirmed by the Durbin-Wu-Haussman test which suggests that beliefs are indeed endogenous, as the residuals in the augmented regression are significantly different from 0 ($F(1,178) = 13.78$; $p = 0.00$). Hence we follow the technique suggested by previous research, using our exogenous treatment variation as an instrument for the desirability of work in age diverse teams (Antonakis et al., 2010; Sajons, 2020). Precisely, we conduct a two-stage least squares estimation in which the first stage regresses the desirability of work in age diverse teams on the variable indicating the diversity policy, and controlling for participants' gender, employment status and experience. The second stage uses the predicted values from the first stage as an exogenous regressor to estimate the effects of the diversity policy on age diversity. However the first-stage F-statistic indicates that the strength of our instrument does not pass the required critical value of 10 (Stock & Yogo, 2005), $F(4,179) = 5.40$. Hence we should focus our interpretation on the reduced form estimates.

Lastly we examine whether the diversity policy fosters unintended consequences by resulting in increased team personality homogeneity (Table 3). As in previous studies, we report that pro-age diversity culture communication does not decrease team diversity in terms of personal characteristics and values ($M = 0.02$, $SD = 0.07$, $t = 0.32$, $p = 0.75$). Moreover, there was no significant effect on the selection of personally similar adults into teams of the participants' gender ($M = -0.13$, $SD = 0.07$, $t = -1.98$, $p = 0.06$), employment status ($M = -0.11$, $SD = 0.10$, $t = -1.15$, $p = 0.25$) and experience ($M = 0.04$, $SD = 0.04$, $t = 1.15$, $p = 0.25$). This time in contrast to previous studies, we find no evidence for the similarity attraction principle, as 51% of participants chose a teammate different in personality to themselves for their team. We suspect that these results are again driven by the fact that initial team diversity is already high in this study, which surely lessens the perceived threat posed by diverse others and hence the need to seek similar others

to counteract this effect. We will investigate further whether initial team diversity is indeed already perceived as high in the next study. Moreover, we find no effect of team incentives on team personality similarity, nor on the effect of the policy on team personality. This provides added evidence for the robustness of our results across incentive schemes.

Discussion. This study identifies a boundary condition for the effectiveness of diversity policies. In fact, pro-age diversity culture communication only has a limited impact in increasing age diversity in teams, certainly due to the fact that the original team was perceived as more diverse than in previous studies. Hence we have identified a boundary condition for the positive effect of the diversity policy, as it has less impact when the original team diversity is higher. Moreover, the incentivization of teammates does not have an effect on the behavioral impact of the policy. This provides evidence for the robustness of our results across incentive schemes. In line with previous studies, we also find evidence that pro-age diversity culture communication is positively related to the perceived desirability of work in age diverse teams. In turn, the perceived desirability of work in age diverse teams is positively related to selecting older adults into work teams. Hence there is evidence that the diversity policy's impact on behavior can be at least partly explained by its positive impact on the perceived desirability of work in diverse teams. Lastly, we find added evidence that pro-age diversity communication does not backlash, as it does not increase team personality similarity. These results hold no matter whether teammates are incentivized or not.

Study 5

Method

Study 4 identified a boundary condition for our previous results, since the diversity policy only had a limited positive impact in increasing team age diversity. To investigate further this effect, we conducted a follow-up study seeking to understand how people perceived the diversity policy as well as the initial team diversity, across all our previous studies.[7] The follow-up study, conducted online on Mechanical Turk, had

51 participants aged under 30 (49% men, mean age 25.55, SD 2.71). 57% were employed full-time, 26% were employed part-time and the remaining 17% were unemployed. Participants successively saw the three diversity policies used in our previous studies, as the policies used in studies 2 and 3 were the same, in random order, and were then asked to rate how likely they would be to act on them, how explicit they found them, how persuasive they found them and how direct they found them, on a 7-Point Likert scale from "not at all" to "very much". In addition, they were shown the initial team descriptions of studies 1 and 4 and had to indicate how diverse they found these teams, in terms of age, personality and gender, on a 7-Point Likert scale from "not diverse at all" to "very diverse". We hypothesize that participants might have perceived the initial team in study 4 as more diverse, explaining the lower impact that the pro-age diversity policy had on behavior in this study.

Results

Table 9 indicates that participants indeed perceived the initial team in study 4 as significantly more diverse in terms of age ($M_{s4} = 5.69$; $M_{s1} = 2.63$; $SD_{s4} = 1.21$; $SD_{s1} = 1.72$; $p = 0.00$) and personality ($M_{s4} = 4.56$; $M_{s1} = 3.56$; $SD_{s4} = 1.03$; $SD_{s1} = 1.2$; $p = 0.00$) than the team from study 1. There was no significant difference in the perceived gender diversity of the teams in studies 1 and 4. We next investigated how the diversity policies were perceived across our studies. Results show that there was no significant difference in how likely participants said they would be to act on the different diversity policies shown in our studies, nor in how persuasive they thought these policies were. However, participants thought the policy from studies 2 and 3 was marginally more explicit than that from study 4 ($M_{s2,3} = 4.77$; $M_{s4} = 4.35$; $SD_{s2,3} = 1.81$; $SD_{s4} = 1.58$; $p = 0.05$) and significantly more explicit than that from study 1 ($M_{s2,3} = 4.77$; $M_{s1} = 4.21$; $SD_{s2,3} = 1.81$; $SD_{s1} = 1.82$; $p = 0.01$). There was no significant difference in how explicit participants perceived the policy from studies 1 and 4 to be. Moreover, participants thought the policy from studies 2 and 3 was significantly more direct than that from study 1 ($M_{s2,3} = 5.46$; $M_{s1} = 4.98$; $SD_{s2,3} =$

1.61 ; SD_{s1} = 1.44 ; p = 0.03) and there was no significant difference in the perceived directness of the policies across the other studies. All in all, these results are in line with our hypothesis that the team in study 4 was perceived as significantly more diverse than that in study 1, providing a likely explanation for the reduced behavioral effect of the policy in this study.

General Discussion

We chart new theoretical and empirical alleys within the diversity literature by investigating the impact of organizational pro-age diversity communication on team age diversity. Building on theories of diversity and symbolic management, we suggested that pro-age diversity culture communication increases the representation of older workers in teams. Our results confirm that pro-age diversity culture communication increases age diversity in teams by increasing the selection of older adults into work teams. Additionally, we identified a boundary condition for the effectiveness of pro-diversity culture communication, showing that its impact is reduced if initial team diversity is perceived to be high. Moreover, we hypothesized that pro-diversity culture communication highlights the desirability of work in diverse teams, and that in turn the perceived desirability of work in diverse teams increases team diversity. We hypothesized that this mechanism is particularly important, as organizational values reflected in pro-diversity culture communication highlight desirable behaviors to be adopted by employees. Our results are in line with these hypotheses, as we show that pro-age diversity culture communication is positively related to the perceived desirability of work in age diverse teams and that the perceived desirability of work in age diverse teams is positively related to selecting older adults into work teams. Finally, we investigated to what extent pro-diversity culture communication may also produce unintended consequences, such as reducing diversity in terms of personality and values due to similarity attraction dynamics. Across our studies, we find no evidence that pro-age diversity culture communication decreases team diversity in terms of personal characteristics and values, showing that increasing diversity in a certain dimension does not

necessarily decrease it in another important dimension.

Overall, several practical implications can be drawn from our research. Firstly, our paper promotes a better understanding of diversity initiatives and of their impact on a yet under-studied dimension of diversity. It provides an actionable recommendation to practitioners, showing that pro-diversity culture communication can be an effective tool to address inequalities. With the current increase in remote working due to the Covid crisis, companies are increasingly likely to disseminate information about their pro-diversity culture communication online in the form of a diversity policy, in a similar way to what was done in this research. It is thus very relevant to show that such simple diversity initiatives, which consist in typing a few words on an email or company website, can have a tangible impact on the diversity of work teams. This is even more important as some businesses have pushed back on the implementation of diversity initiatives, arguing that their impact is too uncertain and that their associated costs are too high. This study shows exactly the opposite and thus constitutes a powerful message to deliver to practitioners aiming to increase their workforce diversity. Secondly, this paper investigates how diversity policies operate, allowing to draw broader conclusions regarding other types of diversity initiatives. In fact, this study uncovers evidence that diversity policies impact behavior as they make more salient the desirability of diversity. Hence there is evidence that crafting other diversity initiatives which also make the desirability of diversity more salient, will have the same sort of positive impact on diversity. Thirdly, this research highlights that diversity policies will not necessarily result in unintended side effects, insofar as they do not lead to decreased personality diversity. Therefore this work provides the reassuring conclusion that efforts to increase a dimension of diversity will not necessarily result in a decrease in another important dimension of diversity. Overall this provides further comfort to practitioners looking to implement effective diversity initiatives in their workplace.

This research has limitations that could be addressed by future studies. We investigated the impact of diversity policies in an online setting, in which participants arguably focus more than in other

circumstances. Hence future work could consider our study as a starting point for a field study, which could consist in an intervention inside a company or an organization. Using a field design with a control group that is not exposed to any diversity initiative and a treatment group that is exposed to a diversity policy before both taking part in a team creation task would be a great way to test the robustness of our results in a different setting. In addition, our studies measured personality using rather succinct measures (world views). Future studies could measure real personality in more extensive ways, using for instance the Big Five, which give a more complete picture of one's personality (John et al., 1991). Also, our research uncovers that our diversity policy impacts diversity choices as it makes the desirability of diversity more salient. It would be interesting to test whether the same mechanism operates for other diversity initiatives. Future research could investigate for instance whether diversity trainings focused at stereotypes and biases are effective in increasing age diversity in teams, and if so whether this is also caused by an increased salience of diversity's desirability. Moreover, it would be important to uncover how long lasting the effects of these diversity initiatives are, in order to formulate recommendations regarding how often they need to be implemented. Lastly, we focused our analysis on young participants, since this is the age group for which ageism is most pronounced (Rupp et al., 2005). Future studies could consider including older participants too, especially in light of the fact that older people might be more often in a position to choose who they wish to work with.

All in all, our analysis reveals that pro-diversity culture communication can be an effective tool to promote diversity. It would now be interesting to build on this observation further and inquire more in depth into the mechanisms driving this effect. For instance, future work could investigate whether the mere fact that an authority figure communicated the pro-diversity culture communication affected compliance, in line with previous results (Schwarzwald & Goldenberg, 1979; Bushman, 1984). In fact, it has been revealed that business justifications by legitimate authority figures influenced discrimination at hiring (Brief et al., 2000), which is closely related to our setting. Across all our studies, the diversity policies were

in fact communicated either by the company for which the participant had to imagine working for, or by the team of researchers. In all these cases, it was thus apparent that the diversity policy was communicated by an instance with a higher role authority, which could further influence compliance. Relatedly, our results reveal that the diversity policy likely increased age diversity in teams by fostering the relative perceived desirability of work in age mixed teams, but not through other changes in the attitudes or perceptions of participants. Thus the effect of the diversity policy might simply amount to a pure compliance effect, perhaps caused by the fact that the diversity policy was communicated by an authority instance. To test this hypothesis, further studies could consider whether the diversity policy is really driving a simple compliance effect. To do so, it would be interesting to analyze whether the compliance effect occurs if an authority figure advocates work diversity on another trivial dimension, say, work with brown-eyed people. In addition, further mechanisms could influence our results. In fact, the attitudes of participants towards older people might have affected their propensity to select older people into their teams. In a closely related setting, prior work has shown that negative attitudes towards older people were negatively related to hiring older people and to the likelihood of selecting the oldest candidate (Fasbender & Wang, 2017). Applied to our work, it is thus a possibility that negative attitudes towards older people influenced negatively the selection of older teammates into teams. As a result, future work could consider adding the attitudes of participants towards older people as another explanatory variable in the analysis, which could shed further light on the mechanisms influencing our results.

In conclusion, by investigating the impact of pro-diversity culture communication on age diversity, this research sheds light on a simple initiative which can be successful at increasing team age diversity. Overall, it provides evidence that short diversity policies have a positive impact on the creation of age diverse teams, without resulting in unintended consequences. This research also highlights that pro-diversity culture communication has a positive behavioral impact because it makes the desirability of diversity more salient. In turn this insight can be useful to apply when devising other types of diversity

initiatives. Future research should further investigate these effects in field studies to measure more accurately the long-term behavioral impact of pro-diversity culture communication, while also considering that of other initiatives aimed at increasing diversity in the workplace.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53*, 63–70.

Alas, R., & Vadi, M. (2006). The impact of organisational culture on organisational learning and attitudes concerning change from an institutional perspective. *International Journal of Strategic Change Management*, *1*(1-2), 155–170.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120.

Avery, D. R. (2003). Reactions to diversity in recruitment advertising–are differences black and white? *Journal of Applied Psychology*, *88*(4), 672.

Avery, D. R., McKay, P. F., Wilson, D. C., & Tonidandel, S. (2007). Unequal attendance: The relationships between race, organizational diversity cues, and absenteeism. *Personnel Psychology*, *60*(4), 875–902.

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, *83*(3), 377.

Barry, B., & Stewart, G. L. (1997). Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied Psychology*, *82*(1), 62.

Bartels, L. K., Nadler, J. T., Kufahl, K., & Pyatt, J. (2013). Fifty years after the civil rights act: Diversity-management practices in the field. *Industrial and Organizational Psychology*, *6*(4), 450–457.

Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., & Briggs, A. L. (2011). Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of Management*, *37*(3), 709–743.

Berrey, E. (2021). *The enigma of diversity*. University of Chicago Press.

Bezrukova, K., Jehn, K. A., & Spell, C. S. (2012). Reviewing diversity training: Where we have been and where we should go. *Academy of Management Learning & Education*, *11*(2), 207–227.

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin*, *142*(11), 1227.

Boye, M. W., & Jones, J. W. (1997). Organizational culture and employee counterproductivity. *Antisocial Behavior in Organizations*, 172–184.

Brief, A. P., Dietz, J., Cohen, R. R., Pugh, S. D., & Vaslow, J. B. (2000). Just doing business: Modern racism and obedience to authority as explanations for employment discrimination. *Organizational behavior and human decision processes*, *81*(1), 72–97.

Bushman, B. J. (1984). Perceived symbols of authority and their influence on compliance 1. *Journal of Applied Social Psychology*, *14*(6), 501–508.

Carter, E. R., Peery, D., Richeson, J. A., & Murphy, M. C. (2015). Does cognitive depletion shape bias detection for minority group members? *Social Cognition*, *33*(3), 241–254.

Chang, E. H., Milkman, K. L., Gromet, D. M., Rebele, R. W., Massey, C., Duckworth, A. L., & Grant, A. M. (2019). The mixed effects of online diversity training. *Proceedings of the National Academy of Sciences*, *116*(16), 7778–7783.

Clore, G. L., & Byrne, D. (1974). A reinforcement-affect model of attraction. *Foundations of Interpersonal Attraction*, 143–170.

Cox, T. H., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *Academy of Management Perspectives*, *5*(3), 45–56.

Devine, P. G., & Ash, T. L. (2022). Diversity training goals, limitations, and promise: a review of the multidisciplinary literature. *Annual Review of Psychology*, *73*, 403–429.

Dobbin, F., & Kalev, A. (2016). Why diversity programs fail. *Harvard Business Review*, *94*(7), 14.

Dobbin, F., & Kelly, E. L. (2007). How to stop harassment: Professional construction of legal compliance in organizations. *American Journal of Sociology*, *112*(4), 1203–1243.

Dover, T. L., Kaiser, C. R., & Major, B. (2020). Mixed signals: The unintended effects of diversity initiatives. *Social Issues and Policy Review*, *14*(1), 152–181.

Dover, T. L., Major, B., & Kaiser, C. R. (2014). Diversity initiatives, status, and system-justifying beliefs: When and how diversity efforts de-legitimize discrimination claims. *Group Processes & Intergroup Relations*, *17*(4), 485–493.

Dover, T. L., Major, B., & Kaiser, C. R. (2016). Members of high-status groups are threatened by pro-diversity organizational messages. *Journal of Experimental Social Psychology*, *62*, 58–67.

Dychtwald, K., Erickson, T., & Morison, B. (2004). It is time to retire retirement. *Public Policy and Aging Report*, *14*(3), 1–28.

Farrell, M. A. (2005). The effect of a market-oriented organisational culture on sales-force behaviour and attitudes. *Journal of Strategic Marketing*, *13*(4), 261–273.

Fasbender, U., & Wang, M. (2017). Negative attitudes toward older workers and hiring decisions: Testing the moderating role of decision makers,Äô core self-evaluations. *Frontiers in Psychology*, *7*, 2057.

Gonzalez, J. A., & DeNisi, A. S. (2009). Cross-level effects of demography and diversity climate on organizational attachment and firm effectiveness. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *30*(1), 21–40.

Gündemir, S., Homan, A. C., Usova, A., & Galinsky, A. D. (2017). Multicultural meritocracy: The synergistic benefits of valuing diversity and merit. *Journal of Experimental Social Psychology*, *73*, 34–41.

Hatch, M. J. (1993). The dynamics of organizational culture. *Academy of Management Review*, *18*(4), 657–693.

Hayes, A. F. (2012). *Process: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling.* University of Kansas, KS.

Heslin, R. (1964). Predicting group task effectiveness from member characteristics. *Psychological Bulletin*, *62*(4), 248.

Hicks-Clarke, D., & Iles, P. (2000). Climate for diversity and its effects on career and organisational attitudes and perceptions. *Personnel Review*.

Hinds, P. J., Carley, K. M., Krackhardt, D., & Wholey, D. (2000). Choosing work group members: Balancing similarity, competence, and familiarity. *Organizational Behavior and Human Decision Processes*, *81*(2), 226–251.

Holmes IV, O., Jiang, K., Avery, D. R., McKay, P. F., Oh, I.-S., & Tillman, C. J. (2020). A meta-analysis integrating 25 years of diversity climate research. *Journal of Management*, 0149206320934547.

Horwitz, S. K., & Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, *33*(6), 987–1015.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*.

Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R. (2013). Presumed fair: Ironic effects of organizational diversity structures. *Journal of Personality and Social Psychology*, *104*(3), 504.

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, *71*(4), 589–617.

Kalinoski, Z. T., Steele-Johnson, D., Peyton, E. J., Leas, K. A., Steinke, J., & Bowling, N. A. (2013). A meta-analytic evaluation of diversity training outcomes. *Journal of Organizational Behavior*, *34*(8), 1076–1104.

Kanter, R. M. (1977). Some effects of proportions on group life. In *The gender gap in psychotherapy* (pp. 53–78). Springer.

Kearney, E., Gebert, D., & Voelpel, S. C. (2009). When and how diversity benefits teams: The importance of team members' need for cognition. *Academy of Management Journal*, *52*(3), 581–598.

Kellough, J. E., & Naff, K. C. (2004). Responding to a wake-up call: An examination of federal agency diversity management programs. *Administration & Society*, *36*(1), 62–90.

Khademi, M., Mast, M. S., Zehnder, C., & De Saint Priest, O. (2021). The problem of demand effects in power studies: Moving beyond power priming. *The Leadership Quarterly*, *32*(4), 101496.

Kirby, T. A., Kaiser, C. R., & Major, B. (2015). Insidious procedures: Diversity awards legitimize unfair organizational practices. *Social Justice Research*, *28*(2), 169–186.

Kulik, C. T., & Roberson, L. (2008). Common goals and golden opportunities: Evaluations of diversity education in academic and organizational settings. *Academy of Management Learning & Education*, *7*(3), 309–331.

Leslie, L. M. (2019). Diversity initiative effectiveness: A typological theory of unintended consequences. *Academy of Management Review*, *44*(3), 538–563.

Marques, J. F. (2010). Colorful window dressing: A critical review on workplace diversity in three major american corporations. *Human Resource Development Quarterly*, *21*(4), 435–446.

McKay, P. F., Avery, D. R., Tonidandel, S., Morris, M. A., Hernandez, M., & Hebl, M. R. (2007). Racial differences in employee retention: Are diversity climate perceptions the key? *Personnel Psychology*, *60*(1), 35–62.

Mumford, M. D., Scott, G. M., Gaddis, B., & Strange, J. M. (2002). Leading creative people: Orchestrating expertise and relationships. *The Leadership Quarterly*, *13*(6), 705–750.

Neuman, G. A., & Wright, J. (1999). Team effectiveness: beyond skills and cognitive ability. *Journal of Applied Psychology*, *84*(3), 376.

O'Reilly III, C. A., Chatman, J., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of Management Journal*, *34*(3), 487–516.

Point, S., & Singh, V. (2003). Defining and dimensionalising diversity: Evidence from corporate websites across europe. *European Management Journal*, *21*(6), 750–761.

Preacher, K. J., & Hayes, A. F. (2004). Spss and sas procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717–731.

Rau, B. L., & Hyland, M. M. (2003). Corporate teamwork and diversity statements in college recruitment brochures: Effects on attraction 1. *Journal of Applied Social Psychology*, *33*(12), 2465–2492.

Roberge, M.-É., Lewicki, R. J., Hietapelto, A., & Abdyldaeva, A. (2011). From theory to practice: Recommending supportive diversity practices. *Journal of Diversity Management (JDM)*, *6*(2), 1–20.

Roberson, L., & Kulik, C. T. (2007). Stereotype threat at work. *Academy of Management Perspectives*, *21*(2), 24–40.

Roberson, L., Kulik, C. T., & Tan, R. Y. (2013). *Effective diversity training.* Oxford University Press.

Rock, D., & Grant, H. (2016). Why diverse teams are smarter. *Harvard Business Review*, *4*(4), 2–5.

Rupp, D. E., Vodanovich, S. J., & Credé, M. (2005). The multidimensional nature of ageism: Construct validity and group differences. *The Journal of Social Psychology*, *145*(3), 335–362.

Sajons, G. B. (2020). Estimating the causal effect of measured endogenous variables: A tutorial on experimentally randomized instrumental variables. *The Leadership Quarterly*, *31*(5), 101348.

Salancik, G. R., & Pfeffer, J. (1978). Uncertainty, secrecy, and the choice of similar others. *Social Psychology*, 246–255.

Schein, E. H. (2010). *Organizational culture and leadership* (Vol. 2). John Wiley & Sons.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*(3), 440.

Schwarzwald, J., & Goldenberg, J. (1979). Compliance and assistance to an authority figure in perceived equitable or nonequitable situations. *Human Relations*, *32*(10), 877–888.

Singh, V., & Point, S. (2006). (re) presentations of gender and ethnicity in diversity statements on european company websites. *Journal of Business Ethics*, *68*(4), 363–379.

Stahl, G. K., Maznevski, M. L., Voigt, A., & Jonsen, K. (2010). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, *41*(4), 690–709.

Staley. (2017). *Why inclusion is a huge issue.* (https://home.barclays/news/2017/03/the-value-of-diversity-jes-staley-tells-mps-why-inclusion-is-a-huge-issue/, Accessed 2021/11/20)

Stevens, F. G., Plaut, V. C., & Sanchez-Burks, J. (2008). Unlocking the benefits of diversity: All-inclusive multiculturalism and positive organizational change. *The Journal of Applied Behavioral Science*, *44*(1), 116–133.

Stock, J., & Yogo, M. (2005). *Asymptotic distributions of instrumental variables statistics with many instruments* (Vol. 6). Chapter.

Suh, J. Y. (2021). Age discrimination in the workplace hurts us all. *Nature Aging*, *1*(2), 147–147.

Tellis, G. J., Prabhu, J. C., & Chandy, R. K. (2009). Radical innovation across nations: The preeminence of corporate culture. *Journal of Marketing*, *73*(1), 3–23.

Tempest, S., Barnatt, C., & Coupland, C. (2002). Grey advantage: New strategies for the old. *Long Range Planning*, *35*(5), 475–492.

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, 227–236.

Thompson, E. P. (1967). Time, work-discipline, and industrial capitalism. *Past & Present*, 56–97.

Thoms, P., Moore, K. S., & Scott, K. S. (1996). The relationship between self-efficacy for participating in self-managed work groups and the big five personality dimensions. *Journal of Organizational Behavior*, *17*(4), 349–362.

UBS. (2017). *What we do: Diversity, equity and inclusion.* (https://www.ubs.com/global/en/our-firm/our-culture/diversity-and-inclusion.html, Accessed 2020/09/30)

Van Vianen, A. E., & De Dreu, C. K. (2001). Personality in teams: Its relationship to social cohesion, task cohesion, and team performance. *European Journal of Work and Organizational Psychology*, *10*(2), 97–120.

Webber, M., Sarris, A., & Bessell, M. (2010). Organisational culture and the use of work–life balance initiatives: Influence on work attitudes and work–life conflict. *The Australasian Journal of Organisational Psychology*, *3*, 54–65.

Wegge, J., Jungmann, F., Liebermann, S., Shemla, M., Ries, B., Diestel, S., & Schmidt, K.-H. (2012). What makes age diverse teams effective? results from a six-year research program. *Work*, *41*(Supplement 1), 5145–5151.

Wilton, L. S., Good, J. J., Moss-Racusin, C. A., & Sanchez, D. T. (2015). Communicating more than diversity: The effect of institutional diversity statements on expectations and performance as a function of race and gender. *Cultural Diversity and Ethnic Minority Psychology*, *21*(3), 315.

Windscheid, L., Bowes-Sperry, L., Kidder, D. L., Cheung, H. K., Morner, M., & Lievens, F. (2016). Actions speak louder than words: Outsiders perceptions of diversity mixed messages. *Journal of Applied Psychology*, *101*(9), 1329.

Footnotes

[1]Additionally, we were initially interested in investigating if team diversity varied depending on whether teams were made for the participant him-herself or for someone else. Hence all our initial studies had two *self-regarding* as well as two *other-regarding* conditions, one of each with the diversity policy. While we do not mention the results of this investigation in this version of the paper, the results are available on request.

[2]We also ran all analyses on the full sample without excluding participants. All results reported are similar insofar as their direction and significance did not change, except for the direction of the effect of the diversity policy on the selection rate of similar team mates, which became negative ($p = 0.63$) ; and the significance of the effect of the desirability of diversity on the selection rate of older teammates, which became insignificant ($p = 0.19$).

[3]We also tested whether the pro-age diversity culture communication had an effect on performance in the quiz. This is not the case across all our studies.

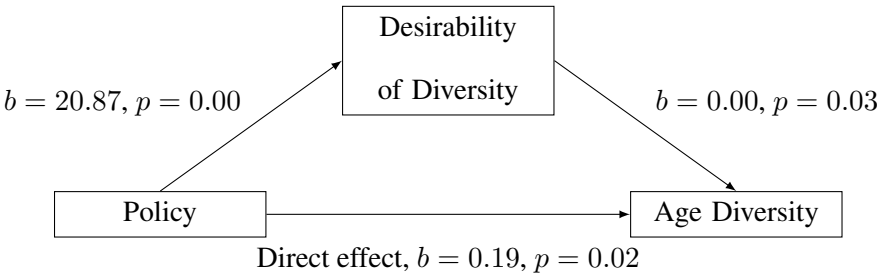
[4]We also ran all analyses on the full sample without excluding participants. All results reported are similar insofar as their direction and significance did not change, except for the significance of the effect of the diversity policy on the perceived closeness of participants, which became significant ($p = 0.03$).

[5]We also ran all analyses on the full sample without excluding participants. All results reported are similar insofar as their direction and significance did not change, except for the significance of the effect of diversity's desirability on the selection rate of older teammates, which became significant ($p = 0.04$).

[6]We also ran all analyses on the full sample without excluding participants. All results reported are similar insofar as their direction and significance did not change, except for the significance of the effect of the diversity policy on perceived age diversity, which became insignificant ($p = 0.35$) ; and the significance of the effect of the diversity policy on perceived closeness to teammates selected, which also became marginally significant ($p = 0.10$).

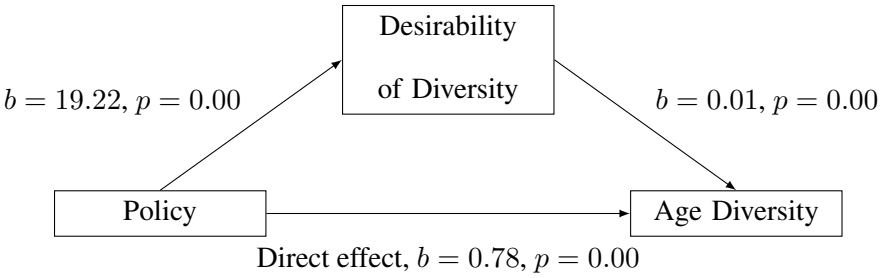
[7]The initial team diversity variable was only collected for studies 1 and 4, as in studies 2 and 3 the initial team just included the participant.

Figure 1: Mediation Model Figure for Age Diversity on Desirability of Diversity through Policy, Study 1



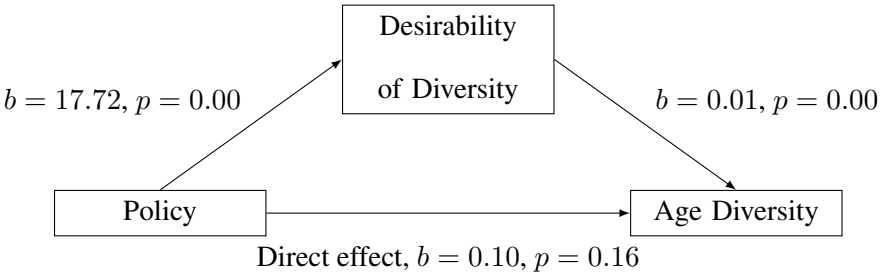
Indirect effect, $b = 0.05$, $95\,\%$ CI [0.00,0.10]

Figure 2: Mediation Model Figure for Age Diversity on Desirability of Diversity through Policy, Study 2



Indirect effect, $b = 0.12$, $95\,\%$ CI [0.02,0.23]

Figure 3: Mediation Model Figure for Age Diversity on Desirability of Diversity through Policy, Study 4



Indirect effect, $b = 0.09$, $95\,\%$ CI [0.03,0.15]

TABLE 1: *Overview of Studies*

	Study 1	Study 2	Study 3	Study 4
Number of people already on the team	2	1	1	2
Number of teammates to choose	2	2	2	1
Number of teammates to choose from	4	8	4	4
Dimensions on which teammates varied	Age Perceived personality similarity	Age Gender Worldview	Age Perceived personality similarity	Age Perceived personality similarity
N	105	156	164	242
% men	55	51	46	55
Mean age	25.51	24.75	24.98	26.87

TABLE 2: *Correlations & Descriptive Statistics, Study 1*

Variables	Older	Diversity Policy	Same	Desirability	Feasibility	IndexAdv	AgeDiv	PersDiv	Nice	Enjoyable	Supervisor	Gender	Experience
Older													
Diversity Policy	0.25												
	(0.02)												
Same	-0.10	0.02											
	(0.35)	(0.84)											
Desirability	0.26	0.38	-0.31										
	(0.02)	(0.00)	(0.00)										
Feasibility	0.14	0.20	-0.20	0.74									
	(0.21)	(0.08)	(0.06)	(0.00)									
IndexAdv	-0.23	-0.06	0.06	0.23	0.37								
	(0.04)	(0.62)	(0.58)	(0.03)	(0.00)								
AgeDiv	0.25	0.24	0.01	0.35	0.29	0.13							
	(0.03)	(0.03)	(0.91)	(0.00)	(0.01)	(0.24)							
PersDiv	-0.04	0.10	-0.69	0.43	0.30	0.19	0.10						
	(0.74)	(0.35)	(0.00)	(0.00)	(0.01)	(0.08)	(0.36)						
Nice	-0.16	-0.07	0.35	-0.11	0.12	0.42	0.20	-0.18					
	(0.14)	(0.51)	(0.00)	(0.34)	(0.29)	(0.00)	(0.07)	(0.10)					
Enjoyable	-0.14	-0.11	0.42	-0.17	0.09	0.39	0.07	-0.24	0.81				
	(0.21)	(0.32)	(0.00)	(0.13)	(0.43)	(0.00)	(0.53)	(0.03)	(0.00)				
Supervisor	0.08	0.00	-0.10	0.04	0.18	0.20	-0.07	0.20	-0.05	-0.02			
	(0.47)	(0.98)	(0.37)	(0.72)	(0.10)	(0.07)	(0.54)	(0.07)	(0.68)	(0.84)			
Gender	-0.00	-0.00	0.12	-0.01	-0.02	-0.28	-0.04	-0.12	0.06	0.02	-0.26		
	(0.98)	(0.99)	(0.28)	(0.93)	(0.86)	(0.01)	(0.74)	(0.30)	(0.62)	(0.86)	(0.02)		
Experience	0.13	0.01	-0.06	0.27	0.34	0.16	0.34	0.17	0.14	0.16	0.25	0.10	
	(0.25)	(0.94)	(0.60)	(0.01)	(0.00)	(0.16)	(0.00)	(0.13)	(0.22)	(0.16)	(0.02)	(0.39)	
mean	1.05	0.52	1.48	64.32	61.93	311.71	5.65	3.80	5.56	5.62	0.23	0.51	2.55
sd	0.38	0.50	0.59	27.71	25.24	97.48	1.37	1.97	1.03	1.04	0.42	0.53	1.21
min	0.00	0.00	0.00	0.00	0.00	30.00	1.00	1.00	4.00	4.00	0.00	0.00	0.00
max	2.00	1.00	2.00	100.00	100.00	500.00	7.00	7.00	7.00	7.00	1.00	2.00	4.00

IndexAdv = sum score of advantages of diversity, rated on a scale from 0 to 500. Participants had to rate on a scale from 0 to 100 how much they thought about the following points when making their teammate selection decision: choosing these people will increase the team's performance, make sure the team is happy working together, bring more talent to the team, not slow down the team's work and is the morally right thing to do.

TABLE 3: *Overview of Main Results: Selection Rates of Older and Personally Similar Team Members*

	Older Team Members	Personally Similar Team Members
Study 1		
Diversity Policy	0.19*	0.03
	(0.02)	(0.84)
Constant	0.85**	1.49**
	(0.00)	(0.00)
Controls for gender, supervisory duties and experience	Yes	Yes
Observations	82	82
Study 2		
Diversity Policy	0.90**	-0.09
	(0.00)	(0.44)
Constant	0.06	1.56**
	(0.71)	(0.00)
Controls for gender, employment status and experience	Yes	Yes
Observations	118	118
Study 3		
Diversity Policy	0.43**	-0.12
	(0.00)	(0.28)
Constant	0.73**	1.05**
	(0.00)	(0.00)
Controls for gender, employment status and experience	Yes	Yes
Observations	133	133
Study 4		
Diversity Policy	0.10	0.02
	(0.17)	(0.75)
Constant	0.24*	0.53**
	(0.05)	(0.00)
Controls for gender, employment status and experience	Yes	Yes
Observations	184	184

Note. Older = older teammate chosen. Older was coded as 2 if the participant chose 2 older teammates, 1 if the participant chose 1 older teammate, 0 otherwise. Same = teammate personality similarity. Same was coded as 2 if the participant chose 2 teammates similar in personality to him-her, 1 if the participant chose 1 teammate similar in personality to him-her, 0 otherwise. Diversity policy = company diversity policy. Diversity policy was coded as 1 if the participant was in the condition where the company had a diversity policy, 0 if not. Gender was coded as 0 = Male, 1 = Female, 2 = Other. Employment status = whether the participant was employed. Employment status was coded as 0, if the participant was unemployed, 1 otherwise. Supervisory duties values were coded as 0, if the participant had no supervisory duties, 1 otherwise. Experience = experience working in teams mixed in terms of age. Experience was coded on a 5-point Likert scale from 0 if the participant had no experience working in teams mixed in terms of age, up until 4 if the participant had a lot of experience working in teams mixed in terms of age. Unstandardized B coefficients are shown ; p statistics derived from robust standard errors in parentheses. * $p < .05$. ** $p < .01$.

TABLE 4: *Linear Regression Analysis: Process and Outcome Variables*

	Desirability	Feasibility	IndexAdv	AgeDiv	PersDiv	Closeness	Nice	Enjoyable	GenderDiv
Study 1									
Diversity Policy	20.87**	9.74	-11.03	0.64*	0.40		-0.16	-0.24	
	(0.00)	(0.07)	(0.59)	(0.03)	(0.35)		(0.50)	(0.31)	
Constant	38.57**	39.30**	305.16**	4.47**	3.03**		5.33**	5.41**	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Controls for gender, supervisory duties and experience	Yes	Yes	Yes	Yes	Yes		Yes	Yes	
Observations	82	82	82	82	82		82	82	
Study 2									
Diversity Policy	19.22**	11.68*	13.99	1.93**	0.63	-0.35			0.48
	(0.00)	(0.03)	(0.39)	(0.00)	(0.12)	(0.18)			(0.24)
Constant	35.46**	36.63**	345.78**	1.82**	2.23**	4.89**			3.71**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)			(0.00)
Controls for gender, employment status and experience	Yes	Yes	Yes	Yes	Yes	Yes			Yes
Observations	118	118	118	118	118	118			118
Study 3									
Diversity Policy	5.44	3.40	0.74	0.54	0.48	-0.18			
	(0.20)	(0.40)	(0.97)	(0.08)	(0.16)	(0.47)			
Constant	49.82**	53.17**	334.65**	4.38**	4.23**	3.59**			
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)			
Controls for gender, employment status and experience	Yes	Yes	Yes	Yes	Yes	Yes			
Observations	133	133	133	133	133	133			
Study 4									
Diversity Policy	17.72**	14.73**	47.00**	0.47*	0.58**	0.65**			
	(0.00)	(0.00)	(0.00)	(0.04)	(0.01)	(0.01)			
Constant	45.89**	50.50**	319.34**	4.16**	4.01**	3.02**			
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)			
Controls for gender, employment status and experience	Yes	Yes	Yes	Yes	Yes	Yes			
Observations	184	184	184	184	184	184			

Note. Desirability = "creating high performing age diverse teams is very desirable", coded from 0 to 100. Feasibility = "creating high performing age diverse teams is easily feasible", coded from 0 to 100. IndexAdv= sum score of advantages of diversity. Agediv = how diverse in terms of age is the team in the participant's opinion, from "not diverse at all" to "very diverse" on a 7-Point scale. Persdiv = how diverse in terms of personality is the team in the participant's opinion, from "not diverse at all" to "very diverse" on a 7-Point scale. Closeness=how close the participant feels to the members of the team created, from "not close at all" to "very close" on a 7-Point scale. Nice=how nice the participant feels the members of the team created are, from "not nice at all" to "very nice" on a 7-Point scale. Enjoyable=how enjoyable the participant feels it would be to work with the members of the team created, from "not enjoyable at all" to "very enjoyable" on a 7-Point scale. Genderdiv = how diverse in terms of gender is the team in the participant's opinion, from "not diverse at all" to "very diverse" on a 7-Point scale. Diversity policy = company diversity policy. Diversity policy was coded as 1 if the participant was in the condition where the company had a diversity policy, 0 if not. Gender was coded as 0 = Male, 1 = Female, 2 = Other. Employment status = whether the participant was employed. Employment status was coded as 0, if the participant was unemployed, 1 otherwise. Supervisory duties values were coded as 0, if the participant had no supervisory duties, 1 otherwise. Experience = experience working in teams mixed in terms of age. Experience was coded on a 5-point Likert scale from 0 if the participant had no experience working in teams mixed in terms of age, up until 4 if the participant had a lot of experience working in teams mixed in terms of age. Unstandardized B coefficients are shown ; p statistics derived from robust standard errors in parentheses. * $p < .05$. ** $p < .01$.

TABLE 5: *Selection Rates of Older and Personally Similar Team Members*

	Older team members
Study 1	
Desirability	0.00*
	(0.01)
Constant	0.77**
	(0.00)
Controls for gender, supervisory duties and experience	Yes
Observations	82
Study 2	
Desirability	0.01**
	(0.00)
Constant	0.13
	(0.48)
Controls for gender, employment status and experience	Yes
Observations	118
Study 3	
Desirability	0.00
	(0.18)
Constant	0.77**
	(0.00)
Controls for gender, employment status and experience	Yes
Observations	133
Study 4	
Desirability	0.01**
	(0.00)
Constant	0.02
	(0.90)
Controls for gender, employment status and experience	Yes
Observations	184

Note. Older = older teammate chosen. Older was coded as 2 if the participant chose 2 older teammates, 1 if the participant chose 1 older teammate, 0 otherwise. Desirability = "creating high performing age diverse teams is very desirable", coded from 0 to 100. Gender was coded as 0 = Male, 1 = Female, 2 = Other. Supervisory duties values were coded as 0, if the participant had no supervisory duties, 1 otherwise. Employment status = whether the participants was employed. Employment status was coded as 0 if the participant was unemployed, 1 otherwise. Experience = experience working in teams mixed in terms of age. Experience was coded on a 5-point Likert scale from 0 if the participant had no experience working in teams mixed in terms of age, up until 4 if the participant had a lot of experience working in teams mixed in terms of age. Unstandardized B coefficients are shown ; p statistics derived from robust standard errors in parentheses. * $p < .05$. ** $p < .01$.

TABLE 6: *Correlations & Descriptive Statistics, Study 2*

Variables	Older	Diversity Policy	Same	Desirability	Feasibility	IndexAdv	AgeDiv	PersDiv	Close	GenderDiv	Employed	Gender	Experience
Older													
Diversity Policy	0.570												
	(0.00)												
Same	-0.02	-0.06											
	(0.82)	(0.52)											
Desirability	0.42	0.27	-0.15										
	(0.00)	(0.00)	(0.11)										
Feasibility	0.31	0.18	-0.02	0.62									
	(0.00)	(0.05)	(0.82)	(0.00)									
IndexAdv	-0.02	0.06	0.25	0.25	0.31								
	(0.83)	(0.54)	(0.01)	(0.01)	(0.00)								
AgeDiv	0.54	0.43	0.01	0.57	0.49	0.19							
	(0.00)	(0.00)	(0.88)	(0.00)	(0.00)	(0.03)							
PersDiv	0.18	0.11	-0.55	0.12	0.03	-0.15	0.10						
	(0.05)	(0.24)	(0.00)	(0.21)	(0.77)	(0.10)	(0.27)						
Close	0.02	-0.11	0.30	0.14	0.14	0.31	-0.03	-0.27					
	(0.81)	(0.24)	(0.00)	(0.13)	(0.13)	(0.00)	(0.77)	(0.00)					
GenderDiv	0.22	0.11	0.02	0.12	0.07	0.20	0.07	0.15	0.02				
	(0.02)	(0.24)	(0.82)	(0.20)	(0.47)	(0.03)	(0.44)	(0.10)	(0.80)				
Employed	0.14	0.06	-0.04	0.00	0.14	0.00	0.19	0.05	-0.14	0.09			
	(0.14)	(0.49)	(0.70)	(0.99)	(0.13)	(0.97)	(0.03)	(0.61)	(0.12)	(0.34)			
Gender	0.04	-0.10	-0.19	0.09	-0.00	-0.08	0.06	0.19	-0.25	-0.03	0.19		
	(0.64)	(0.29)	(0.04)	(0.32)	(0.99)	(0.37)	(0.51)	(0.04)	(0.01)	(0.75)	(0.04)		
Experience	0.13	-0.10	0.05	0.23	0.21	0.23	0.24	0.09	0.02	0.12	0.28	0.04	
	(0.17)	(0.29)	(0.56)	(0.01)	(0.03)	(0.01)	(0.01)	(0.34)	(0.81)	(0.20)	(0.00)	(0.63)	
mean	0.97	0.56	1.46	62.12	60.18	386.28	4.53	3.39	4.23	4.66	0.78	0.48	2.54
sd	0.76	0.50	0.64	30.00	29.57	88.02	2.09	2.23	1.43	2.11	0.42	0.50	1.17
min	0.00	0.00	0.00	0.00	0.00	46.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
max	2.00	1.00	2.00	100.00	100.00	582.00	7.00	7.00	7.00	7.00	1.00	1.00	4.00

IndexAdv = sum score of advantages of diversity, rated on a scale from 0 to 600. Participants had to rate on a scale from 0 to 100 how much they thought about the following points when making their teammate selection decision: working with the people I chose will increase the team's performance, the people I chose will bring more talent to the team, not slow down the team's work, make sure that the team's work satisfaction will not decrease, make the team's outcome less uncertain, selecting the people I chose was the morally right thing to do.

TABLE 7: *Correlations & Descriptive Statistics, Study 3*

Variables	Older	Diversity Policy	Same	Desirability	Feasibility	IndexAdv	AgeDiv	PersDiv	Close	Employed	Gender	Experience
Older												
Diversity Policy	0.39											
	(0.00)											
Same	-0.03	-0.09										
	(0.77)	(0.30)										
Desirability	0.14	0.13	-0.09									
	(0.11)	(0.14)	(0.30)									
Feasibility	0.17	0.10	-0.22	0.57								
	(0.05)	(0.27)	(0.01)	(0.00)								
IndexAdv	-0.05	-0.00	0.03	0.28	0.19							
	(0.60)	(0.99)	(0.72)	(0.00)	(0.03)							
AgeDiv	0.34	0.17	-0.01	0.36	0.36	0.09						
	(0.00)	(0.06)	(0.86)	(0.00)	(0.00)	(0.31)						
PersDiv	0.11	0.12	-0.55	0.23	0.23	-0.00	-0.00					
	(0.23)	(0.19)	(0.00)	(0.01)	(0.01)	(0.96)	(0.96)					
Close	-0.15	-0.05	0.14	0.16	0.22	0.30	0.05	0.09				
	(0.08)	(0.55)	(0.12)	(0.07)	(0.01)	(0.00)	(0.55)	(0.32)				
Employed	0.08	-0.04	0.09	0.09	0.02	0.19	0.05	0.08	0.10			
	(0.37)	(0.61)	(0.28)	(0.32)	(0.81)	(0.03)	(0.60)	(0.38)	(0.26)			
Gender	0.15	0.03	0.03	0.12	0.06	-0.08	0.06	-0.04	-0.02	-0.04		
	(0.08)	(0.76)	(0.73)	(0.17)	(0.52)	(0.35)	(0.49)	(0.68)	(0.83)	(0.63)		
Experience	-0.01	0.14	0.05	0.12	0.14	0.05	0.05	-0.03	0.10	0.12	0.10	
	(0.88)	(0.11)	(0.55)	(0.18)	(0.11)	(0.55)	(0.58)	(0.75)	(0.23)	(0.16)	(0.23)	
mean	1.05	0.55	1.19	63.98	63.20	373.49	5.00	4.50	4.05	0.80	0.55	2.66
sd	0.53	0.50	0.62	23.10	21.59	101.25	1.64	1.87	1.41	0.40	0.53	1.06
min	0.00	0.00	0.00	0.00	0.00	13.00	1.00	1.00	1.00	0.00	0.00	0.00
max	2.00	1.00	2.00	100.00	100.00	600.00	7.00	7.00	7.00	1.00	2.00	4.00

IndexAdv = sum score of advantages of diversity, rated on a scale from 0 to 600. Participants had to rate on a scale from 0 to 100 how much they thought about the following points when making their teammate selection decision: working with the people I chose will increase my team's performance, the people I chose will bring more talent to my team, not slow down my team's work, make sure that my team's work satisfaction will not decrease, make my team's outcome less uncertain, selecting the people I chose was the morally right thing to do.

TABLE 8: *Correlations & Descriptive Statistics, Study 4*

Variables	Older	Diversity Policy	Same	Desirability	Feasibility	IndexAdv	AgeDiv	PersonalityDiv	Closeness	Employed	Gender	Experience	Team Incentives
Older													
Diversity Policy	0.10												
	(0.17)												
Same	0.01	0.02											
	(0.89)	(0.76)											
Desirability	0.29	0.32	-0.10										
	(0.00)	(0.00)	(0.18)										
Feasibility	0.22	0.28	-0.10	0.54									
	(0.00)	(0.00)	(0.18)	(0.00)									
IndexAdv	-0.02	0.21	0.09	0.31	0.28								
	(0.77)	(0.00)	(0.21)	(0.00)	(0.00)								
AgeDiv	0.39	0.15	-0.02	0.47	0.46	0.19							
	(0.00)	(0.05)	(0.79)	(0.00)	(0.00)	(0.01)							
PersonalityDiv	-0.07	0.19	-0.44	0.22	0.14	0.10	0.13						
	(0.35)	(0.01)	(0.00)	(0.00)	(0.06)	(0.16)	(0.07)						
Closeness	-0.08	0.19	0.15	0.22	0.09	0.38	0.12	0.02					
	(0.28)	(0.01)	(0.05)	(0.00)	(0.21)	(0.00)	(0.12)	(0.77)					
Employed	-0.02	-0.04	-0.05	0.05	0.01	0.11	0.06	-0.00	0.12				
	(0.80)	(0.58)	(0.53)	(0.49)	(0.91)	(0.13)	(0.40)	(0.97)	(0.11)				
Gender	-0.02	0.03	-0.13	0.02	0.01	-0.21	-0.07	0.05	-0.12	-0.14			
	(0.79)	(0.68)	(0.08)	(0.82)	(0.85)	(0.00)	(0.36)	(0.53)	(0.09)	(0.06)			
Experience	0.20	-0.01	0.07	0.06	-0.02	0.12	0.09	-0.05	0.07	0.27	-0.04		
	(0.01)	(0.94)	(0.36)	(0.46)	(0.79)	(0.12)	(0.21)	(0.50)	(0.33)	(0.00)	(0.61)		
Team Incentives	-0.03	0.01	-0.10	-0.05	0.02	0.05	-0.09	0.01	0.02	-0.02	-0.04	-0.10	
	(0.73)	(0.85)	(0.18)	(0.48)	(0.80)	(0.49)	(0.21)	(0.88)	(0.78)	(0.80)	(0.59)	(0.17)	
mean	0.47	0.53	0.49	61.45	58.33	360.99	4.73	4.26	3.67	0.80	0.50	2.51	0.47
sd	0.50	0.50	0.50	27.50	26.39	105.30	1.55	1.50	1.62	0.40	0.54	1.06	0.50
min	0.00	0.00	0.00	0.00	0.00	86.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00
max	1.00	1.00	1.00	100.00	100.00	600.00	7.00	7.00	7.00	1.00	2.00	4.00	1.00

IndexAdv = sum score of advantages of diversity, rated on a scale from 0 to 600. Participants had to rate on a scale from 0 to 100 how much they thought about the following points when making their teammate selection decision: working with the people I chose will increase my team's performance, the people I chose will bring more talent to my team, not slow down my team's work, make sure that my team's work satisfaction will not decrease, make my team's outcome less uncertain, selecting the people I chose was the morally right thing to do.

TABLE 9: *Descriptive Statistics, Study 5*

	AgeDiv	PersDiv	GenderDiv	ActionOrientation	Explicitness	Persuasiveness	Directness
Study 1							
mean	2.62	3.56	3.50	4.88	4.21	4.10	4.98
sd	1.72	1.20	1.54	1.52	1.82	1.63	1.44
min	1.00	1.00	1.00	1.00	1.00	1.00	2.00
max	7.00	7.00	7.00	7.00	7.00	7.00	7.00
Studies 2 and 3							
mean				4.62	4.77	4.08	5.46
sd				1.63	1.81	1.60	1.61
min				1.00	1.00	1.00	1.00
max				7.00	7.00	7.00	7.00
Study 4							
mean	5.69	4.56	3.69	4.75	4.35	3.98	5.12
sd	1.21	1.03	1.03	1.51	1.58	1.52	1.55
min	2.00	2.00	1.00	1.00	1.00	1.00	2.00
max	7.00	7.00	6.00	7.00	7.00	7.00	7.00

Note. AgeDiv = perceived age diversity of the initial team in the study, coded on a 7-Point Likert scale from "not diverse at all" to "very diverse". PersDiv = perceived personality diversity of the initial team in the study, coded on a 7-Point Likert scale from "not diverse at all" to "very diverse". GenderDiv = perceived gender diversity of the initial team in the study, coded on a 7-Point Likert scale from "not diverse at all" to "very diverse". ActionOrientation = how likely participants said they were to act on the diversity policy in the study, from a scale from "not at all" to "very much". Explicitness = how explicit participants found the diversity policy in the study, from a scale from "not at all" to "very much". Persuasiveness = how persuasive participants found the diversity policy in the study, from a scale from "not at all" to "very much". Directness = how direct participants found the diversity policy in the study, from a scale from "not at all" to "very much".