AN EXPERIENCE THAT MAKES ME FEEL SUCCESSFUL AND PROUD:

ESSAYS ON SELF-AFFIRMATION AND PRODUCTIVITY, AND ON LOSS AVERSION IN CONTESTS

Inauguraldissertation zur Erlangung des Doktorgrades der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2025

vorgelegt von

Sebastian Schneiders, M. Sc.

Referent*in: Prof. Dr. Bettina Rockenbach Korreferent*in: Prof. Dr. Matthias Sutter Tag der Promotion: 25.02.2025

ACKNOWLEDGEMENTS

First and foremost, I want to thank Bettina Rockenbach for her excellent supervision throughout my doctoral studies. I am deeply grateful for your support, advice, encouragement, inspiration, and your trust over such a long time. You created an environment that gave me exceptional academic and personal freedom. This thesis would not have been possible without you. I am also indebted to the other members of the committee: Matthias Sutter and Sebastian Siegloch.

I especially thank Sebastian Tonke. Working and spending time with you has always been a great pleasure, both professionally and personally. I am grateful for all your feedback and your invaluable tips and insights.

I thank Arne R. Weiss for countless ideas and discussions as a co-author and your great support before and during my doctoral studies. I also thank my other co-authors, Björn Vollan and Marcin Waligora, and all the colleagues at the university: Suparee Boonmanunt, Linda Ethner, Susanna Grundmann, Sören Harrs, Heike Kirch, Felix Kölle, Thomas Lauer, Viet A Nguyen, Lukas Reinhardt, Anne Schielke, Eugenio Verrina, Lukas Wenner, Christopher Zeppenfeld, Jarid Zimmermann, and Nina Zimmermann. I also wish to thank Alexander Morell and Roland Imhoff, it is a pity that our projects did not work out in the end. I finally thank Michael Kurschilgen for introducing me to the great fun of experimental research and for your constant advice and support.

The field research in this dissertation would not have been possible without the help of many people. I am grateful for the outstanding research assistance of Jonathan Kempen and Barbara Scharfbillig in Namibia and for their extensive consultation, together with Michael Bollig and Ute Dieckmann, the Legal Assistance Center (LAC), and the Desert Research Foundation of Namibia (DRFN). I am also very thankful to all research assistants involved, who contributed a lot to the success of this project: Johannes Abusema, Aaron Anton, Johannes Dam, Appolia Dabe, Hanna Fuhrmann, Samuel Hartenberg, Markus Hendrik, Erastus Itembu, Charlotte Krömer, Martha Lubhaan, and Paulina Swaartbooi.

I thank the Cologne Graduate School in Management, Economics, and Social Sciences (CGS), the Center for Social and Economic Behavior (C-SEB), and the Chair of Behavioral and Experimental Economics for providing financial support of various kinds.

Finally, I thank my family and everyone around me who always, sometimes, or never believed that this endeavor would finally come to a happy end. Thank you for your patience, your love, and your support. This thesis is dedicated to all of you.

TABLE OF CONTENTS

Chapter	r 1. Introduction	1
Chapter	r 2. Self-Affirmation and Productivity: Backfiring Among Those Who	Could
Benefit	the Most	6
2.1.	Introduction	7
2.2.	Field Experiment	10
2.2	2.1. Experimental Design	10
2.2	2.2. Conceptual Framework	14
2.3.	Results of the Field Experiment	16
2.3	.1. Treatment Effects	16
2.3	2.2. Exploring the Backfiring: Difficulties in Reporting Successful Even	ıts 17
2.4.	Online Experiment	19
2.4	.1. Experimental Design	19
2.4	.2. Main Results of the Online Experiment	20
2.4	.3. Additional Treatments: Timer	23
2.5.	Conclusion and Discussion	24
Apper	ndix A: Additional Figures and Tables	25
Apper	ndix B: A Simple Theoretical Framework	40
Apper	ndix C: Alternative Explanations (Field Experiment)	42
Apper	ndix D: Psychological Questionnaire (Field Experiment)	44
Apper	ndix E: Experimental Protocol and Materials	45
Chapter	r 3. Backfiring or Benefits? Heterogeneous Effects of Self-Affirmation	on
Product	tivity	54
3.1.	Introduction	55
3.2.	Three Potential Moderators	58
3.3.	Experimental Design	59
3.4.	Theory and Hypotheses	61
3.5.	Empirical Strategy	63
3.6.	Results	65
3.6	.1. Heterogeneous Effects: Gender	66
3.6	2.2. Heterogeneous Effects: Socioeconomic Status	69
3.6	.3. Heterogeneous Effects: Self-efficacy	71

3.7. Limitations and Conclusion	74
Appendix A: Additional Tables	
Appendix B: Instructions	
Chapter 4. Pushing the Bad Away: Reverse Tullock Contests	
References	
List of Applied Software	

CHAPTER 1

INTRODUCTION

Economics has a long tradition of integrating psychological insights, dating back to Adam Smith, who already explored concepts now recognized as loss aversion, overconfidence, and present bias (Thaler 2016). However, over time, psychological realism of economic theories diminished in favor of mathematically tractable models based on the famous *homo economicus* assumptions. In the last 40 years, economics reembraced its roots. Economists have studied systematic deviations from purely rational behavior and incorporated human psychology back into their models, creating a more evidence-based behavioral economics (see, e.g., DellaVigna 2009; Rabin 1998; 2013; Thaler 2016).

But not only have economists recognized that human psychology affects economic decision-making. It has recently been acknowledged that economic circumstances can also systematically affect human psychology: living in poverty, for example, has direct consequences for the cognitive functions of the poor, giving rise to a specific "psychology of poverty" (Haushofer and Fehr 2014; Schilbach, Schofield, and Mullainathan 2016; E. Dean, Schilbach, and Schofield 2018; Kremer, Rao, and Schilbach 2019; Haushofer and Salicath 2023). Scarcity of economic resources can impede the poor's cognitive capacity, as persistent thoughts about limited financial resources consume scarce mental resources, preventing their more productive use (Mani et al. 2013; Mullainathan and Shafir 2013). This mental burden can, in turn, have direct consequences for economic outcomes, such as impaired labor productivity (Kaur et al. 2024).

Living in poverty is often associated with other non-monetary deprivations that shape the specific psychology of the poor and affect their economic outcomes, including higher levels of stress (Haushofer and Fehr 2014), greater exposition to loud noise (J. Dean 2024), and poor sleeping conditions (Bessone et al. 2021). Moreover, poverty is also often associated with negative stereotypes, stigmatization, and feelings of shame, creating further mental burdens on the poor (Hall, Zhao, and Shafir 2014; Ghosal et al. 2022). Addressing the psychological roots of poverty and its psychological consequences calls for policies that go beyond the traditional economic policy toolbox. Consequently, a recent literature has explored psychological interventions explicitly tailored to the specific psychology of the poor, ranging from one-shot, light-touch interventions, such as showing aspirational videos, to multi-week programs involving professional therapy sessions (for an overview, see Haushofer and Salicath 2023).

This dissertation proposes self-affirmation, a widely used light-touch, low-cost, and easy-to-implement psychological intervention aimed at creating a positive self-image, as a potential policy to counteract the negative stereotypes and stigma associated with poverty (Hall et al. 2014; Cohen and Sherman 2014). Chapters 2 and 3 examine how self-affirmation affects productivity in a field experiment and in an online experiment. Labor productivity is of central interest to economics and policymakers alike; yet, it has hitherto received limited attention in the literature on the psychological antecedents and consequences of poverty. The findings presented in this dissertation suggest that standard self-affirmation interventions have to be treated with caution as they may be ineffective and can even backfire for those who could benefit the most. Nevertheless, the findings provide clear evidence that labor productivity is malleable through psychological interventions, paving the way for the development of more effective policies in the future.

Chapter 2 (*"Self-Affirmation and Productivity: Backfiring Among Those Who Could Benefit the Most"*) is joint work with Bettina Rockenbach, Sebastian Tonke, Björn Vollan, and Arne R. Weiss.¹ In a field experiment among an extremely poor and stigmatized group in Namibia, we study the effect of a standard self-affirmation intervention, i.e., asking participants to recall experiences that made them feel successful and proud, on labor productivity. In a control condition, we simply ask participants to recall their daily routines. In both experimental conditions, participants then work on a real agricultural crop production job with piece-rate incentives. Surprisingly, the self-affirmation intervention backfires: participants are substantially less productive than those in the control condition. An analysis of participants' answers to the self-affirmation intervention suggests that the backfiring can be explained by difficulties in recalling positive experiences.

To provide causal evidence for this conjecture, we conduct a well-powered online follow-up experiment with U.S. participants, who are either assigned to variants of a selfaffirmation intervention or to a neutral control condition. The self-affirmation treatments exogenously vary the difficulty of recalling experiences. We ask participants to recall either three (easy) or eight (difficult) experiences that made them feel successful and proud. In all

¹ The chapter is based on Rockenbach et al. (2024). The paper currently has a "revise and resubmit" at the *Economic Journal*.

conditions, participants then work on a standard real-effort task with piece-rate incentives. Our online experiment replicates the findings from the field experiment in Namibia in a completely different context and shows that self-affirmation can backfire, especially when it is too difficult. Findings from a psychological questionnaire substantiate the mechanism: difficulties in recalling successful events can disaffirm participants, increase mental burdens and consequently reduce productivity.

Chapter 3 (*"Backfiring or Benefits? Heterogeneous Effects of Self-Affirmation on Productivity"*) is single authored. I follow up on the previous chapter and further analyze data from the online experiment to understand whether the negative effect of self-affirmation varies systematically across important sociodemographic and psychological dimensions and whether there are subgroups of participants for which self-affirmation has the intended positive effects. As pre-registered, I investigate how the intervention's effects differ by gender, socioeconomic status, and baseline self-efficacy. I study potential heterogeneous effects of the two self-affirmation treatments (easy and difficult) and a control treatment.

The results suggest that there is some heterogeneity in the effects of self-affirmation on productivity. Specifically, the difficulty of self-affirmation protocols appears to be a sensitive factor influencing their effects, and this chapter shows heterogeneity in this mechanism. While a difficult self-affirmation protocol impairs productivity for some groups, it has no such effect on others. Strikingly, self-affirmation does not benefit any of the analyzed subgroups, also when recalling self-affirming experiences is easy. Instead, even the easy self-affirmation protocol backfires for some participants. These insights have important practical implications for the design of future self-affirmation interventions. Researchers and practitioners who want to implement psychological interventions in the field should carefully test their protocols to understand and consider potential heterogeneity in order to avoid unintended negative effects and to create interventions that benefit their recipients.

This dissertation also contributes to a more classical branch of modern behavioral economics by studying behavior in gain frames versus loss frames in its last chapter. Standard economic theory predicts no differences in behavior when equal changes in wealth are framed as gains or losses. However, already Adam Smith knew that "pain ... is, in almost all cases, a more pungent sensation than the opposite and correspondent pleasure." (Smith 1792, p. 305; also see Thaler 2016). Prospect theory (Kahneman and Tversky 1979), one of the corner stones of behavioral economics, models this sentiment as "loss aversion", recognizing that losses create larger

changes in utility than equally sized gains. Chapter 4 applies this theory to behavior in rentseeking contests. The chapter presents surprising findings and discusses several explanations that could advance a behavioral theory of decision-making in contest.

Chapter 4 ("*Pushing the Bad Away: Reverse Tullock Contests*") is joint work with Bettina Rockenbach and Marcin Waligora.² We study behavior in contests for avoiding losses as compared to contests for achieving gains, both theoretically and in a laboratory experiment. Contests for avoiding losses or for avoiding something bad coming are omnipresent. Imagine, for example, that one of the primary schools in a city needs to be closed down. To avoid long travel times, different neighborhoods will start lobbying against closing their school. Similarly, imagine that two potential locations for a waste disposal have been identified and the government needs to decide which one to pick. Again, local communities will be ready and determined to spend resources to avoid their community being chosen. However, such 'reverse' contests are largely understudied. While under standard preferences investments in reverse and conventional contests should not be different, loss aversion predicts contests for avoiding losses to be fiercer than conventional ones.

Our findings are surprising: the difference in investments between conventional and reverse contests is small and statistically insignificant, in contrast to our predictions. This may be due to a theory of decision-making in contests that is still too simple and neglects further "seemingly irrelevant factors" (Thaler 2016). For example, expanding the prospect theory model with joy of winning, a psychological motivation commonly found in contest behavior, can explain the results post-hoc. Future research can build on these results and try to better understand the importance of non-standard motives, advancing a more psychologically realistic model that accurately predicts contest behavior in both gain and loss frames.

Contributions. My contributions to the respective chapters of this dissertation are as follows.

Chapter 2 is joint work with Bettina Rockenbach, Sebastian Tonke, Björn Vollan, and Arne R. Weiss. I provided major contributions to the conceptualization and the design of the field and the online experiments. I prepared the field experiment and collected the data in Namibia. I programmed the online experiment and collected the data. I contributed significantly to the data analysis of both experiments and to writing the paper.

² This chapter is based on Rockenbach, Schneiders, and Waligora (2018), published in the *Journal of the Economic Science Association*.

Chapter 3 is single authored. I conceptualized the research, analyzed the data from the online experiment of Chapter 2 with respect to the research questions of this paper and I wrote the paper.

Chapter 4 is joint work with Bettina Rockenbach and Marcin Waligora. Bettina and Marcin designed the original experiment, collected and analyzed the original data, and wrote the first version of the paper. Together, we conceptualized a modified version of the experiment. I programmed the new experiment and collected and analyzed the data presented in the final paper. I significantly contributed to writing the final version of the paper.

SELF-AFFIRMATION AND PRODUCTIVITY: BACKFIRING AMONG THOSE WHO COULD BENEFIT THE MOST

Joint work with Bettina Rockenbach, Sebastian Tonke, Björn Vollan, and Arne R. Weiss

Abstract. In a field experiment among an extremely poor and stigmatized group in Namibia, we study the effect of a standard self-affirmation intervention – i.e., asking participants to recall experiences that made them feel successful and proud – on labor productivity. In a control condition, we simply ask participants to recall their daily routines. In both experimental conditions, participants then work on a real job with piece-rate incentives. Surprisingly, the self-affirmation intervention backfires: participants are substantially less productive than those in the control condition. An analysis of participants' answers to the self-affirmation intervention suggests that the backfiring can be explained by difficulties in recalling positive experiences. To provide causal evidence for this conjecture, we conduct a well-powered online follow-up experiment with U.S. participants, which replicates the backfiring in a completely different context. Thus, our study provides evidence that standard self-affirmation interventions should be treated with some caution, as they may be ineffective and even counterproductive for those who could most need a boost in self-image and productivity. The unintended result notwithstanding, we provide clear evidence that productivity is malleable through psychological interventions.

Acknowledgements: We thank Alexander Cappelen, Johannes Haushofer, Frank Schilbach, Matthias Sutter, and Bertil Tungodden, as well as participants at the Ammersee Workshop, M-BEES in Maastricht, GfEW in Kassel, SEEDEC in Wageningen, ESA in Berlin, ASFEE in Nice, and seminar participants in Cologne, Erfurt, Munich, Bergen, and Bonn for their comments. The field research and study design was approved by the Namibian Ministry of Land Reform and was developed upon consultation with the Legal Assistance Center (LAC), the Desert Research Foundation of Namibia (DRFN), and social anthropologists with extensive experience in the region (especially Michael Bollig, Ute Dieckmann, and Jonathan Kempen). We are grateful for their tremendous help. We thank Jonathan Kempen and Barbara Scharfbillig for their outstanding research assistance in the field, Hanna Fuhrmann, Samuel Hartenberg, and Charlotte Krömer for interview codings, Appolia Dabe for interview translations and codings as well as our great team of local research assistants in Namibia (Johannes Abusema, Aaron Anton, Johannes Dam, Appolia Dabe, Markus Hendrik, Erastus Itembu, Martha Lubhaan, and Paulina Swaartbooi). IRB approval for the online experiment was obtained from the University of Cologne (230035BR). The online experiment was preregistered at the AEA registry (AEARCTR-0012142). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 39083886.

2.1. Introduction

A growing literature argues that poverty leads to mental burdens, such as stress, anxiety, and worry (see, for example, Haushofer and Fehr 2014; Ridley et al. 2020). One important source of mental burdens stems from the stigma of being poor, such as being viewed as lazy, unsuccessful, and unmotivated. Fighting these stereotypes and other mental burdens consumes cognitive resources that could be used more productively otherwise. In line with this, poverty and stigmatization have been shown to lead to lower performance in cognitive tests and reduced academic achievement (E. Dean et al. 2018; Haushofer and Fehr 2014; Mani et al. 2013; Steele 1988). Recent evidence shows that these detrimental effects on performance are also relevant for productivity in the labor market (Kaur et al. 2024). Naturally, penalties on productivity are particularly detrimental for those who are already poor. Hence, it is highly desirable to find interventions that can buffer the negative effects of mental burdens and thereby boost productivity.

In this paper, we test the potential of a widely used and promoted low-cost psychological intervention — self-affirmation — to reduce mental burdens. Self-affirmation theory posits that people are motivated to maintain a positive self-image (Aronson, Cohen, and Nail 2019; Sherman and Cohen 2006; Steele 1988). Typically, self-affirmation interventions aim to improve self-image by letting people write about personal values or recall events that made them proud. Affirming the self reduces mental burdens resulting from events that threaten self-image, such as fighting a stereotype of being lazy while working (Hall et al. 2014). Consequently, those freed-up mental resources can be used more productively. While several studies show that affirming the self can indeed boost cognitive performance, there is a lack of evidence on central economic outcomes such as labor productivity. We address this gap through a field experiment and test whether a self-affirmation intervention can increase productivity. To assess the generalizability of the field-experimental results, we furthermore conducted a follow-up online experiment in a different subject pool.

Self-affirmation interventions have shown positive effects across a range of domains for over more than three decades of research, such as in health, politics, inter-group conflict, and academic performance (for a recent overview, see Sherman et al. 2021). They are promoted in workplaces and schools worldwide, especially for disadvantaged groups (Cohen et al. 2006; 2009; Sherman et al. 2021; Walton 2014). Self-affirmation, therefore, seems a natural candidate for an intervention to improve productivity among those who strongly suffer from mental burdens.

7

To provide evidence from a group that suffers particularly strongly from mental burdens, we conducted a field experiment among 116 extremely poor and negatively stereotyped participants in rural villages in the Omaheke region of Namibia. The population of the Omaheke region suffers from high levels of absolute poverty and has limited access to educational and health facilities.¹ Our experimental sample mainly comprises ethnic groups of the Khoisan language family such as the "San" and "Damara", who are subject to strong negative stereotypes of social inferiority and laziness (Dieckmann et al. 2014).

We implemented a typical self-affirmation intervention, in which participants were asked to recall and report personal experiences that made them feel successful and proud. As in Hall et al. (2014), participants answered orally because of high illiteracy rates, and we recorded participants' answers via digital voice recorders. In the control condition, participants were simply asked to recall and report their daily routines (as in Hall et al. 2014, and Baranov, Haushofer, and Jang 2020).

Immediately afterwards, we offered all participants the opportunity to work for one hour on a typical agricultural task in this region: processing peanuts. This allows us to measure the effect of the self-affirmation intervention on labor productivity. At the same time, the task confronted participants with a domain in which the poor often have to deal with negative stereotypes: performance in the labor market. Participants were paid under a highly incentivized piece-rate scheme so that their productivity directly affected their earnings. Participants knew that the processed peanuts were sold to a local farmer afterwards. Our setting thus combined high degrees of experimental control and realism.²

We hypothesized – in line with the literature – that the self-affirmation intervention would make participants more productive in the task. Instead, we find that the self-affirmation intervention backfired: it *decreased* productivity by a magnitude of 15.3 percent or about 0.6 standard deviations. To understand this backfiring effect, we analyze the content of the recorded interviews. We find that many participants had difficulties reporting experiences of pride and success. Our intervention thus likely induced the opposite of its goal: The difficulty of recalling successful experiences further depleted cognitive resources and consequently impaired productivity. In line with this interpretation, we find correlational evidence that participants

¹ According to Namibia's official poverty mapping (Republic of Namibia National Planning Commission 2015), people are poor when they are not able to afford typical non-food consumption above a basic nutritional basket. Severely poor people cannot even afford this basic nutritional basket required for healthy survival. The most recently estimated poverty lines when we conducted our study were 2.67 \$PPP / day (poor) and 1.96 \$PPP / day (severely poor) per adult in 2010.

 $^{^2}$ Furthermore, our experimental design meets the three moderators identified in the literature for a self-affirmation intervention to produce strong effects: a psychological threat, an opportunity for the psychological effects to manifest themselves (in our case, the task), and timing the self-affirmation intervention right before the threat.

who had difficulties answering the self-affirmation question performed significantly worse than those for whom it was easier. We do not find a similar heterogeneous effect related to the difficulty of reporting daily routines in the control treatment.

Given the unexpected results, we pre-registered and conducted a well-powered followup online experiment in a different subject pool (recruited from the platform Prolific in the U.S.). The design of the online experiment parallels the field experiment. Participants are either assigned to a neutral control condition or to variations of the self-affirmation intervention. The self-affirmation treatments exogenously vary the difficulty of recalling stories of pride and success, following the "ease of retrieval" paradigm in social psychology (Schwarz et al. 1991; Schwarz and Vaughn 2002): participants have to recall and report either three (easy) or eight (difficult) experiences that made them feel successful and proud. These variations allow us to provide causal evidence for our conjecture that the backfiring can be explained by difficulties in recalling positive experiences. Afterwards, participants can work on a real-effort task with piece-rate incentives. The online experiment replicates the results from the field experiment. In line with the hypothesized mechanism, productivity in the difficult-recall condition is lower than in the easy-recall condition. Further, participants in the difficult-recall condition felt less successful and proud after the intervention.

Our paper makes three main contributions: First, self-affirmation interventions are widely promoted to improve psychological well-being, interpersonal relationships, health, and educational outcomes. We provide evidence that these interventions should be treated with some caution. In our setting, a standard self-affirmation protocol backfires among a group for whom a reduction in mental burdens and a boost in productivity could be strongly needed. This backfiring effect is replicated in a second study in a different subject pool. Our paper thus responds to the call to understand the "boundary conditions" of self-affirmation interventions (Cohen and Sherman 2014; Hall et al. 2014; Sherman et al. 2021).³ In order to achieve positive effects and minimize the risk of backfiring, self-affirmation protocols would likely need to be carefully adjusted to the target population to ensure that they allow people to manifest positive aspects of themselves.

³ In this regard, our paper is related to Baranov, Haushofer, and Jang (2020), who find no overall treatment effect of a combination of several positive-psychology interventions, including self-affirmation, on cognitive control, intertemporal choice, and a battery of psychological outcomes in Kenya. We add to this by isolating the effects of a self-affirmation intervention and by showing that positive-psychology interventions may not only be ineffective but can also backfire.

Second, we contribute to the literature on psychological interventions in low- and middle-income countries in general.⁴ While several studies underscore their benefits, there are instances where these psychological interventions yield no significant results (e.g., Haushofer, Mudida, and Shapiro 2020; Baranov et al. 2020; McKenzie, Mohpal, and Yang 2022). Our paper even documents a case where a positive-psychology intervention had an adverse effect, emphasizing that these types of interventions need to be designed very carefully.

Third, most previous literature studies the relationship between mental burdens and psychometric tests or laboratory measures of preferences and decision-making (see Kremer et al. (2019) for an overview). In our paper, we study the effects of alleviating such mental burdens on labor productivity. Labor productivity is of central interest to economics and policymakers alike; yet, it has hitherto received limited attention in the literature on the psychological consequences of poverty (Schilbach et al. 2016). Kaur et al. (2024) provide recent evidence that liquidity constraints, as a facet of poverty, may indeed reduce the capacity to focus at work. While we find the opposite of the desired effect, we provide clear evidence that productivity is malleable through psychological interventions.

2.2. Field Experiment

2.2.1. Experimental Design

Sample, Recruitment, and Experimental Set-up. Our field experiment was conducted in the Omaheke region of Namibia. The region has high levels of absolute poverty, high illiteracy rates, typically no regular income, and only limited access to health and educational facilities (Republic of Namibia Central Bureau of Statistics 2008). For example, 28.9 percent of the population in the region never attended school and only 23 percent completed secondary school. Employment opportunities in this region are typically short-term and constitute farm or domestic work. The study population mostly comprises indigenous ethnic groups who speak a Khoisan language and are socially marginalized, such as the "San" and "Damara". According to the multi-dimensional poverty index, households of this language group are almost

⁴ These interventions range from psychotherapeutic approaches such as cognitive behavioral therapy (Barker et al. 2022; Blattman, Jamison, and Sheridan 2017; Haushofer, Mudida, and Shapiro 2020), workshop sessions aiming to improve financial and entrepreneurial decision-making (Ghosal et al. 2022; Ashraf et al. 2024; McKenzie, Mohpal, and Yang 2022; Campos et al. 2017), to interventions with specific goals such as raising aspirations (Bernard et al. 2023; Orkin et al. 2023; Riley 2024) or visualizing and planning the future (John and Orkin 2022). The intensity of these interventions ranges from multi-week programs with trained personnel to one-shot interventions that can be implemented without training, for example, from repeated therapy sessions to showing short videos.

universally considered poor (93 percent), which is noticeably higher than the national average of 43 percent (Namibia Statistics Agency 2021). They are subject to strong negative stereotypes of social inferiority (Sullivan and Ganuses 2020; Suzman 2001), laziness, and alcoholism (Dieckmann et al. 2014). Our study population, therefore, likely experienced mental burdens in their daily lives, which may contribute to a perpetuation of poverty. In this setting, affirming the self may allow the participants to make better use of the scarce opportunities for economic progress.⁵

We conducted the experiment in September and October 2015. Given the low population density of the Omaheke region, we traveled to eight research locations spanning an area of about 14,000 square kilometers to recruit our study sample. Figures A1 and A2 in the Appendix provide maps of the research locations. To provide a real short-term work opportunity, we cooperated with a local farmer. The work comprised cracking and sorting peanuts. Local NGOs supported us in accessing public buildings, such as small community centers or schools in the villages, which allowed us to conduct our study in a controlled environment. We recruited participants by approaching them in their settlements and asking whether they would like to participate in our study later the same day at a certain time and place. There were no exclusion criteria for participation. The study was introduced as research of the University of Cologne in collaboration with our local partners about the livelihoods in the region. Participants were informed about the provision of snacks and soft drinks as well as a monetary compensation for their participation. Upon recruitment, participants completed a short questionnaire, which we mainly used to stratify the treatments by age.⁶

At the beginning of each session, participants were jointly welcomed and assigned to groups of four, with one local research assistant responsible for each group.⁷ Participants were placed in small cubicles made from cardboard to prevent them from observing others' performance and communicating with each other. All participants in a group received the same treatment, and each session comprised two groups per treatment. The research assistant called out each participant of the group one after another in random order and conducted the respective

⁵ The self-affirmation intervention is intended to benefit our participants. Since they are members of a marginalized group, we had to take special care in designing and conducting the experiment. We therefore closely developed our study in consultation with social anthropologists who had extensive experience in the region, as well as with two local NGOs, the Legal Assistance Center (LAC) and the Desert Research Foundation of Namibia (DRFN). Both NGOs strive to improve the livelihoods of the most disadvantaged groups in Namibia, for example, through education, research, legal advice, and representation. Furthermore, the field research and study design were approved by the relevant ministry for the study region, the Namibian Ministry of Land Reform. The experimental task was adapted to participants' capabilities, interaction was carried out through locals and in the local environment. Participation was voluntary and under informed consent, and with opt-out possibilities at any time.

⁶ Age seemed to correlate with productivity in pilot sessions. For details of the pre-intervention questionnaire, see Appendices D and E.

⁷ Some groups only had three participants because of a lower show-up rate in that location.

interview one-on-one in private. After all interviews were finished, the job opportunity was announced, peanuts were distributed, and participants cracked and sorted peanuts for 60 minutes. Participants were informed that communication between participants was prohibited so that all participants could focus on their task. As explained to the participants before the task, after the 60 minutes, the research assistants weighed the peanuts using precision scales and paid participants according to their output. All processed nuts were sold to the farmer after the experiment.

The experimental instructions were given in Afrikaans, a language that is widespread in the local population, although not the native language of most of our participants. This mimics a typical working environment in the region, as Afrikaans is spoken by most farm owners, who are the main employers and offer similar short-term jobs as in our experiment.⁸ The experimental instructions can be found in Appendix E.

Treatments. Our experimental design comprises two treatments: self-affirmation and control. In both treatments, local research assistants conducted a personal one-on-one interview with each participant. The treatments only differed with respect to the interview question. In the self-affirmation treatment, participants were asked to describe an experience that made them feel successful and proud. This form of self-affirmation is based on Hall et al. (2014) and was also used (in a slightly different form) by Klein, Blier, and Janze (2001), as well as by Larsen et al. (2023). In the control treatment, participants were asked to describe what they typically did on weekdays (as in Hall et al. 2014, and Baranov et al. 2020).⁹

The research assistants were instructed to conduct interviews of about two minutes in length. They were instructed to ask follow-up questions if necessary to keep the conversation going. Due to high illiteracy rates in the population, the interviews were conducted verbally and recorded using digital voice recorders. We initially planned to let the participants record their responses on voice recorders on their own as in Hall et al. (2014). However, participants showed high levels of unfamiliarity with voice recorders during pilot sessions and hardly spoke at all. Participants were more talkative when research assistants operated the voice recorders and guided the interviews.

⁸ The research assistants were also able to conduct the interviews and questionnaires in the local languages Khoekhoegowab and Ju/'hoansi in case participants had problems understanding Afrikaans.

⁹ We also conducted a third treatment that is not analyzed in this paper. These interviews asked participants to "describe what the government or other organizations should do for [them]."

Randomization. 116 participants in eight experimental locations took part in the two treatments, 58 in self-affirmation and 58 in control. This sample size is sufficient to detect a moderate effect size (Cohen's d = 0.53) at conventional levels of power (80%) and two-sided alpha (0.05) with a two-sample t-test. Moderate effect sizes (Cohen's d between 0.45 and 0.72) of a self-affirmation intervention were also found by Hall et al. (2014). We stratify treatments by experimental location and age, i.e., within each of the eight research locations the same fractions of participants were assigned to the self-affirmation and the control treatments. In addition, we randomized the assignments of research assistants to treatment conditions, which avoids research-assistant fixed effects confounding our results.

	(1)	(2)	(3)	(4)
	Full sample	Self-affirmation	Control	p-value
Age	35.58	35.90	35.26	0.763
8	(13.64)	(13.73)	(13.65)	
Years of schooling	4.940	5.345	4.534	0.229
6	(3.707)	(3.492)	(3.899)	
Male	0.414	0.397	0.431	0.706
No work	0.647	0.569	0.724	0.080
Housing: Corrugated iron	0.595	0.621	0.569	0.860
No means of transport	0.836	0.914	0.759	0.233
No toilet at home	0.612	0.552	0.672	0.474
Observations	116	58	58	

 TABLE 1: BALANCE TABLE

Note: The table presents means and standard deviations (when meaningful) in parentheses. We report modal characteristics of housing (Corrugated iron), type of toilet (No toilet) and transportation (No means of transportation). We test for treatment differences using two-sided Wilcoxon-Mann-Whitney test for Age and Years of schooling, which are interval-scaled. We use two-sided Chi-square tests for the remainder of the variables (ordinally scaled categories). Note that the tests for statistical differences between treatments reported in column (4) include all categories of housing (no housing, plastic bags and natural materials, corrugated iron, brick house, brick house with more than three rooms), type of toilet (none, dry toilet or long-drop toilet, flush-by-hand toilet, flush toilet) and mode of transportation (no means of transportation, motorbike, donkey cart, other), instead of the displayed modal characteristic.

Table 1 summarizes the most important sample characteristics. The average participant is 36 years old and has completed five years of schooling. The majority of participants has no work, lives in houses built of corrugated iron, does not own any means of transportation (bicycle, donkey, etc.) and has no access to either flush or dry toilets. Participants stem mainly from two

negatively-stereotyped Khoisan language groups in Namibia: Ju/'hoansi and Khoekhoegowab.¹⁰

Main Outcome: Labor Productivity. After the interviews, participants were offered a job opportunity for a local farmer, who was interested in purchasing shelled peanuts. The processing of agricultural products such as beans and nuts constitutes typical farm work in the region. Participants' task was to remove the shells and sort the peanuts with a piece-rate incentive for 60 minutes. They removed the shells and sorted the nuts into boxes for undamaged peanuts and damaged peanuts. Participants were only paid for undamaged peanuts. Measuring the weight of damaged peanuts separately allows us to assess the quality of work. Participants were paid a piece-rate of 10 Namibian Dollars (N\$) for every 100 grams of undamaged peanuts (about 0.60 EURO at the time).¹¹ The average hourly wage for agricultural labor in the region is about 8 N\$ based on local information. Participants were hence well-paid and the task highly incentivized. Take-up of the work task was indeed 100 percent in both treatments. Participants earned on average 20.42 N\$ in the work task plus a flat compensation of 10 N\$ for participating in the study. Additionally, all participants received a soft drink and a packet of snacks upon arrival as an additional compensation for participating.

2.2.2. Conceptual Framework

In order to think about possible treatment effects more precisely, we borrow a simple framework from Kaur et al. (2024). Individuals gain utility from the output they produce. For simplicity, and without loss of generality, we model only one period such that we can abstract from savings and debt. We consider output to be an increasing function of two inputs: an effortful and consciously chosen input e and an automatic input a. While e is under control of the individual (e.g., how hard she tries to work), a depends on the available cognitive resources and cannot be directly chosen by the individual. Most relevant for our set-up, automatic input captures the capacity to pay attention (e.g., how to carefully peel peanuts without breaking them) and to stay focused on the task by inhibiting distracting thoughts or stimuli. Automatic input a can be considered as the effectiveness factor of effort, i.e., how well consciously chosen effort translates into actual output.

¹⁰ About 26 percent of our sample indicated that they spoke the Khoisan language Ju/'hoansi, a clear marker of San ethnicity. The majority of 60 percent indicated that they spoke Khoekhoegowab, a Khoisan language primarily associated with the Damara ethnicity, which is widespread in the area and often used as a common language.

Consequently, for a given level of effort, increasing the available cognitive resources *a* directly increases output (e.g., by reducing attentional errors or making less unintentional breaks because of a loss of focus). Further, under standard assumptions on the functional form (concave production function with complementarity in inputs, convex effort costs, and separability), it follows straightforwardly that an increase in *a* also increases output indirectly by increasing optimal effort levels (see Appendix B for a more detailed exposition).

Crucially, we assume mental burdens to reduce a. As individuals cannot consciously change a and may not even be aware of its importance (as suggested by J. Dean (2024), increasing available cognitive resources must be the result of an external factor, such as a policy intervention.

The novel part of our experiment is to test one such intervention: self-affirmation. Based on self-affirmation theory and the plethora of previous experimental results cited earlier, affirming the self can reduce those mental burdens that are due to "ego-defenses": efforts to maintain a positive self-image when the self-image is threatened by exposure to stereotypes or stressful events. By affirming the self prior to an otherwise threatening event, the need to maintain a positive self-image when exposed to the threat is lowered, which frees up mental resources. Based on the results by Hall et al. (2014), self-affirmation may thereby increase what E. Dean et al. (2018) describe as inhibitory control, i.e., the ability to control impulses and block out distractions, sometimes used interchangeably with self-control. In terms of our experimental task, self-affirmation should improve participants' ability to stay focused on the tedious task.

In the context of the simple theoretical framework, the effects of self-affirmation can be modeled as an increase in *a*, which consequently boosts labor productivity and raises output. This effect is conditional on the ex-ante presence of mental burdens from ego-defenses. It is highly likely that the extremely poor, unemployed and negatively-stereotyped participants in our sample suffer from such mental burdens in their daily lives and in particular in the domain of labor markets. This leads to our main hypothesis: *Participants in the self-affirmation treatment are more productive than participants in the control treatment*.

2.3. Results of the Field Experiment

2.3.1. Treatment Effects

In both experimental conditions, all participants worked on the job and none of them chose to drop out. Mean productivity in the control group is 243.02 grams of undamaged peanuts (SD = 60.81), whereas mean productivity in the self-affirmation treatment is only 204.72 grams (SD = 58.50). The negative treatment effect of 38.3 grams corresponds to a standardized effect size (Cohen's *d*) of 0.64 and is highly significant (both using a t-test and a non-parametric Mann-Whitney-U test: p < 0.001 and p = 0.001).

TABLE 2. TREATMENT EFFECTS ON PRODUCTIVITY			
	(1)	(2)	
	Nuts in grams	Nuts in grams	
Self-affirmation treatment	-38.281***	-37.069***	
	(12.042)	(12.797)	
Has no mode of transport		-8.762	
		(16.569)	
Years of schooling		-0.201	
		(1.916)	
Male		1.038	
		(14.673)	
Age		-0.667	
		(0.559)	
Has a job		6.835	
		(13.778)	
Session Fixed Effects	Yes	Yes	
Interviewer Fixed Effects	Yes	Yes	
Observations	116	116	
R-squared	0.289	0.305	

Notes: This table shows treatment effects on productivity. The outcome variable is processed undamaged nuts in grams. We provide robust standard errors in parentheses. *** p < 0.01.

We now turn to regression statistics, which allow us to control for covariates. Both regressions presented in Table 2 control for interviewer and research location fixed effects. In Column (2), we also add a range of sociodemographic controls. Column (2) of Table 2 suggests that the self-

affirmation treatment reduces the number of processed undamaged nuts by about 15.3 percent (p = 0.005).¹²

Main Result Field Experiment: *The self-affirmation intervention reduced productivity by about 15.3 percent compared to the control group.*

2.3.2. Exploring the Backfiring: Difficulties in Reporting Successful Events

Our experimental findings do not match our hypothesis. An explanation for the negative effect of our self-affirmation is that the intervention may have actually failed to work as planned because many participants had difficulties sharing experiences of success and pride and, in many instances, even reported negative stories (see in more detail below). Therefore, the intervention may have *disaffirmed* participants' self. The increased mental burdens then decreased productivity, in line with the mechanism underlying the hypothesis. This explanation links to a strand in the psychological literature showing that the ease of recalling positive experiences affects positive self-assessments (Schwarz et al. 1991; an overview is provided in Schwarz and Vaughn 2002).¹³

To explore the plausibility of this interpretation, we measured participants' ease or difficulty in reporting experiences that made them feel successful and proud ("Ease"). Three research assistants who were unaware of the research question and hypothesis independently coded answers for both treatments with respect to the question "how easy was it for the interviewed person to answer the questions concerning success stories (daily routines)" on a ten-point scale. Another research assistant counted the number of words that participants spoke during the interviews as well as the number of follow-up questions to keep the conversation going.

We find that participants are less at ease when answering the self-affirmation question than they are when answering the control question (5.6 vs 6.2; p = 0.019, t-test; see Table A2 in the Appendix). Interviewers also ask more follow-up questions in the self-affirmation treatment to keep the interview going (3.9 vs 2.8, p = 0.002), which results in a similar word count across both conditions (p = 0.524). This suggests that participants in the self-affirmation treatment struggled to fill the time allocated for the interview. Moreover, despite explicitly

¹² The results remain robust when clustering standard errors at the group level to account for potential dependency within groups (see Table A1).

¹³ Schwarz et al. (1991) find that participants who had to recall twelve examples of assertive behavior (difficult) perceived themselves as significantly less assertive than participants who had to recall only six examples (easy).

being asked to recall successful stories, about 40 percent of participants in the self-affirmation treatment talked about instances that research assistants coded as negative experiences.¹⁴ We hence conjecture that the backfiring effect of the self-affirmation intervention may be due to difficulties in reporting successful events, which disaffirmed participants and increased their mental burdens.

To further explore the potential role of the difficulty of answering the interview question, we run regressions that interact Ease with the treatment. While Ease does not cleanly capture the same construct in the two conditions, it nevertheless entails a measurement of fluency in and comfort with the interview situation. Lower Ease is only associated with lower productivity in the self-affirmation treatment, but not in the control group (Table A4 in the Appendix). In other words, differences in interview fluency per se do not correlate with productivity, while difficulties in coming up with positive experiences in the self-affirmation treatment do correlate with productivity. We can also use education as an exogenous proxy for the ease of telling success stories. While having above-median education strongly correlates with the Ease codings in the self-affirmation treatment (p = 0.010, t-test), it does not significantly correlate with productivity in the self-affirmation treatment but not in the control condition (Table A5). These analyses provide correlational evidence that the overall negative treatment effect may be related to the mental burdens that are caused by the difficulties in reporting stories of success and pride.

To investigate whether lower productivity is driven by "attentional lapses" (as in Kaur et al. 2024), we analyze the quality of the work provided. We evaluate the fraction of broken peanuts among all processed peanuts. Recall that participants were only paid for undamaged nuts, so being careless while opening peanuts has direct consequences on productivity and earnings. Table A6 shows that the fraction of broken nuts is statistically insignificantly different between treatments. Hence, the decrease in productivity in our self-affirmation treatment is not driven by more mistakes. Our finding suggests that mental burdens can also influence productivity in different ways than clear "attentional lapses", such as a less effective or slower work pace caused by a lack of self-control to stay focused on the task (E. Dean et al. 2018).

Finally, we ask whether our findings can be explained by different mechanisms. In Appendix C, we examine two potential channels and argue that these are unlikely to drive the

¹⁴ An overview of topics covered during the interviews can be found in Tables A3.1 and A3.2 in the Appendix. Participants talked about different topics across treatments. For instance, housework and subsistence work are the overwhelming topics in the control condition, whereas participants in the self-affirmation treatment mainly talked about education, hobbies, family, and employment.

treatment effects. First, we argue that the self-affirmation treatment did not cause perceptions of being overqualified for the peanut-cracking job. As a second explanation, we discuss whether the control treatment was not neutral and show that this explanation is inconsistent with our data.

2.4. Online Experiment

Given the unexpected findings, one may wonder about the generalizability of the results beyond the field setting in Namibia. Furthermore, our conjecture to explain the backfiring effect relies on a correlational analysis based on the coding of the interview content. To address these two concerns, we conducted a pre-registered (AEARCTR-0012142) online experiment with 1,848 U.S. participants, who were either assigned to a neutral control condition or to variations of the self-affirmation intervention. The self-affirmation treatments exogenously vary the difficulty of recalling stories of pride and success in order to test the conjectured mechanism directly. After the treatment manipulation, participants are given the opportunity to work on a standard real-effort task with piece-rate incentives.

2.4.1. Experimental Design

Treatments and Task. At the beginning of the experiment, participants were informed that there were two stages of the study. In the first stage we elicited baseline self-efficacy and asked participants to write about personal experiences (treatment manipulations), and the second stage offered the possibility of earning additional money (real-effort task). In our treatment manipulations, we ask participants to recall either three (treatment "Easy") or eight (treatment "Difficult") experiences that made them feel successful and proud. This manipulation follows a standard paradigm in psychology on the effects of ease of retrieval (Schwarz et al. 1991; Schwarz and Vaughn 2002) and is calibrated based on pilots.¹⁵ A control condition asks participants to "describe [their] daily routine on a typical day of the week", similar to our control treatment in the field experiment ("Control"). Participants are asked to spend five minutes in total to describe their experiences.

In the real-effort task, participants were instructed to count the number of zeros in randomly generated 3x10 matrices of zeros and ones for ten minutes (as used in, e.g., Abeler et

¹⁵ In a pilot session without being prompted to write a particular number of stories, most participants were able to recall three stories, whereas only few could recall eight.

al. 2011). Each correctly solved task earned participants a bonus payment of 0.05 USD.¹⁶ Participants were informed that aborting the study implied that neither their show-up fee nor a bonus on the second part of the experiment would be earned. Participants could press a skip button to skip either part of the study without penalty.

In a final questionnaire after the task, we included psychological outcomes as manipulation checks. They capture the two main dimensions that are intended to be manipulated: ease of recall ("It was easy / challenging to recall and describe all experiences") and, consequently, feeling successful and proud ("When recalling the experiences, I felt very successful / I did not feel proud at all"). We hypothesized that ease of recall positively influences feeling successful and proud and, thereby, productivity.

Sample and Randomization. We implemented our online experiment in Qualtrics and recruited U.S. participants on Prolific. As self-affirmation interventions are supposed to provide benefits especially to disadvantaged groups, we targeted participants with a self-reported socioeconomic status below the scale median (i.e., 1-5 on a scale from 1-10, with 10 being the highest position) and with an after-tax household income approximately in the lowest quintile.¹⁷ Randomization is stratified by sex. Table A7 summarizes the sample characteristics and shows that the treatments were balanced. Participants were paid a show-up fee of 2.07 USD and earned an average bonus payment from working on the real-effort task of 1.83 USD. In line with Prolific's rules and our experimental instructions, only participants who finished the experiment were paid.

2.4.2. Main Results of the Online Experiment

Table 3 presents our main results. As pre-registered, we focus our analysis on all participants who saw the treatment screen (ITT sample), i.e., who were asked to either recall experiences or describe their daily routines.¹⁸ Participants were around 18 percent less productive (exact value depending on the specification) in Difficult than in the Control (p < 0.001). The negative effect is of similar size as in the field experiment and robust to including demographic control

¹⁶ The payment was calibrated based on pilot studies such that participants should earn a substantial bonus on average. Payments on Prolific are calculated in GBP. As our sample only consists of U.S. participants, we only report values converted to USD with the exchange rate at the time.

¹⁷We initially recruited participants with an after-tax household income of less than 40,000 USD. In line with our pre-registration, we slightly increased the income threshold to reach the desired sample size. In our final questionnaire, 84 percent of participants report an income below 50,000 USD.

¹⁸ Note that our main analysis focuses on the two self-affirmation treatments Easy and Difficult and on the Control treatment (n = 1,109). Two further exploratory treatments (n = 739) are discussed in Section 4.3 below.

variables and to different ways to deal with outliers: top-coding the outcome variable at the 99th percentile and running a robust regression with MM-estimator (based on Yohai (1987) and as recommended by Wilcox 2017).¹⁹ Our online experiment hence replicates the findings from the field experiment in Namibia. In line with our pre-registered hypotheses, we also find significant differences in productivity between Easy and Difficult of around 12 percent (depending on the exact specification, all p < 0.05), i.e., participants who are asked for eight success stories are less productive than participants who are asked for only three stories. Somewhat disappointingly, self-affirmation does not lead to positive effects on productivity, even in the Easy treatment. If at all, the effects are negative in Columns (2) and (3).

	(1)	(2)	(3)
	Correct tasks	Correct tasks	Correct tasks
Easy	-1.693	-2.849*	-3.516*
	(2.332)	(1.678)	(1.954)
Difficult	-6.347***	-6.361***	-7.567***
	(1.730)	(1.716)	(2.071)
H ₀ (Easy=Difficult): p-value	0.048	0.039	0.045
Comparison mean	36.463	36.461	36.696
Sample	ITT	ITT	ITT
Controls	NO	YES	YES
Outlier correction	NO	Winsorized	Robust Reg.
Observations	1,109	1,109	1,109
(Pseudo) R-squared	0.008	0.029	0.029

TABLE 3. TREATMENT EFFECTS IN COMPARISON TO CONTROL (ONLINE)

Notes: Treatment coefficients in comparison to control group. Column (2): Outcome variable topcoded at the 99th percentile. Column (3): Robust regression with MM-estimator and 85%-efficiency. Controls include all variables that are available for the ITT sample: sex, self-efficacy, number of previous studies on Prolific and whether the attention check was passed. Robust standard errors in parentheses. *** p<0.01, * p<0.1.

To further substantiate the mechanisms underlying our findings, we analyze the impact of our interventions on key psychological outcomes in Table 4.²⁰

¹⁹ A few participants seemingly did not count zeros in the real-effort task but entered the expected value very rapidly each time (for each new randomly generated 3x10 table, the expected value of zeros is 15). This resulted in extreme values of attempts and mistakes. We use two ways to account for them: top-coding (winsorizing) at the 99th percentile as well as robust regressions with MM-estimator in the recommended default setting of the ROBREG package (Jann 2021), which combines a high breakdown point and efficiency. We do not include a log-transformation to account for outliers due to the presence of zeros (around 10.5%) in the main outcome variable. Accounting for extreme outliers explains the change in the estimated coefficient of Easy from (1) to (2) and (3).

²⁰ The questionnaire including the psychological outcomes was administered after the real-effort task at the end of the experiment. Hence, we have data on these measures only for those participants who completed the experiment (i.e. 1,014 out of the 1,109 participants in the ITT sample). Note that we also observe the backfiring of Difficult

	(1)	(2)	(3)	(4)	(5)	(6)
	Word count	Number of	Recall easy	Recall	Feeling	Feeling not
		stories	-	challenging	successful	proud
						-
Difficult	70.762***	3.087***	-0.960***	0.967***	-0.348***	0.323***
	(9.077)	(0.151)	(0.128)	(0.145)	(0.107)	(0.112)
Comp. mean	117.584	2.654	5.079	3.238	5.504	2.191
Observations	660	660	660	660	660	660
R-squared	0.100	0.411	0.244	0.164	0.138	0.124
Sample	Finisher	Finisher	Finisher	Finisher	Finisher	Finisher
Controls	YES	YES	YES	YES	YES	YES

TABLE 4. MANIPULATION CHECK: EASY VS. DIFFICULT (ONLINE)

Notes: OLS estimates in comparison to the Easy treatment. Robust standard errors in parentheses. The outcome variables of Columns (3) to (6) are measured on a 7-point Likert scale. Note that due to the timing of the elicitation, we only have data on the psychological measures for participants who finished the experiment. Controls include age, income, socioeconomic status, sex, self-efficacy, number of previous studies on Prolific and whether the attention check was passed. *** p<0.01.

As hypothesized, participants responded differently to the treatments Easy and Difficult: Participants wrote more words and more stories in Difficult, in which they were asked to recall eight as compared to three stories (Columns 1 and 2, all p < 0.001, n = 660). Yet, participants perceived recalling stories as harder and more challenging in Difficult (Columns 3 and 4, all p < 0.001). As a consequence, participants in Difficult felt significantly less successful and proud after the intervention (Columns 5 and 6, all p < 0.01). These results support our conjecture that difficulties in recalling successful events disaffirm participants, increase mental burdens and consequently reduce productivity.

Main Results Online Experiment. The backfiring of a self-affirmation intervention can be replicated in a U.S. subject pool. Exogenously increasing the difficulty of recalling positive experiences disaffirms participants and decreases their productivity.

A potential concern is that participants in Difficult are less productive because they are mentally exhausted from writing more words compared to their counterparts in Easy and Control (163.0 vs 109.0 and 111.2 words, respectively, p < 0.001). However, we find that participants who wrote more were significantly more productive in all conditions. Controlling for the number of words written consequently increases the estimated treatment differences between Difficult and both Control and Easy (Table A9). This suggests that writing more did not lead to mental

compared to Control when excluding those who dropped out and when controlling for the additional sociodemographic variables that were also collected at the end of the experiment (Table A8).

exhaustion and cannot explain the backfiring effect. If anything, the analysis suggests that the backfiring of Difficult might have been even more pronounced if participants had written the same number of words in Difficult as they did in Easy or Control.

2.4.3. Additional Treatments: Timer

In an additional exploratory variation of the self-affirmation treatments, we furthermore tested a second explanation through which the intervention in the field may have disaffirmed participants: struggling to fill the time allocated to share success stories. The higher number of follow-up questions in the interviews of the field experiment suggests that participants in the self-affirmation treatment struggled to fill the time allocated to talk about positive experiences. To test this possible mechanism, we ran two variations of Easy and Difficult introducing a five-minute timer that is displayed on the screen (treatments "Easy+Timer" and "Difficult+Timer"). The timer automatically ends the writing stage after five minutes. We hypothesized that in Easy+Timer, participants may struggle to write enough to fill the available time, which is made salient through the timer. Contrarily, in Difficult+Timer, the five-minute timer more likely interrupts participants from writing, suggesting that they could have recalled more stories, resulting in a positive affirmation.²¹

Table A10 presents the regression results for the treatments with a timer. While the direction of the treatment differences between Easy and Difficult reverses when introducing a timer (as hypothesized), those differences are not statistically significant.²² The results suggest that ease of retrieval is a better explanation for the observed backfiring effect in the field than the failure to fill the allocated time.

On a positive note, the backfiring of the self-affirmation intervention disappears when adding a salient timer and the coefficient of Difficult+Timer turns positive (but not statistically significant). Consequently, productivity is higher in Difficult+Timer than in Difficult (p < 0.01 in all regressions in Table A12). These effects highlight that seemingly subtle procedural details of the self-affirmation intervention matter.

²¹ Alternatively, however, it is also conceivable that not being able to think of successful experiences quickly enough, i.e., failing to complete the task in the given time, is perceived as having difficulties in reporting successful events, which can disaffirm participants and result in increased mental burdens in Difficult+Timer.

²² We observe a marginally significant difference between Easy+Timer and Difficult+Timer when excluding those who dropped out before the end of the experiment (Table A11). This might be driven by the fact that, as hypothesized, participants in the Difficult+Timer are more often interrupted by the timer than in the Easy+Timer treatment (56.5 percent vs. 26.4 percent of participants in the Finisher sample, p < 0.001).

2.5. Conclusion and Discussion

Low-cost psychological interventions that explicitly address the mental burdens of the poor hold great promise (Haushofer and Fehr 2014; Ridley et al. 2020; Haushofer and Salicath 2023). Our study shows that interventions that aim to alleviate mental burdens should be treated with caution. In a field experiment, we find that a self-affirmation intervention backfired among those who could need a boost in labor productivity most. It is therefore imperative to understand the conditions under which negative results may occur – rather than merely null effects, as found in a small number of previously published studies (see the discussion in Sherman et al. (2021) and the critical re-assessment by Serra-Garcia, Hansen, and Gneezy (2020) of the highly-cited study by Miyake et al. (2010)). Our analysis suggests that the backfiring is rooted in the failure of participants to report what the intervention asked them for: experiences that made them feel successful and proud. It is therefore worth re-emphasizing an important point made by some of the founders of self-affirmation theory: "Self-affirmation works not by giving people something that they lack but by allowing people to manifest what they already have [...]" (Sherman et al. 2021, p.67). Our results suggest that prompting people to manifest what they do *not* have can even backfire.

The population from which we recruited our participants may be particularly prone to lack experiences of success and pride, given the structural poverty in the region and a long history of racial discrimination under South Africa's Apartheid regime. In order to test whether the backfiring mechanism extends to other contexts, we run an online experiment in a U.S. subject pool, in which we manipulate the difficulty of recalling stories of success and pride. We indeed replicate the backfiring effect of self-affirmation interventions on productivity in a real-effort task. The negative effect tends to be stronger in the treatment with higher difficulty of recall. While none of the self-affirmation treatments tested produces positive effects compared to a control group, seemingly subtle manipulations (introducing a salient time limit) can – in line with the pre-registered hypotheses – stop the backfiring effect. This again highlights that this type of intervention needs to be designed very carefully and with thorough attention to detail.

On the flipside, our results provide evidence that labor productivity is malleable by changing mental burdens through low-cost psychological interventions. More research is needed to test the productivity-increasing potential of other delivery techniques of self-affirmation – such as asking people to reflect on their core values – or positive-psychology interventions more generally.

Appendix A: Additional Figures and Tables



FIGURE A1: LOCATION OF THE OMAHEKE REGION (RED FRAME). Source: Google Earth / Google Maps.



FIGURE A2: RESEARCH SITES (RED STARS) IN THE OMAHEKE REGION. Source: Google Earth / Google Maps.

CLUSTERED S.E. AT OROUP LEVEL			
	(1)	(2)	
	Nuts in grams	Nuts in grams	
Self-affirmation treatment	-38.281**	-37.069**	
	(15.398)	(15.789)	
Has no mode of transport		-8.762	
		(15.799)	
Years of schooling		-0.201	
		(1.652)	
Male		1.038	
		(14.521)	
Age		-0.667	
-		(0.560)	
Has a job		6.835	
		(11.772)	
Session Fixed Effects	Yes	Yes	
Interviewer Fixed Effects	Yes	Yes	
Number of clusters	32	32	
Observations	116	116	
R-squared	0.289	0.305	

TABLE A1. TREATMENT EFFECTS ON PRODUCTIVITY	Y
CLUSTERED S.E. AT GROUP LEVEL	

Notes: This table shows treatment effects on productivity. The outcome variable is processed undamaged nuts in grams. Robust standard errors in parentheses are clustered at the group level. ** p < 0.05.

TABLE A2. INTERVIEW ANALYSIS				
	(1)	(2)	(3)	
	Ease (1-10)	Number of follow-up questions	Word count	
Control	6.236 (0.967)	2.759 (1.720)	83.379 (33.540)	
Self-affirmation (SA)	5.600 (1.744)	3.860 (1.917)	87.920 (40.192)	
H_0 (SA = Control): p-value	0.019	0.002	0.524	

Notes: Three interviews could not be coded due to technical errors of the voice recorders. Five interviews are excluded because participants did not respond to the interview questions. Excluding unresponsive participants from our analysis does not change our main results. We report standard deviations in brackets and p-values from two-sided t-tests.

	Share of
	participants
No. of positive stories > 0	1
Education	0.34
Sports / hobbies / competition	0.34
Children	0.34
Job	0.52
Property	0.28
Good life in general	0.06
Good social relations	0.26
Helping others	0.20
Receiving help	0.16
Religiosity / God's help	0.10
Other positive	0.14
No. of negative stories > 0	0.40
Education negative	0.08
Sports / hobbies / competition negative	0.02
Children negative	0.06
Job negative	0.20
Property negative	0.14
Bad life in general	0.04
Bad social relations	0.12
General complaints	0.16
Religiosity / No help from God	0
Other negative	0.02
Observations	50

TABLE A3.1. INTERVIEW CONTENTSELF-AFFIRMATION TREATMENT

Notes: Two research assistants independently coded the interview content into pre-defined categories. We defined meaningful categories based on reading all transcribed interviews, trying to make a largely comprehensive list. Assigning one interview to multiple categories was allowed. We treat an interview as belonging to a category if at least one of the two coders assigned its content to that category.

	Share of
	participants
Housework	0.862
Subsistence work	0.707
Procurements (e.g., collecting	0.414
firewood, water)	
Employment	0.310
Self-employment	0.207
Job search	0.121
Unemployment	0.241
Children	0.397
Social relations	0.379
Helping others	0.224
Education (own)	0.069
Leisure time	0.259
Church	0.207
Problems in general	0.086
Lethargy in general	0.138
Other	0
Observations	58

TABLE A3.2. INTERVIEW CONTENTCONTROL TREATMENT

Notes: Two research assistants independently coded the interview content into pre-defined categories. We defined meaningful categories based on reading all transcribed interviews, trying to make a largely comprehensive list. Assigning one interview to multiple categories was allowed. We treat an interview as belonging to a category if at least one of the two coders assigned its content to that category.
	(1)	(2)
	Nuts in grams	Nuts in grams
Self-affirmation treatment (SA)	-48.951***	-49.224***
()	(11.795)	(12.021)
Ease	-7.415	-10.121
	(10.092)	(9.901)
SA * Ease	18.223*	20.239*
	(10.555)	(10.572)
No means of transport		5 731
No means of transport		(12,772)
Veers of schooling		(13.772)
rears of schooling		-1.770
Mala		(1.800)
Iviale		-9.000
A		(12.207)
Age		-1.102^{+1}
Has a lab		(0.409)
nas a job		(12,259)
Lin our Combinations		(12.238)
Linear Combinations		
Ease + SA * Ease	10.807**	10.118**
	(4.619)	(4.509)
	()	(
Session Fixed Effects	Yes	Yes
Interviewer Fixed Effects	Yes	Yes
Observations	108	108
R-squared	0.402	0.447

TABLE A4 HETEROGENEOUS	TREATMENT FEFECTS	S BV FASE OF ANSWERING
IADLE AT, HEILKOOLNLOUS	INLAIMENT LITECT	DI LASL OI MINSWERING

Notes: This table shows treatment effects on productivity. The outcome variable is processed undamaged nuts in grams. Robust standard errors are shown in parentheses. Ease is mean-centered by coder and averaged over the three coders. * p<0.1; ** p<0.05; *** p<0.01.

	(1)	(2)
	Nuts in grams	Nuts in grams
Control	53.260***	52.341***
	(18.401)	(19.520)
Above median education	26.374*	24.770*
	(14.553)	(14.296)
Control * Above median education	-30.667	-28.849
	(22.514)	(22.903)
Has no mode of transport		-5.808
Male		(17.123) 2.677
Age		(14.663) -0.511
Has a job		(0.527) 9.919
		(13.673)
Session Fixed Effects	Yes	Yes
Interviewer Fixed Effects	Yes	Yes
Observations	116	116
R-squared	0.308	0.321

|--|

Notes: This table shows treatment effects on productivity. The outcome variable is processed undamaged nuts in grams. We provide robust standard errors in parentheses. * p<0.1 ** p<0.05 *** p<0.01.

I ABLE AO. I REAL	MENT EFFECTS ON QUAL	11 Y
	(1)	(2)
	Fraction of broken	Fraction of broken
	peanuts	peanuts
Self-affirmation treatment	0.003	0.003
	(0.004)	(0.005)
Has no mode of transport		-0.002
1		(0.005)
Years of schooling		0.000
C C		(0.001)
Male		0.004
		(0.004)
Age		-0.000
		(0.000)
Has a job		0.002
		(0.005)
Session Fixed Effects	Yes	Yes
Interviewer Fixed Effects	Yes	Yes
Observations	116	116
R-squared	0.213	0.243

TABLE A6. TREATMENT EFFECTS ON QUALITY

Notes: This table shows treatment effects on the fraction of unbroken nuts among all processed peanuts (broken and unbroken). We provide robust standard errors in parentheses.

TABLE A7. BALANCE TABLE (ONLINE)							
	(1) Full	(2) Easy	(3) Difficult	(4) E+T	(5) D+T	(6) Control	(7) p-value
Sample: ITT							
Male	0.503	0.503	0.503	0.501	0.503	0.504	1.000
Self-efficacy, baseline (scale 1-7)	5.085 (1.222)	5.111 (1.229)	5.075 (1.218)	5.084 (1.184)	5.048 (1.292)	5.104 (1.187)	0.963
Number of previous studies on Prolific	1,329.0 (1,457.1)	1,458.1 (1,542.1)	1,297.0 (1,410.1)	1,359.1 (1,520.1)	1,271.2 (1,424.2)	1,260.1 (1,381.2)	0.351
Attention check: failed	0.028 (0.164)	0.027 (0.162)	0.022 (0.146)	0.033 (0.178)	0.030 (0.170)	0.027 (0.163)	0.915
Observations	1,848	370	370	367	372	369	
Sample: Finisher							
Age	37.66 (13.77)	38.75 (14.12)	36.45 (13.62)	38.78 (14.29)	36.90 (13.01)	37.31 (13.67)	0.092
Net household income (USD)	34,647.7 (21,677.4)	33,137.8 (21,631.4)	36,332.3 (23,335.0)	34,856.7 (20,245.8)	33,735.3 (22,426.9)	35,254.2 (20,769.7)	0.377
Net household income per capita (USD)	17,728.8 (13,757.7)	17,231.6 (12,657.4)	17,506.4 (14,435.6)	17,586.2 (12,941.0)	17,436.3 (13,873.7)	18,829.6 (14,799.2)	0.604
Socioeconomic status, in questionnaire (1-10)	3.765 (1.379)	3.710 (1.374)	3.859 (1.402)	3.748 (1.369)	3.694 (1.460)	3.816 (1.292)	0.494
Education: some college	0.310	0.358	0.288	0.289	0.338	0.277	0.094
Employment: full-time	0.414	0.399	0.408	0.415	0.426	0.421	0.955
Ethnicity: White	0.675	0.683	0.674	0.685	0.688	0.644	0.733
Country of birth: USA	0.915	0.906	0.925	0.917	0.929	0.898	0.578
Nationality: U.S.	0.947	0.938	0.940	0.957	0.950	0.949	0.810
First language: English	0.932	0.965	0.928	0.928	0.929	0.910	0.018
Observations	1,703	341	319	349	340	354	

Notes: The table provides means and standard deviations (when meaningful) in parentheses. Statistics on modal characteristics are reported for the categorical variables education, employment, ethnicity, country of birth, nationality and first language. Columns (5) and (6) report characteristics of the timer treatments Easy+Timer (E+T) and Difficult+Timer (D+T). Column (7) reports p-values of F-tests for equality of means in all treatments (i.e., columns 2-6). Note that there are no significant treatment differences in any categorical variable using χ^2 -tests including all categories. Participants indicated net household income in income brackets, the bracket mean is used as their income here. Net household income per capita divides net household income by participants' household size.

FINISHER SAMPLE				
	(1)	(2)	(3)	
	Correct tasks	Correct tasks	Correct tasks	
Easy	-0.281	-1.763	-2.622	
	(2.404)	(1.646)	(1.912)	
Difficult	-3.168*	-3.510**	-4.416**	
	(1.723)	(1.706)	(2.061)	
H ₀ (Easy=Difficult): p-value	0.237	0.303	0.376	
Comparison mean	38.008	38.006	38.740	
Sample	Finisher	Finisher	Finisher	
Controls	NO	YES	YES	
Outlier correction	NO	Winsorized	Robust Reg	
Observations	1,014	1,014	1,014	
(Pseudo) R-squared	0.002	0.028	0.026	

TABLE A8. TREATMENT EFFECTS IN COMPARISON TO CONTROL (ONLINE) –
EDUCTED SAMPLE

Notes: OLS estimates in comparison to control group. Column (2): Outcome variable top-coded at the 99th percentile. Column (3): Robust regression with MM-estimator and 85%-efficiency. Controls include age, income, socioeconomic status, sex, baseline self-efficacy, number of previous studies on Prolific and whether the attention check was passed. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	menues minime		
	(1)	(2)	(3)
	Correct tasks	Correct tasks	Correct tasks
Easy	-0.463	-1.896	-2.923
	(2.338)	(1.620)	(1.793)
Difficult	-5.549***	-6.097***	-8.305***
	(1.821)	(1.798)	(2.132)
Words written (demeaned)	0.069***	0.070***	0.095***
	(0.018)	(0.018)	(0.021)
Easy * Words written (demeaned)	-0.009	-0.001	-0.003
	(0.029)	(0.027)	(0.027)
Difficult * Words written (demeaned)	-0.040**	-0.039*	-0.051**
	(0.020)	(0.020)	(0.023)
H ₀ (Words+Difficult*Words=0): p-value	< 0.001	< 0.001	< 0.001
H ₀ (Easy=Difficult): p-value	0.039	0.019	0.009
Comparison mean	38.008	38.006	38.770
Sample	Finisher	Finisher	Finisher
Controls	NO	YES	YES
Outlier correction	NO	Winsorized	Robust Reg
Observations	1,014	1,014	1,014
(Pseudo) R-squared	0.030	0.082	0.091

TABLE A9. TREATMENT EFFECTS	5 AND WORDS WRITTE	n (Online) –	FINISHER SAMPLE

Notes: OLS estimates in comparison to control group. Column (2): Outcome variable top-coded at the 99th percentile. Column (3): Robust regression with MM-estimator and 85%-efficiency. Controls include age, income, socioeconomic status, sex, baseline self-efficacy, number of previous studies on Prolific and whether the attention check was passed. The variable "Words written" is demeaned using the mean of the Finisher sample across all treatments. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

TABLE ATU. TIME.	TABLE ATO, TIMER TREATMENTS IN COMPARISON TO CONTROL (ONLINE)						
	(1)	(2)	(3)				
	Correct tasks	Correct tasks	Correct tasks				
Easy+Timer (E+T)	-1.518	-1.496	-2.396				
	(1.742)	(1.695)	(1.990)				
Difficult+Timer (D+T)	0.233	0.062	-0.288				
	(1.747)	(1.699)	(2.031)				
H ₀ (E+T=D+T): p-value	0.328	0.361	0.291				
Comparison mean	36.463	36.461	36.636				
Sample	ITT	ITT	ITT				
Controls	NO	YES	YES				
Outlier correction	NO	Winsorized	Robust Reg.				
Observations	1,108	1,108	1,108				
(Pseudo) R-squared	0.001	0.031	0.033				
(rseudo) K-squared	0.001	0.031	0.035				

TABLE A10. TIMER TREATMENTS IN COMPARISON TO CONTROL (ONLINE)

Notes: OLS estimates in comparison to control treatment. Column (2): Outcome variable top-coded at the 99th percentile. Column (3): Robust regression with MM-estimator and 85%-efficiency. Controls include all variables that are available for the ITT sample: sex, baseline self-efficacy, number of previous studies on Prolific and whether the attention check was passed. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

FINISHER SAMPLE					
	(1)	(2)	(3)		
	Correct tasks	Correct tasks	Correct tasks		
Easy+Timer (E+T)	-1.261	-1.031	-2.323		
•	(1.724)	(1.660)	(1.966)		
Difficult+Timer (D+T)	2.142	1.722	1.151		
	(1.709)	(1.662)	(1.981)		
H ₀ (E+T=D+T): p-value	0.053	0.098	0.077		
Comparison mean	38.008	38.006	38.893		
Sample	Finisher	Finisher	Finisher		
Controls	NO	YES	YES		
Outlier correction	NO	Winsorized	Robust Reg.		
Observations	1,043	1,043	1,043		
(Pseudo) R-squared	0.004	0.045	0.041		

TABLE A11. TIMER TREATMENTS IN COMPARISON TO CONTROL (ONLINE) – Finished Sample

Notes: OLS estimates in comparison to control treatment. Column (2): Outcome variable top-coded at the 99th percentile. Column (3): Robust regression with MM-estimator and 85%-efficiency. Robust standard errors in parentheses. Controls include age, income, socioeconomic status, sex, baseline self-efficacy, number of previous studies on Prolific and whether the attention check was passed. *** p < 0.01, ** p < 0.05, * p < 0.1.

TABLE A12. ALL TREATMENTS IN COMPARISON TO CONTROL (ONLINE)					
	(1)	(2)	(3)		
	Correct tasks	Correct tasks	Correct tasks		
Easy	-1.693	-2.792*	-3.447*		
-	(2.332)	(1.676)	(1.953)		
Difficult	-6.347***	-6.357***	-7.551***		
	(1.730)	(1.715)	(2.048)		
Easy+Timer	-1.518	-1.523	-2.416		
-	(1.742)	(1.695)	(1.959)		
Difficult+Timer	0.233	0.035	-0.313		
	(1.747)	(1.698)	(1.994)		
Comparison mean	36.463	36.461	36.650		
Sample	ITT	ITT	ITT		
Controls	NO	YES	YES		
Outlier correction	NO	Winsorized	Robust Reg.		
Observations	1,848	1,848	1,848		
(Pseudo) R-squared	0.007	0.034	0.035		

TABLE A12. ALL TREATMENTS IN COMPARISON TO CONTROL (ONLINI
--

Notes: OLS estimates in comparison to control treatment. Column (2): Outcome variable top-coded at the 99th percentile. Column (3): Robust regression with MM-estimator and 85%-efficiency. Robust standard errors in parentheses. Controls include age, income, socioeconomic status, sex, self-efficacy, number of previous studies on Prolific and whether the attention check was passed. *** p<0.01, ** p<0.05, * p<0.1.

Appendix B: A Simple Theoretical Framework

We present a simple model building on Kaur et al. (2024). Let an individual derive utility from consuming the output she produces. Her production function f comprises two inputs: an effortful and consciously chosen input e and an automatic input a. While e is under control of the individual (e.g., how hard she tries to work), a depends on the available cognitive resources (e.g., the capacity to stay focused on the task) and cannot directly be chosen by the individual. Exerting effort e is costly to the individual. We assume separability of consumption utility and costs of effort. For simplicity, and without loss of generality, we model only one period such that all output is consumed and we can abstract from savings and debt. The individual chooses an optimal effort level $e^* \ge 0$ facing the following decision problem:

$$\max_{e} u(e) = \max_{e} \left[f(e, a) - c(e) \right]$$
(1)

We apply standard assumptions on the functional form, as in Kaur et al. (2024). The production function has positive and diminishing marginal returns to inputs e and a ($f'_e > 0$, $f''_{ee} < 0$ and $f'_a > 0$, $f''_{aa} < 0$). Inputs are complements ($f''_{ea} > 0$), i.e., increasing available cognitive resources improves marginal returns to effort. Effort costs are convex ($c''_{ee} > 0$). We do not model labor supply on the extensive margin assuming incentives are such that there is no participation constraint.

The self-affirmation treatment presented in the paper aims to increase available cognitive resources and thus to exogenously increase the automatic input a in the individual's production function. Our simple model provides two channels through which increasing a also increases produced output.

First, automatic input *a* can be considered as the effectiveness factor of effort in our model, i.e., how well effort translates into actual output. For any given level of effort (e.g., a consciously chosen work speed), increasing the available cognitive resources (e.g., the capacity to focus on the task) therefore directly increases output (e.g., by reducing attentional errors).

Second, exogenously altering automatic input *a* also impacts output indirectly by changing optimal effort levels. To see this, we derive the first-order condition of (1), yielding a standard equation of marginal benefits and marginal costs that characterizes optimal effort e^* :

$$f'_{e}(e^{*},a) = c'_{e}(e^{*})$$
⁽²⁾

We obtain the effect of manipulating *a* on optimal effort e^* by differentiating the first-order condition with respect to *a* and solving for $\frac{\partial e^*}{\partial a}$.

$$f''_{ee} \frac{\partial e^*}{\partial a} + f''_{ea} \frac{\partial a}{\partial a} = c''_{ee} \frac{\partial e^*}{\partial a}$$
$$\frac{\partial e^*}{\partial a} = \frac{f''_{ea}}{c''_{ee} - f''_{ee}}$$
(3)

Given the assumptions outlined above (complementarity of inputs, convex costs, and diminishing marginal returns), the sign of $\frac{\partial e^*}{\partial a}$ in (3) is positive: increasing the availability of cognitive resources, i.e., increasing automatic input *a*, increases optimal effort levels. Note that the level of *a* is set exogenously before individuals choose their effort level. This is mirrored in our experiment by administering the treatment before the job is announced.

 \Leftrightarrow

Appendix C: Alternative Explanations (Field Experiment)

This section discusses two alternative explanations for our treatment effects and argues that they are unlikely to explain the backfiring of the self-affirmation treatment. The first alternative explanation proposes that our self-affirmation intervention worked as planned and made participants feel proud and successful. The consequence of this, however, might not be an improved utilization of available skills and cognitive resources for productivity on the peanutcracking job. Rather, it may have induced a perception of overqualification for this simple task that in turn reduced motivation and satisfaction with the job (Peiró, Agut, and Grau 2010). Participants who feel better about themselves and see their potential in a better light might want to distance themselves from the "menial" task of cracking nuts. This perception of overqualification should be especially present among participants with higher education.

To evaluate whether more highly educated individuals were less productive in the selfaffirmation treatment, we interact the treatment dummy with a median split of education (more than five years of schooling) in Table A5. For this regression, we use the self-affirmation treatment as the omitted category, which allows a more straightforward interpretation of the coefficient of the above median education variable. We find that, in the self-affirmation treatment, participants with an above-median education (five years of schooling or more), if anything, displayed a higher level of productivity (p < 0.1). Observing a less negative effect of the self-affirmation intervention among more highly educated participants is inconsistent with the overqualification explanation. For the control treatment, we do not find such an additional effect. In fact, column (2) suggests a small and negative point estimate for above-median education in the control treatment. Instead, this regression provides additional support for the finding that the treatment effects are driven by the ease of answering. Our ease coding strongly correlates with having above-median education. In the self-affirmation treatment, having above-median education increases the Ease score by 1.26 points (p = 0.010, t-test). There is no significant correlation between Ease and having above-median education in the control treatment, however (p = 0.315).

A second explanation for our backfiring could be that the control treatment was, in fact, not neutral. In this treatment, we asked participants to describe aspects of their daily life. It is possible that by simply showing interest in the life and circumstances of marginalized people in the context of a research project, our interviews could have alleviated the participants' mental burdens. The interviews by strangers who explicitly visited participants' villages to learn about their lives potentially gave their stories and experiences an importance that they did not feel

before, which in turn made them feel proud and, in line with our conjecture, also positively impacted their productivity in the nut-cracking job.

If anything, our content analysis suggests that even in the control treatment participants talked about negative experiences such as being unemployed (see also Table A.3.2). If this had a similar effect on productivity as the self-affirmation treatment, we would potentially even underestimate the degree of backfiring in comparison to a truly untreated group. Furthermore, we find an insignificant negative point estimate of ease of answering in the control treatment in the regressions presented in Table A4. This is also inconsistent with the explanation that our control treatment was not neutral.

Appendix D: Psychological Questionnaire (Field Experiment)

Before the interventions in the field experiment, participants completed a short questionnaire that elicited age and other demographic information, as well as psychological measures of selfesteem (Robins, Hendin, and Trzesniewski 2001) and locus of control (Kovaleva 2012). The psychological items were also elicited immediately after the intervention. The one-item selfesteem scale comprised the statement "*I have a high self-image*", and the four-item locus of control scale comprised "*I am my own boss*" and "*If I work hard, I will succeed*" as measures of internal locus of control. External locus of control was measured by asking "*Whether at work or in my private life: what I do is mostly determined by others*" and "*Fate always gets in the way of my plans*". For both scales, participants were asked to indicate verbally their approval on a five-point Likert scale ranging from "not true for me at all" to "very true for me".

Our local research assistants reported that participants had severe problems understanding the Likert scales and even the questions themselves. This might have resulted in relatively high overall approval (all items except for the third locus of control item have an average approval larger than four), leaving little variance in the data.

These observations mirror the results by Laajaj and Macours (2021), who systematically investigate the reliability and validity of several psychological scales, which were originally developed in high-income countries, to measure cognitive and non-cognitive skills in a rural low-income context. Measurements of non-cognitive skills, such as self-esteem and locus of control, often suffer from large and non-classical measurement error and low internal validity leading to low reliability. These problems can be induced by research assistants during the verbal administration of questionnaires, participants' difficulties in understanding the questions, or response biases such as acquiescence bias ("yes-saying" especially among participants who do not fully understand the questions), extreme response biase, or social desirability bias.

Appendix E: Experimental Protocol and Materials

[Translated from Afrikaans]

General Protocol:

Introduction to the Experiment for all Participants

Welcome. Thank you so much for coming.We are a group from the University of Cologne in Germany.We work with the Legal Assistance Center (LAC) in Namibia.

Today, we are going to hold some interviews about livelihood in Omaheke. We are very happy that you are participating.

We are interested in your experiences and your opinions on various topics. We will be holding short interviews with every one of you. To make interviews easier, we will now divide you into smaller groups. Then the interviewers will explain everything to you.

Thank you very much. [Then send people to different groups.]

Introduction of the Interviewers in the Groups

Welcome!

My name is _____ and I will now interview you.

I am only going to talk to one person at a time.

When it's not your turn, please stay in your seat and do not talk to each other.

Please do not disturb the interviews.

Stay with your group when you are done.

Thank you very much.

Introduction and Explanation of the Job

The interviews are now complete.

Thank you so much for your stories about [personal success / what you usually do during the weekdays].

Before we leave, we will offer you a job. If you want, you can make money with that job. We work with a German farmer in Omaheke. He's looking for peeled peanuts. We have a lot of peanuts you need to peel. You can work for an hour. After that, we have to leave.

We will give you the peanuts in their skins.

Your job is to peel the peanuts with your hands. Put the whole peanuts in the bowl and the broken ones in the other bowl.

Please do not bite or eat the peanuts.

Please put the peels in the large bowl.

In the end, we will weigh the peanuts.

You can work for an hour. After that, we have to leave.

For every 10 grams of unbroken peanuts in your bowl, you will receive N\$ 1. This means the more peanuts you peel, the more money you can get. If you peel about 14 peanuts, you get N\$ 1.

It is important that you separate the peanuts and peels. Put the unbroken peanuts in the bowl, the broken ones in the other bowl, and the peels in the big bowl.

We have separated the table with cardboard boxes, so everyone has their own workplace. Please do not talk to each other during work, so that you do not disturb your neighbors and can concentrate better on your work.

Are there any questions?

I am now going to hand out the peanuts. [When everyone has nuts:] You can now start peeling.

After 60 minutes:

Stop. Please stop working now.

[Put a cover on the bowls.]

I will now weigh the peanuts and pay you for the job. Please bring all the peanuts to me.

[After payment:]

Thank you very much again for your participation and your personal stories in the interviews. They are of great interest to the university in Germany and the Legal Assistance Center. The people are looking forward to listening to all of your stories!

Experimental Materials below.

- **1. Recruiting Survey**
- 2. Experimental Survey including Treatment Question (SELF-AFFIRMATION)
- 3. Experimental Survey including Treatment Question (CONTROL)

1. Recruiting Survey [Place, Date]

Welcome. My name is		
We are a group from the University of G	Cologne in Germany.	
We work with the Legal Assistance Cer	nter (LAC) in Namibia	1.
Today, we are going to hold some interv	views about livelihood	l in Omaheke.
The interviews will take place at	(time) at	(location).
We will provide snacks and soft drinks.	You will also receive	some money.
Are you interested in participating?		

[IF YES:] I already want to ask you some questions, is it ok?

Name:				
Phone:				
Age:		Sex:	Man	Woman
What is your home	language?			
Khoekhoegowab	Ju/'hoansi	Afrika	ians	<i>Other</i> :
What was your higl	hest grade in s	school?	Grade	
What kind of work	do you do?			

Please describe if the following sentence is true for you.

I like competing with others.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

What material is the house you live in mostly made of?

I do not live in a house or apartment - plastic bags and natural materials - corrugated iron bricks - bricks with more than three rooms

What kind of toilet do the members of your household use?

flush toilet – flush-by-hand toilet - Dry toilet / Long-drop - none

Do you have a means of transportation?

None – Motorbike – Donkey Cart – Other:

Please describe if the following sentences are true for you.

1. I have a high self-esteem.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

2. I am my own boss.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

3. If I work hard, I will be successful.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

4. Whether it's in my job or in my private life: what I do is mostly decided by others.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

5. Fate is often in the way of my plans.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

2. Experimental Survey [<i>Place, Date</i>] (Interviewer:	// Number:)
Name:		
Age:		

Please describe an experience when you felt successful and proud of yourself.

(Time: _____ // Translated? Yes No // Answer Afrikaans? Yes No)

(How many questions? *none* 1-2 3-5 *more*)

Please describe if the following sentences are true for you.

1. I have a high self-esteem.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

2. I am my own boss.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

3. If I work hard, I will be successful.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

4. Whether it's in my job or in my private life: what I do is mostly decided by others.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

5. Fate is often in the way of my plans.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

Thank you so much for your stories of personal success!

Understood:	//	Anything special? Child – Drunk - No Afrikaans
at all		
Comments:		

Name:_____

How satisfied	l are you wit	h the number	of peanut	peels you	have peeled?
	v		1		1

A - B - C - D - E - F - G - U

What do you think: how much money have you earned? N\$ _____

Do you still remember how much money you get for 10 grams of unbroken peanuts?

N\$ _____

Total amount of whole nuts: _		gram
Total amount of broken nuts: _		gram
Total income:	10 +	N\$

How many of the people in your group do you know? 0 - 1 - 2 - 3

3. Experimental Survey [<i>Place</i> , <i>Date</i>] (Interviewer:	// Number:)	ı
Name:		
Age:		

Please describe what you usually do during each of the weekdays.

(Time: _____ // Translated? Yes No // Answer Afrikaans? Yes No)

(How many questions? *none* 1-2 3-5 *more*)

Please describe if the following sentences are true for you.

1. I have a high self-esteem.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

2. I am my own boss.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

3. If I work hard, I will be successful.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

4. Whether it's in my job or in my private life: what I do is mostly decided by others.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

5. Fate is often in the way of my plans.

Not true of me - Slightly true - Somewhat true - Mostly true - Completely true of me

Thank you so much for your stories about what you do during the weekdays.

Understood:	//	Anything special? Child – Drunk - No Afrikaans
at all		
Comments:		

Name:

How satisfied are you with the number of peanut peels you have peeled?

A - B - C - D - E - F - G - U

What do you think: how much money have you earned? N\$ _____

Do you still remember how much money you get for 10 grams of unbroken peanuts? N\$

Total amount of whole nuts: _		gram
Total amount of broken nuts:		gram
Total income:	10 +	N\$

How many of the people in your group do you know? 0 - 1 - 2 - 3

BACKFIRING OR BENEFITS? HETEROGENEOUS EFFECTS OF SELF-AFFIRMATION ON PRODUCTIVITY

Abstract. Self-affirmation interventions hold great promise in improving outcomes, especially for disadvantaged groups in society. However, Rockenbach et al. (2024) demonstrate that a standard self-affirmation intervention can backfire and impair productivity. I analyze data from their online experiment (N=1,109) to understand potential heterogeneous effects of two self-affirmation treatments, in which participants had to describe either three (easy) or eight (difficult) experiences that made them feel successful and proud, and a control treatment, in which participants had to describe their daily routines. The results suggest some heterogeneity in the effects of self-affirmation on productivity with respect to gender, socioeconomic status, and baseline self-efficacy. Specifically, there is heterogeneity in how the difficulty of the self-affirmation does not benefit any of the analyzed subgroups, even when recalling self-affirming experiences is easy. These insights have important practical implications for the design of future self-affirmation interventions.

Acknowledgements: I thank Sebastian Tonke for his comments and suggestions. The experiment was preregistered at the AEA RCT Registry (AEARCTR-0012142). IRB approval for the study was obtained from the Ethics Committee of the Faculty of Economic and Social Sciences, University of Cologne (Number 230035BR).

3.1. Introduction

Self-affirmation is a light-touch, low-cost, and easy-to-implement psychological intervention designed to create and enhance a positive self-image. Self-affirmation can improve outcomes in various domains such as education, health, and inter-group conflict (for extensive overviews, see Cohen and Sherman 2014; Sherman et al. 2021). Often, these interventions have been particularly effective for disadvantaged groups facing negative identity-based stereotypes or specific psychological threats in a given context. For instance, self-affirmation has improved women's performance in stereotypically male-dominated tasks such as mathematics and physics tests (Martens et al. 2006; Miyake et al. 2010; Shapiro, Williams, and Hambarchyan 2013) or improved short- and long-term educational outcomes for other treated groups in these studies, such as men or non-minority students, were typically unaffected by the interventions. The effects of self-affirmation thus seem to depend on specific individual characteristics. Understanding this heterogeneity is crucial when designing, implementing, and evaluating self-affirmation interventions, and tailoring them to specific groups and contexts.¹

The current paper follows up on Rockenbach et al. (2024), where we examine the effects of a standard self-affirmation intervention on labor productivity. In our self-affirmation protocol, participants were asked to recall and describe experiences that made them feel successful and proud. We find that self-affirmation backfired and substantially decreased productivity both in a crop production task in rural Namibia and in a real-effort task in an online experiment with U.S. participants. We argue that this backfiring occurs when recalling selfaffirming experiences is difficult compared to when recall is easy, a mechanism known as "ease of retrieval" (Schwarz et al. 1991; Schwarz and Vaughn 2002). When self-affirmation is too difficult, it may induce a negative self-image rather than a positive one and thereby exacerbate psychological stress rather than alleviate it. Defending against this stress consumes scarce mental resources and impairs labor productivity. While Rockenbach et al.'s (2024) findings demonstrate that self-affirmation interventions can backfire on average, they do not answer whether these effects vary systematically across important sociodemographic and psychological dimensions. Importantly, they also do not answer whether there are subgroups of participants for which self-affirmation has the intended positive effects. In the current paper, I provide evidence on these questions.

¹ Sherman et al. (2021) refer to the importance of the "three Ts": targeting the correct person, tailoring the intervention to the need, and timing the intervention right.

I study heterogeneous effects of a self-affirmation intervention on productivity using data from the online experiment of Rockenbach et al. (2024). As pre-registered, I investigate how the intervention's effects differ by gender, socioeconomic status, and baseline self-efficacy, three dimensions derived from the literature as potential moderators.² I focus on heterogeneity in the effects of two self-affirmation treatments, in which participants had to describe either three ("Easy") or eight ("Difficult") experiences that made them feel successful and proud, and a control treatment ("Control"), in which participants had to describe their daily routines. Specifically, I study heterogeneous effects of the two self-affirmation treatments on productivity by analyzing differences (e.g., between men and women) in the effects of Easy and Difficult compared to Control. I study heterogeneity in how ease of retrieval influences the effects of self-affirmation by analyzing differences (e.g., between men and women) in the treatment difference between Easy and Difficult.

I find statistically significant heterogeneity with respect to gender: the treatment difference between Easy and Difficult is significantly larger for women than for men. Women are about 20 percent more productive in Easy than in Difficult, whereas there is almost no treatment difference for men. This suggests that ease of retrieval more strongly affects women in this experiment. Beyond that, I find no statistically significant heterogeneity in any treatment comparisons with respect to gender, socioeconomic status, and baseline self-efficacy.

However, I argue that the absence of statistically significant heterogeneous effects does not necessarily imply evidence for the absence of heterogeneity. The point estimates in this study are often not tightly estimated zeros but are large and close to statistical significance at conventional levels. Moreover, power calculations suggest that while the sample size in this study (N = 1,109) is large compared to existing studies of self-affirmation and is sufficient to detect relatively small main effects, it may be insufficient to detect heterogeneous effects with sufficient power.

I conduct additional subsample analyses to identify cases in which self-affirmation produces significant treatment differences for one group but not for the other, even when the difference between the groups is not statistically significant. These analyses provide suggestive evidence of further heterogeneity. First, self-affirmation significantly reduces productivity in the Difficult treatment compared to Control only for women. There is no statistically significant backfiring of any self-affirmation treatment for men. Second, self-affirmation significantly

² Note that there is also a large literature studying the effects of self-affirmation on disadvantaged ethnic groups, especially in education (for an overview, see Sherman et al. 2021). I do not analyze heterogeneous effects with respect to ethnicity in this paper as there is not enough variation in the sample, which consists mostly of white participants (67% White, 12% Black, 6% Asian). For this reason, ethnicity was also not pre-registered as a category of interest.

reduces productivity in the Difficult treatment compared to Control only for participants with relatively high socioeconomic status but not for participants with low socioeconomic status. There is no statistically significant effect of any self-affirmation treatment for participants with low socioeconomic status. Third, even in the Easy treatment, participants with relatively low baseline self-efficacy are significantly less productive compared to Control, while there is no significant difference for high self-efficacy participants. Strikingly, the self-affirmation treatments show no significant positive effects compared to Control for any subgroup studied in this paper, and point estimates are negative in eleven out of twelve treatment comparisons.

This paper highlights the importance of individual sociodemographic and psychological characteristics as moderators of the effects of self-affirmation interventions. The results suggest that gender, and potentially also differences in socioeconomic status and baseline self-efficacy, are candidates for such moderators. Specifically, the difficulty of self-affirmation protocols appears to be a sensitive factor influencing their effects, and this paper shows heterogeneity in this mechanism. While a difficult self-affirmation protocol impairs productivity for some groups, it has no such effect on others. Moreover, the self-affirmation protocol studied in this paper does not benefit any of the analyzed subgroups, even when recalling self-affirming experiences is easy. Instead, the easy self-affirmation protocol backfires for participants with low baseline self-efficacy and appears to still be too difficult for this group. These insights have important practical implications for the design of future self-affirmation interventions. Target populations must be carefully selected, and protocols must be rigorously tested and tailored to these populations, not only in terms of content but also in terms of difficulty.

Heterogeneity in the effects of self-affirmation and similar psychological interventions is often attributed to situational differences that make interventions meaningful in a given context but not in others, that facilitate learning, and that provide opportunities for change (see, e.g., Walton and Wilson 2018). This paper provides evidence on individual differences, rather than situational differences, as potential sources of heterogeneity. Previous evidence on the influence of the specific individual differences studied in this paper is limited and mixed (see Section 2). The paper therefore also contributes to a broader understanding of the manifold conditions shaping self-affirmation interventions' success and failure (see Cohen and Sherman 2014; Sherman et al. 2021; Escobar-Soler et al. 2024).

The remainder of this paper is organized as follows. Section 2 reviews previous evidence on the three potential moderators of self-affirmation analyzed in this paper. Section 3 introduces the experimental design. Section 4 discusses theory and hypotheses. Section 5 introduces the empirical strategy. Section 6 presents the results. Section 7 discusses limitations and concludes.

3.2. Three Potential Moderators

Gender. Self-affirmation interventions have been successfully implemented to improve the performance of women in male-stereotyped contexts. Martens et al. (2006) show that explicit gender stereotypes impaired women's performance in math tests but that this effect disappears following a self-affirmation intervention. They find no effect on men, suggesting that self-affirmation specifically mitigates identity-based threats and is not just a simple performance boost. Similarly, Miyake et al. (2010) show in a widely cited field experiment that self-affirmation reduced the gender gap in academic achievement in a college physics course by significantly improving female students' performance. However, as Serra-Garcia et al. (2020) highlight, these findings critically depend on the empirical strategy and only hold for a small subset of participants.³ Their reanalysis suggests that, in the full sample, the gender gap was reduced because self-affirmation impaired male students' performance but did not affect female students. A related literature demonstrates that the effect of stereotype threats on women's performance strongly depends on gender differences in the reaction to informational context and feedback (Berlin and Dargnies 2016; Iriberri and Rey-Biel 2017).

The existing evidence offers no clear consensus regarding heterogeneous effects of selfaffirmation on men and women. Hence, the current paper contributes to this literature by analyzing potential gender differences in the effect of self-affirmation on productivity in a realeffort task, i.e., in a context with no obvious gender-specific stereotype threat.

Socioeconomic Status. Self-affirmation interventions are often employed to improve outcomes for socially and economically disadvantaged groups. For example, Hall et al. (2014) find that self-affirmation improved executive control, fluid intelligence, and help-seeking behavior among low-income visitors of a soup kitchen in the U.S. Contrarily, Rockenbach et al. (2024) find that a very similar self-affirmation intervention backfired and reduced productivity in a crop production task among an extremely poor and marginalized group in rural Namibia. Baranov et al. (2020) find no effect of a psychological intervention including self-affirmation on a battery of psychological and economic outcomes in Kenya. Ghosal et al. (2022) study a closely related intervention inspired by self-affirmation, which aimed to improve self-image and fight negative stigma, that showed positive effects on savings and health choices among sex workers in India.

³ In particular, Miyake et al. (2010) base their results on covariate-adjusted interaction effects that only provide a reasonable interpretation for men and women with no differences in these covariates (in their case, SAT scores and stereotype endorsement), which constitute only 28 percent of their sample.

As self-affirmation interventions often target socially and economically marginalized groups but there is mixed evidence on the effects of self-affirmation in this domain, differences in socioeconomic status constitute a natural candidate for potential heterogeneity.

Self-Efficacy. Self-affirmation theory argues that self-affirmation helps people maintain a positive self-image of self-integrity, defined as "a sense of global efficacy, an image of oneself as able to control important (...) outcomes in one's life." (Cohen and Sherman 2014, p. 336). Consistently, Albalooshi et al. (2020) find that self-affirmation improved self-control among participants because it created an "efficacious self-view", a belief that "one can carry out goals and influence the environment despite challenges" (p. 197). Altering perceptions of self-efficacy, therefore, seems to be one major function of self-affirmation interventions. Consequently, differences in individuals' baseline self-efficacy perceptions might influence how they respond to self-affirmation. Individuals with low baseline self-efficacy might experience a greater boost in self-image as there is more room for improvement. However, previous studies suggest that self-affirmation and similar positive interventions could be meaningful and effective only for those with high baseline self-perceptions (Creswell et al. 2005) and could even backfire for those with low self-perceptions (Wood, Perunovic, and Lee 2009).

3.3. Experimental Design

Treatments and Task. This paper analyzes data from the online experiment presented in Rockenbach et al. (2024). At the beginning of this experiment, participants were informed that the study consisted of two stages. In the first stage, baseline self-efficacy perceptions were elicited and participants were then asked to write about personal experiences (treatment manipulations). The second stage offered the possibility of earning additional money (real-effort task) and concluded with psychological and demographic questionnaires.

Participants were randomly assigned to one of three conditions. In two self-affirmation treatments, participants were asked to describe either three (treatment "Easy") or eight (treatment "Difficult") experiences "that made them feel successful and proud". This self-affirmation is similar to the protocols in Hall et al. (2014), Klein et al. (2001), and Larsen et al. (2023). The novel easy-vs.-difficult self-affirmation manipulation follows a standard paradigm in psychology on the effects of ease of retrieval (Schwarz et al. 1991; Schwarz and Vaughn

2002).⁴ In the control condition, participants were asked to "describe [their] daily routine on a typical day of the week", similar to the control conditions in Hall et al. (2014) and Baranov et al. (2020) and the field experiment of Rockenbach et al. (2024). In all conditions, participants were asked to spend five minutes in total to describe their experiences.⁵

In the subsequent real-effort task, participants were instructed to count the number of zeros in randomly generated 3x10 matrices of zeros and ones for ten minutes (as used in, e.g., Abeler et al. 2011). Participants earned a bonus payment of \$0.05 for each correctly solved matrix.⁶ Participants were informed that leaving the study early implied that neither their show-up fee nor their bonus from the second stage of the experiment would be paid. However, in both the writing stage and the real-effort task, they could press a skip button to continue with the next section without penalty.

Two final questionnaires were administered after the task. A psychological questionnaire included questions on the two primary dimensions the treatments were intended to manipulate: ease of recall ("It was easy / challenging to recall and describe all experiences") and feeling successful and proud ("When recalling the experiences, I felt very successful / I did not feel proud at all"). A demographic questionnaire concluded the study.

Sample and Randomization. The online experiment was implemented in Qualtrics with U.S. participants recruited on Prolific. As self-affirmation interventions are expected to particularly benefit disadvantaged groups, the sample was restricted to participants with after-tax household incomes approximately in the lowest quintile⁷ and to participants with a self-reported socioeconomic status in their Prolific profile below the scale median (i.e., 1-5 on a scale from 1-10 with 10 being the highest position).⁸ Treatment assignment was stratified by sex.

⁴ The treatments are calibrated based on pilot sessions. In a first pilot, without being prompted to write a particular number of stories, most participants were able to recall three stories, whereas only few could recall eight. In further pilots, recalling three stories was perceived as significantly easier than recalling eight.

⁵ The experimental design of Rockenbach et al. (2024) also includes two additional exploratory treatments that add a salient five-minute timer to Easy and Difficult, which automatically ends the writing stage after five minutes. However, these additional treatments do not significantly affect productivity. This paper therefore focuses on the treatments without a salient timer. Note that there are also no statistically significant heterogeneous effects of the timer treatments on productivity with respect to gender, socioeconomic status, and self-efficacy in the data.

⁶ The payment was calibrated based on pilot studies such that participants earn a substantial bonus on average. Payments on Prolific are calculated in GBP. As the sample only consists of U.S. participants, I report values converted to USD with the exchange rate at the time.

⁷ We initially recruited participants with after-tax household incomes of less than \$40,000. In line with our preregistration, we slightly increased the income threshold to reach the desired sample size. In the final questionnaire, 84 percent of participants report an income below \$50,000.

⁸ Participants also report their self-assessed socioeconomic status in the final questionnaire of the experiment. Socioeconomic status in their Prolific profile and in the final questionnaire are significantly but not perfectly correlated (Spearman's rho = 0.583, p < 0.001 in the full sample; highly similar in the single treatments with Spearman's rho between 0.574 and 0.593). For the analysis in this paper, I use socioeconomic status elicited in the experimental questionnaire as this constitutes participants' most recent assessment of their status. In the

Data collection took place in October and November 2023. The ITT sample in this paper consists of 1,109 participants in Easy, Difficult and Control, of whom 1,014 participants completed the experiment (Finisher sample). Table A1 summarizes the sample characteristics and shows that treatments were balanced. Participants were paid a show-up fee of \$2.07 and earned an average bonus payment of \$1.80 from the real effort task. In line with Prolific's rules and our experimental instructions, only participants who completed the experiment were paid.

3.4. Theory and Hypotheses

This section briefly introduces the general mechanism of self-affirmation interventions based on the simple theoretical model developed in Rockenbach et al. (2024) and motivates how heterogeneous effects can emerge.

A simple conceptual framework. Let an individual's production of output be an increasing function of two complementary inputs: consciously chosen costly effort e and automatic input a. The individual choses effort levels, for example how hard she tries to work, to maximize utility from production. In contrast, automatic input a is outside the individual's control and reflects available cognitive resources, capturing aspects such as the capacity to focus on a task by inhibiting distracting thoughts and stimuli.

Self-affirmation theory argues that individuals generally strive to maintain a sense of self-integrity, i.e., a positive self-image as morally adequate person, capable of controlling important aspects of their life (Steele 1988; Cohen and Sherman 2014). Negative stereotypes and stressful events can threaten this self-image and evoke defensive responses that consume scarce mental resources. Self-affirmation interventions buffer against these threats. They let people reflect on important aspects of their life, typically by describing core values or successful experiences. This creates and expands a positive self-image, which reduces stress and the need to fight psychological threats by reducing their perceived relative importance, freeing mental resources for more productive purposes. In the context of labor productivity, self-affirmation can help to enhance self-control, the ability to control impulses, block distractions, and stay focused on a tedious task (Rockenbach et al. 2024).

In the model, self-affirmation interventions increase the availability of cognitive resources *a*. This has two effects. First, it directly increases output for any given effort level

experimental questionnaire, 90 percent of participants report a socioeconomic status of 5 or lower on the scale from 1-10.

because, for example, increased focus makes the same effort more productive. Second, it indirectly raises optimal effort levels under standard assumptions (diminishing marginal returns, complementarity, convex costs, and separability; see Appendix B of the previous chapter for a detailed theoretical exposition).

Heterogeneous effects of self-affirmation can arise for several reasons in this framework. First, psychological threats and the associated mental burdens might be specific to certain groups and only these groups can benefit from the intervention, i.e., self-affirmation affects *a* only for a subset of participants in a given context. This is the standard case in many studies of self-affirmation that improve outcomes for disadvantaged, negatively stereotyped groups (e.g., Cohen et al. 2006; 2009; Miyake et al. 2010; Hall et al. 2014).

Second, even when a psychological threat affects all groups equally, diminishing marginal returns to *a* in the production function let individuals with low baseline *a* experience larger productivity gains from a similar increase in *a* than individuals with high baseline *a*. For instance, individuals with fewer baseline available cognitive resources would benefit more from a similar increase in available resources, ceteris paribus.

Third, individuals might respond differently to self-affirmation interventions for reasons exogenous to the model. Differences in the susceptibility to self-affirmation might lead to different increases in a from the same intervention. For example, there might be heterogeneity in the influence of ease of retrieval, i.e., differences in how the difficulty of a self-affirmation protocol influences its effects on a.

Hypotheses. There are no pre-registered hypotheses regarding the direction of potential heterogeneous effects for different subgroups. For example, it is unclear whether explicit negative stereotypes exist in this experimental setup that could constitute a threat only to a specific group. While there is a common stereotype that women perform worse than men in math-related tasks, which self-affirmation interventions have previously addressed (Martens et al. 2006; Shapiro et al. 2013), it is not clear whether this stereotype also exists and affects performance in an abstract real-effort task such that self-affirmation could have a positive effect for women in this setup. Similarly, participants of low socioeconomic status might encounter negative stereotypes related to labor productivity (e.g., "the poor are poor because they are lazy"). However, it remains unclear whether already a simple experimental task can evoke these stereotypes strongly enough for self-affirmation to have an effect.

Working on an incentivized real-effort task under pressure might generally constitute a stressful event, and self-affirmation could have differential effects across groups then. For

62

instance, participants with low self-efficacy perceptions might benefit more from a boost in their self-image than those with already high self-efficacy perceptions when faced with a challenging task. This would be consistent with the assumption of diminishing marginal returns to increases in cognitive resources in the model.

This paper also studies heterogeneity in ease of retrieval, i.e., heterogeneity in how the difficulty of the self-affirmation protocol influences its effect. Again, there are no directed hypotheses and the analysis is exploratory.

3.5. Empirical Strategy

To study heterogeneity in the treatment effects, I estimate regressions with interaction terms between treatments and the respective group indicators. The following regression model serves as an example for analyzing heterogeneous effects with respect to gender and is used to derive the hypotheses of interest. The same logic applies to all other analyses of heterogeneity in this paper with binary dummy variables indicating group membership (high vs. low socioeconomic status and high vs. low baseline self-efficacy perception).

Let $Easy_i$ and $Difficult_i$ be 1 if participant *i* was assigned to the respective treatment, and 0 otherwise. Let $Male_i$ be 1 if participant *i* is a man, and 0 if she is a woman. To study heterogeneous effects with respect to gender, I estimate to following regression with interaction effects:

$$Y_{i} = \beta_{0} + \beta_{E} \times Easy_{i} + \beta_{D} \times Difficult_{i} + \beta_{M} \times Male_{i} + \beta_{ME} \times Male_{i} \times Easy_{i} + \beta_{MD} \times Male_{i} \times Difficult_{i} + \varepsilon_{i}, \qquad \text{with } i = \{1, ..., N\}$$
(1)

For women, the effect of Easy compared to Control is:

$$\Delta Y_{W,E} = E[Y \mid Easy = 1, Difficult = 0, Male = 0]$$

- E[Y \mid Easy = 0, Difficult = 0, Male = 0] = \beta_E (2)

For men, the effect of Easy compared to Control is:

$$\Delta Y_{M,E} = E[Y | Easy = 1, Diff. = 0, Male = 1] - E[Y | Easy = 0, Diff. = 0, Male = 1]$$

= $\beta_E + \beta_M + \beta_{ME} - \beta_M = \beta_E + \beta_{ME}$ (3)

Hence, the first heterogeneous effect of interest is whether there are gender differences in the effect of Easy compared to Control:

$$\Delta Y_{M,E} - \Delta Y_{W,E} = (\beta_E + \beta_{ME}) - \beta_E = \beta_{ME}$$
(4)

By the same logic, gender differences in the effect of Difficult compared to Control are given by:

$$\Delta Y_{M, D} - \Delta Y_{W, D} = \beta_{MD} \tag{5}$$

The third heterogeneous effect of interest is whether there are gender differences in the difference between Easy and Difficult, i.e., whether men and women are differently affected by ease of retrieval. This is given by:

$$(\Delta Y_{M, D} - \Delta Y_{M, E}) - (\Delta Y_{W, D} - \Delta Y_{W, E}) = (\beta_D + \beta_{MD} - (\beta_E + \beta_{ME})) - (\beta_D - \beta_E)$$
$$= \beta_{MD} - \beta_{ME}$$
(6)

Consequently, I test three formal hypotheses. Are there heterogeneous effects of Easy compared to Control (H₀: $\beta_{ME} = 0$)? Are there heterogeneous effects of Difficult compared to Control (H₀: $\beta_{MD} = 0$)? Is there heterogeneity in the difference between Easy and Difficult (H₀: $\beta_{MD} - \beta_{ME} = 0$)? The first two hypotheses examine whether the self-affirmation treatments affect men and women differently when compared to the control treatment. The third hypothesis tests whether ease of retrieval differently affects men and women.

Variables. I analyze heterogeneity across three dimensions: gender, self-assessed socioeconomic status, and baseline self-efficacy perceptions. All analyses include binary dummy variables of the respective categories.

Gender. Participants indicate their sex in their Prolific profile ("What is your sex, as recorded on legal/official documents?"), which allows a binary choice between male and female. As standard in economics, I use the term "gender" in this paper to study behavioral and psychological differences between men and women (see, e.g., Croson and Gneezy 2009).⁹

Socioeconomic status. Participants self-assessed their socioeconomic status in the postexperimental questionnaire, indicating "where [they] would put [themselves] on the socioeconomic ladder" on a ten-point scale from 1 (very low) to 10 (very high). I created a dummy variable for (relatively) low vs. high socioeconomic status using a median split at 4 (see footnote 10).

Self-efficacy. Baseline self-efficacy perceptions were elicited pre-treatment using an eight-item scale (Chen, Gully, and Eden 2001, see Table A2 for the items). The scale showed high internal consistency (Cronbach's alpha = 0.950), implying that the items reliably measure the same underlying construct. Consequently, the questions were collapsed to generate a

⁹ According to the Merriam-Webster dictionary, sex refers to the biological "state of being male or female" whereas gender encompasses the "behavioral, cultural, or psychological traits typically associated with one sex."

combined measure of self-efficacy. For the analyses, I created a dummy variable indicating low vs. high self-efficacy using a median split at 5.25 on a scale from 1-7 (see footnote 13).

Analytical procedure. As pre-registered, the analyses are conducted in the ITT sample whenever possible. Participants' sex is known for all participants because male and female participants were simultaneously invited to different studies to achieve balanced treatments. Self-efficacy perceptions were elicited pre-treatment and they are available for all participants in the ITT sample. Socioeconomic status was only elicited in the post-experimental questionnaire, however, restricting the analysis of heterogeneity with respect to socioeconomic status to the Finisher sample.

Following the analysis of main effects in Rockenbach et al. (2024), I first examine heterogeneous treatment effects on productivity for each category (gender, socioeconomic status, and self-efficacy) by comparing the self-affirmation treatments Easy and Difficult to Control. I then examine heterogeneous effects in ease of retrieval by comparing Easy and Difficult. If the results suggest heterogeneity in ease of retrieval, I use data from the psychological questionnaire to further analyze the underlying mechanism.

3.6. Results

The self-affirmation intervention significantly backfired in the Difficult treatment on average (for a detailed exposition of the main effects, see Rockenbach et al. 2024). Productivity was also significantly lower in Difficult than in Easy. The post-experimental psychological questionnaire revealed that participants felt significantly less successful and proud in Difficult than in Easy. These findings support ease of retrieval as a mechanism: increasing the difficulty of recalling positive experiences in the self-affirmation intervention disaffirmed participants and impaired productivity.

The following sections analyze heterogeneous effects with respect to gender, socioeconomic status, and baseline self-efficacy perceptions. As there are no pre-registered directed hypotheses, potential heterogeneous effects are examined using two-sided t-tests for single coefficients and Wald-tests for linear combinations of coefficients.

TABLE 1. GENDER					
HETEROGENOUS TREATMENT EFFECTS					
	(1)	(2)	(3)		
	Correct tasks	Correct tasks	Correct tasks		
	FULL SAMPLE	WOMEN	MEN		
Easy	-2.273	-2.094	-3.507		
	(2.176)	(2.183)	(2.546)		
Difficult	-9.154***	-9.365***	-3.808		
	(2.243)	(2.245)	(2.586)		
	1 1 4 4				
Male * Easy (ME)	-1.144				
	(3.349)				
Male * Difficult (MD)	5.545				
	(3.424)				
Male	-1.240				
	(2.393)				
H_0 (ME = MD): p-value	0.048				
Comparison mean	37.355	37.355	36.115		
Sample	ITT	ITT	ITT		
Controls	YES	YES	YES		
Outlier correction	Winsorized	Winsorized	Winsorized		
Observations	1,109	551	558		
R-squared	0.033	0.065	0.017		

3.6.1. Heterogeneous Effects: Gender

Notes: OLS estimates with treatment coefficients in comparison to Control. Base category in (1): female participants. The outcome variable is top-coded at the 99th percentile. Controls include all variables that are available for the ITT sample: baseline self-efficacy, number of previous studies on prolific and whether the attention check was passed. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1 presents regression analyses of heterogeneous effects with respect to gender. Column (1) presents the main regression in the full sample. Columns (2) and (3) present regressions in the subsamples of men and women for illustrative purposes. The outcome variable, number of correctly solved matrices, is top-coded at the 99th percentile to account for the presence of extreme outliers.¹⁰ All reported tests refer to the full regression in column (1), following the empirical strategy derived in Section 5.

Heterogeneous effects of self-affirmation on productivity. For women, the self-affirmation intervention in the Difficult treatment significantly reduces the number of correctly solved tasks compared to Control (p < 0.001). The Easy treatment shows no statistically significant effect

¹⁰ A few participants seemingly did not count zeros in the real-effort task but entered the expected value very rapidly each time. The most extreme outlier was a male participant in Easy who correctly solved 628 matrices, corresponding to 17 times the treatment median. See Table A3 for different regression specifications as in Rockenbach et al. (2024) that do or do not account for outliers as robustness checks.
for women compared to Control. The insignificant coefficient of *Male* indicates that there is no gender difference in productivity in the Control treatment.

There are no statistically significant heterogeneous effects for men of Easy or Difficult compared to Control: both interaction terms *Male* * *Easy* and *Male* * *Difficult* are not significantly different from zero at conventional levels (p = 0.733 and p = 0.106, respectively). Yet, the estimated coefficient of *Male* * *Difficult* is relatively large, close to being statistically significant, and in the opposite direction of the treatment effect for women. This results in an estimated treatment effect of Difficult for men that is about 60 percent smaller than the estimated effect for women and statistically insignificant (p = 0.163 in a test of H_0 : *Difficult* + *Difficult* * *Male* = 0). Hence, other than for women, self-affirmation does not significantly backfire for men.

Finally, note that self-affirmation does not provide any significant benefits: the point estimates of all treatment comparisons are negative for both men and women.

Heterogeneity in ease of retrieval. There is statistically significant heterogeneity in the treatment difference between Easy and Difficult (p = 0.048 in a test of H_0 : Male * Easy = Male * Difficult). Women are about 20 percent less productive in Difficult than in Easy (p = 0.002 in a test of H_0 : Easy = Difficult). In contrast, there is almost no productivity difference between Easy and Difficult for men (p > 0.940 in a test of H_0 : Easy = Difficult + Male * Difficult). These findings suggest that ease of retrieval of positive experiences more strongly affects the productivity of women than men in this context.

Table 2 presents an analysis of differences between men and women in the psychological outcomes of the intervention. This analysis is conducted in the Finisher sample, for whom data on the psychological questionnaire is available.

Women, the base category, write significantly more words and more stories in Difficult than in Easy (columns (1) and (2)) and perceive recalling stories as harder and more challenging in Difficult (columns (3) and (4)). Consequently, in line with ease of retrieval as mechanism, they feel significantly less successful and proud in Difficult (columns (5) and (6)). The interaction term *Male* * *Difficult* is not significantly different from zero for any outcome variable, indicating no statistically significant heterogeneous effects for men.

		I ADL	E 2. OENDER			
PSYCHOLOGICAL OUTCOMES (EASY VS. DIFFICULT)						
	(1) (2) (3) (4) (5) (6)					
	Word count	Number of	Recall easy	Recall	Feeling	Feeling not
		stories		challenging	successful	proud
Difficult	74.482***	3.082***	-1.044***	1.069***	-0.420***	0.446***
	(13.016)	(0.217)	(0.185)	(0.211)	(0.152)	(0.150)
Male * Difficult	-7.187	0.008	0.164	-0.198	0.138	-0.236
	(18.088)	(0.302)	(0.252)	(0.286)	(0.213)	(0.218)
Male	-13.538	-0.129	-0.132	0.110	-0.166	0.362**
	(8.836)	(0.083)	(0.160)	(0.190)	(0.139)	(0.142)
Comparison Mean	124 526	2 719	5 170	3 181	5 602	2 000
Observations	660	660	660	660	660	660
R-squared	0.101	0.411	0 244	0.165	0 1 3 9	0.126
Sample	Finisher	Finisher	Finisher	Finisher	Finisher	Finisher
Cantrala	VEC	VEC	VEC	VEC	VEC	VEC
Controls	YES	YES	YES	YES	YES	YES
11 01 0 1			- D	a 1		

TADLE 2 GENDER

Notes: OLS estimates in comparison to the Easy treatment. Base category: female participants. Robust standard errors in parentheses. The outcome variables of Columns (3) to (6) are measured on a 7-point Likert scale. Note that due to the timing of the elicitation, we only have data on the psychological measures for participants who finished the experiment. Controls include age, income, socioeconomic status, self-efficacy, number of previous studies on prolific and whether the attention check was passed. *** p<0.01.

However, the direction of the estimated coefficients suggests that the effect might be smaller for men. In fact, the treatment difference between Easy and Difficult for men, estimated by the joint coefficient *Difficult* + *Male* * *Difficult*, is only marginally significant in column (5) (p =0.061) and not significantly different from zero in column (6) (p = 0.194). Unlike women, men do not feel significantly less proud in Difficult compared to Easy. Moreover, men already feel significantly less proud in Easy compared to women (p = 0.011). These results provide further suggestive evidence that women might be more strongly affected by ease of retrieval in this experiment. This may explain why the difference in productivity between Easy and Difficult is larger for women than for men.

Result 1 (Gender). The treatment difference in productivity between Easy and Difficult is significantly larger for women than for men. Supported by results from the psychological questionnaire, this suggests that women are impacted more strongly by ease of retrieval as a mechanism underlying the backfiring of self-affirmation. A subsample analysis further reveals that self-affirmation in the Difficult treatment only significantly backfires for women, but not for men.

3.6.2. Heterogeneous Effects: Socioeconomic Status

Table 3 presents the analysis of heterogeneous effects with respect to socioeconomic status (SES). Column (1) presents the main regression in the full sample, and columns (2) and (3) present regressions in the subsamples of participants with relatively high SES (median or above median) and low SES (below median) for illustrative purposes, similar to Table 1.¹¹ All tests refer to the full regression in column (1).

HETEROO	GENOUS TREATMENT	EFFECTS	
	(1)	(2)	(3)
	Correct tasks	Correct tasks	Correct tasks
	FULL SAMPLE	HIGH SES	LOW SES
Easy	-3.355	-3.557	0.606
	(2.159)	(2.163)	(2.554)
Difficult	-5.034**	-4.948**	-1.331
	(2.198)	(2.193)	(2.703)
	2.021		
Low SES * Easy (LE)	3.931		
	(3.336)		
Low SES * Difficult (LD)	3.881		
	(3.491)		
Low SES	-2.188		
	(2.404)		
H_0 (LE = LD): p-value	0.989		
Comparison mean	38.609	38.609	36.036
Sample	ITT	ITT	ITT
Controls	YES	YES	YES
Outlier correction	Winsorized	Winsorized	Winsorized
Observations	1,014	576	438
R-squared	0.028	0.032	0.037

TABLE 3. SOCIOECONOMIC STATUS (SES) (SES)
HETEROGENOUS TREATMENT FEFESTS

Notes: OLS estimates with treatment coefficients in comparison to Control. Base category in (1): participants with high SES. The outcome variable is top-coded at the 99th percentile. Controls include: gender, age, baseline self-efficacy, number of previous studies on prolific and whether the attention check was passed. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Heterogeneous effects of self-affirmation on productivity. Participants with high SES, the base category, were significantly less productive in Difficult compared to Control (p = 0.022).

¹¹ The median reported SES is 4 on a scale from 1-10. Of the 1,014 participants in the Finisher sample, 438 have below median SES (low SES) and 576 have median or above median SES (high SES). Participants with low SES have an average SES of 2.537 and participants with high SES have an average SES of 4.750. The difference is highly significant (p < 0.001, two-sided t-test).

The productivity difference between Easy and Control is not statistically significant (p = 0.120).¹²

There are no statistically significant heterogeneous effects of Easy and Difficult compared to Control for participants with low SES: both interaction terms *Low SES* * *Easy* and *Low SES* * *Difficult* are not significantly different from zero (p = 0.239 and p = 0.266). However, the estimated coefficients of the interaction terms are relatively large compared to the estimated treatment effects for high SES participants and are in the opposite direction. Consequently, for participants with low SES, there are no statistically significant differences between Easy and Control and Difficult and Control (p = 0.821 and p = 0.671 in tests of the respective linear combinations). Although the insignificant interaction terms do not allow to conclude that treatment effects differ between participants with low and high SES, the results from the subsample analysis show that self-affirmation significantly backfired only for participants with high SES, suggesting that SES is a potential moderator of the effects of self-affirmation.

Finally, note that self-affirmation does not provide any significant benefits: for participants with high SES, the point estimates of all treatment comparisons are negative. For participants with low SES, only the point estimate of Easy is positive but close to zero.¹³

Heterogeneity in ease of retrieval. There is no statistically significant heterogeneity between Easy and Difficult with respect to socioeconomic status (p = 0.989 in tests of H_0 : Low SES * Easy = Low SES * Difficult). The difference in productivity between Easy and Difficult is not significantly different from zero for participants with high SES (p = 0.450) and for participants with low SES (p = 0.512). These findings suggest no heterogeneity in ease of retrieval.

Result 2 (Socioeconomic Status). There are no statistically significant differences in the treatment effects for participants with high vs. low socioeconomic status (SES). A subsample analysis shows that self-affirmation in the Difficult treatment significantly backfires only for participants with high SES, but not for participants with low SES. This provides suggestive evidence for SES being a potential moderator of the effects of self-affirmation.

¹² Note that in other regression specifications presented as robustness checks in Table A4, the difference between Easy and Control is marginally significant with 0.05 .

¹³ At least, a positive coefficient of Easy for participants with low SES would be in line with the conjecture that self-affirmation may especially benefit disadvantaged groups.

3.6.3. Heterogeneous Effects: Self-efficacy

Table 4 presents the analysis of heterogeneous effects with respect to baseline self-efficacy. Column (1) presents the main regression in the full sample, and columns (2) and (3) present regressions in the subsamples of participants with relatively high self-efficacy (above median) and low self-efficacy (median or below median) for illustrative purposes, similar to Tables 1 and 3.¹⁴ All tests refer to the full regression in column (1).

TABLE 7. SELF-LITICACT				
HETEROGENOUS TREATMENT EFFECTS				
	(1)	(2)	(3)	
	Correct tasks	Correct tasks	Correct tasks	
	FULL SAMPLE	HIGH SELF-EFF	LOW SELF-EFF	
Easy	-1.009	-1.003	-4.648*	
	(2.332)	(2.341)	(2.404)	
Difficult	-7.132***	-7.218***	-5.836**	
	(2.396)	(2.403)	(2.447)	
Low self-efficacy * Easy (LE)	-3.665			
	(3.350)			
Low self-efficacy * Difficult (LD)	1.380			
•	(3.423)			
Low self-efficacy	3.181			
·	(2.386)			
H_0 (LE = LD): p-value	0.137			
· · · ·				
Comparison mean	34.984	34.984	38.205	
Sample	ITT	ITT	ITT	
Controls	YES	YES	YES	
Outlier correction	Winsorized	Winsorized	Winsorized	
Observations	1,109	542	567	
R-squared	0.034	0.037	0.028	

TABLE 4. SELF-EFFICACY

Notes: OLS estimates with treatment coefficients in comparison to Control. Base category in (1): participants with high baseline self-efficacy. The outcome variable is top-coded at the 99th percentile. Controls include gender, number of previous studies on prolific and whether the attention check was passed. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Heterogeneous effects of self-affirmation on productivity. Participants with high selfefficacy, the base category, were significantly less productive in Difficult compared to Control

¹⁴ The median participant has a baseline self-efficacy perception of 5.250 on a scale from 1-7. Of the 1,109 participants in the ITT sample, 542 have high self-efficacy perceptions (above median) and 567 have low self-efficacy perceptions (median or below median). High self-efficacy participants have an average score of 6.037 and low self-efficacy participants have an average score of 4.198. The difference is highly significant (p < 0.001, two-sided t-test).

(p = 0.003). There is no significant difference between Easy and Control for high self-efficacy participants (p = 0.666).¹⁵

There are no statistically significant heterogeneous effects of Easy or Difficult compared to Control with respect to baseline self-efficacy: both interaction terms *Low Self-Efficacy* * *Easy* and *Low Self-Efficacy* * *Difficult* are not significantly different from zero (p = 0.274 and p = 0.687). However, the estimated heterogeneous effect of Easy for low self-efficacy participants is relatively large compared to the estimated treatment effect of Easy for high self-efficacy participants. Consequently, for low self-efficacy participants, the Easy treatment also backfires significantly (p = 0.052 in a test of H_0 : *Easy* + *Low Self-Efficacy* * *Easy* = 0). The estimated treatment effect of Difficult compared to Control remains statistically significant for low self-efficacy participants as well (p = 0.019). Hence, while the Difficult treatment impairs productivity for all participants, even the Easy treatment backfires for low self-efficacy participants as well.

Finally, note that self-affirmation does not provide any significant benefits: the point estimates of all treatment comparisons are negative for both participants with low and high baseline self-efficacy.

Heterogeneity in ease of retrieval. There is no statistically significant heterogeneity in the difference between Easy and Difficult with respect to baseline self-efficacy (p = 0.135 in a test of *H*₀: Low Self-Efficacy * Easy = Low Self-Efficacy * Difficult). However, while for high self-efficacy participants, there is a significant and large difference in productivity between Easy and Difficult (p = 0.009), the difference between Easy and Difficult is small and not significantly different from zero for low self-efficacy participants (p = 0.661). Although the insignificant hypothesis test prevents the conclusion that there exists a heterogeneous effect, these results suggest that ease of retrieval could more strongly affect participants with high self-efficacy.

Table 5 presents an analysis of heterogeneity in the psychological outcomes with respect to baseline self-efficacy. High self-efficacy participants, the base category, write significantly more words and stories in Difficult than in Easy and find recalling stories less easy and more challenging (all p < 0.001). They consequently feel more successful (p = 0.096) and proud (p = 0.027) in Easy than in Difficult. There is no significant heterogeneity in these differences for low self-efficacy participants (all p > 0.240) and all treatment differences are also statistically significant for low self-efficacy participants (all p < 0.053). These results are in line with

¹⁵ See Table A5 for other regression specifications as robustness checks. All results remain unchanged.

observing no significant heterogeneity in productivity differences between Easy and Difficult and suggest that there is no heterogeneity in ease of retrieval with respect to baseline selfefficacy.

		TABLE 5. SE	ELF-EFFICACY			
	PSYCHOLOG	GICAL OUTCO	mes (Easy vs	. DIFFICULT)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Word count	Number of	Recall easy	Recall	Feeling	Feeling not
		stories		challenging	successful	proud
Difficult	68.688***	3.096***	-0.992***	1.125***	-0.225*	0.322**
	(13.197)	(0.211)	(0.180)	(0.213)	(0.135)	(0.145)
Low self-efficacy * Diff.	4.145	-0.014	0.049	-0.309	-0.253	0.006
	(18.076)	(0.301)	(0.257)	(0.288)	(0.215)	(0.221)
Low self-efficacy	-11.641	-0.187**	-1.050***	1.036***	-0.669***	0.687***
·	(8.957)	(0.088)	(0.169)	(0.197)	(0.147)	(0.143)
Comparison mean	123.278	2.780	5.642	2.671	5.850	1.827
Observations	660	660	660	660	660	660
R-squared	0.101	0.412	0.207	0.158	0.107	0.110
Sample	Finisher	Finisher	Finisher	Finisher	Finisher	Finisher
Controls	YES	YES	YES	YES	YES	YES

Notes: OLS estimates in comparison to the Easy treatment. Base category: participants with above median self-efficacy (high). Robust standard errors in parentheses. The outcome variables of Columns (3) to (6) are measured on a 7-point Likert scale. Note that due to the timing of the elicitation, we only have data on the psychological measures for participants who finished the experiment. Controls include age, income, socioeconomic status, sex, number of previous studies on prolific and whether the attention check was passed. *** p < 0.01, ** p < 0.05, * p < 0.1.

However, given that there is no heterogeneity in ease of retrieval, the insignificant treatment difference in productivity between Easy and Difficult for low self-efficacy participants must be driven by other factors. Indeed, there is significant and systematic heterogeneity in the psychological questionnaire between high and low self-efficacy participants in the Easy treatment. Low self-efficacy participants write significantly fewer stories (p = 0.034) and find recalling stories less easy and more challenging compared to high self-efficacy participants (both p < 0.001). As a result, low self-efficacy participants feel significantly less successful and proud in Easy compared to high self-efficacy participants (both p < 0.001).

These findings from the psychological questionnaire can explain why low self-efficacy participants are significantly less productive in Easy compared to Control, while there is no significant difference for high self-efficacy participants. This could also explain the smaller and insignificant difference in productivity between Easy and Difficult for low self-efficacy participants, given that there is no heterogeneity in the effects of Difficult on productivity. Taken together, the results provide evidence that high and low self-efficacy participants react differently to the Easy treatment, i.e., the treatment that also more closely resembles standard self-affirmation protocols. This suggests that baseline self-efficacy is a potential moderator of the effects of self-affirmation interventions.

Result 3 (Self-Efficacy). There are no statistically significant differences in the treatment effects for participants with high vs. low baseline self-efficacy. A subsample analysis shows that self-affirmation significantly backfires even in the Easy treatment for participants with low self-efficacy, but not for participants with high self-efficacy. This is supported by systematic differences in the psychological questionnaire, suggesting that baseline self-efficacy is a potential moderator of the effects of self-affirmation.

The results in this section do not support the conjecture that self-affirmation might particularly benefit participants with low baseline self-efficacy, who would have the most room for improvement. Instead, the finding that self-affirmation significantly impairs outcomes for low self-efficacy participants in the Easy treatment suggests that even the easy self-affirmation protocol was too difficult for these participants. This aligns with previous evidence on backfiring of light-touch psychological interventions for participants with low self-perceptions (Creswell et al. 2005; Wood et al. 2009).

3.7. Limitations and Conclusion

In the following, I discuss three limitations that may influence the conclusions drawn from this paper.

Statistical Power. Analyzing heterogeneous effects by testing the statistical significance of interaction terms in regressions often suffers from insufficient statistical power because sample sizes typically have to be much larger to achieve the same power compared to main effects (Sommet et al. 2023). The sample size of the experiment analyzed in this paper was calculated to detect even relatively small treatment main effects with sufficient power but was not calibrated for tests of heterogeneous effects (see Rockenbach et al. 2024). This limitation might explain the insignificance of some hypothesis tests in this paper.

The following example illustrates the problem. In Table 1, the estimated coefficient of the interaction term *Male* * *Difficult*, which captures heterogeneity in the effect of Difficult compared to Control for men, is not significantly different from zero at conventional levels (p = 0.106). However, the estimated coefficient is relatively large compared to the estimated

coefficient of *Difficult* (the treatment effect for women), with the estimated treatment effect being about 60 percent smaller for men.

In a regression including only Difficult and Control, the achieved power to find a significant interaction term with $\alpha = 0.05$ is only 0.362 (n = 739).¹⁶ That means, assuming the point estimates reflect the true parameters, there is only a 36.2 percent chance to find a significant heterogeneous effect of this magnitude with the given sample size. Consequently, to find a heterogeneous effect of at least the size of the estimated effect with the conventional power of 80 percent and $\alpha = 0.05$, one would need 2,243 instead of 739 observations in this case, i.e., three times the number of observations.¹⁷

Most hypothesis tests for heterogeneity in this paper yielded insignificant results and it seems plausible that at least some of these results may be caused by insufficient statistical power. Hence, I also conducted additional subsample analyses to identify cases in which selfaffirmation produced significant differences for one group but not for the other. While these findings are not definitive, they provide suggestive evidence for heterogeneity, highlighting sociodemographic and psychological factors that potentially influence the effects of selfaffirmation interventions.

Multiple Hypothesis Tests. As discussed in Section 5, there are nine main hypothesis tests for heterogeneous effects, three hypothesis tests for each of the three dimensions gender, socioeconomic status, and baseline self-efficacy. I report additional tests on subsamples and tests on psychological outcomes. However, the probability of falsely rejecting a null hypothesis and falsely claiming a significant finding (i.e., a type I error) increases rapidly in the number of hypothesis tests.¹⁸ Even when I only focus on the nine main hypothesis tests, the probability of falsely rejecting at least one null hypothesis is approximately 37 percent.

¹⁶ As in column (1) of Table 1, I regress the top-coded outcome variable on Difficult, Male, their interaction, and a set of control variables. I calculate power with an F-test on the coefficients of variation (R^2) of the unrestricted regression model and a restricted model without the interaction term using the "power rsquared" command in Stata 18. I obtain very similar results when comparing the full regression including all treatments (i.e., column (1) in Table 1) to a restricted version of this regression without the *Male* * *Difficult* interaction. However, using the simplified regression here with only Difficult and Control makes sample size comparisons easier.

¹⁷ The INTxPOWER web app by Sommet et al. (2023) yields a very similar required sample size using standardized effect sizes for the difference between Difficult and Control derived from the data. The approximate effect sizes for the treatment difference in the top-coded outcome variable are Cohen's d = -0.421 for women and Cohen's d = -0.146 for men, yielding a Cohen's d = -0.138 of the interaction term, i.e., a small effect size (see Sommet et al. (2023) on how to calculate effect sizes of interaction terms). Sommet et al. (2023) also show that the required sample sizes for high-powered studies of interaction effects strongly depend on the shape and size of the interaction, which depend on the shapes and sizes of the effects in the respective subgroups.

¹⁸ This probability is given by $1 - (1 - \alpha)^N$, with significance level α and N hypothesis tests (see, e.g., List, Shaikh, and Xu 2019).

A standard procedure would be to adjust p-values for multiple hypothesis testing. Before adjustments, only one of the nine main hypothesis tests is statistically significant (H_0 : Male * Easy = Male * Difficult with p = 0.048). Applying the standard Bonferroni-Holm method to adjust p-values for multiple hypothesis tests leaves none of the tests significant (the lowest corrected p-value is now p = 0.432).¹⁹

This paper suffers from a relatively small sample size given that the effect sizes of the interaction terms are also relatively small. With a larger sample, p-values would also be smaller and more robust to adjustments for multiple hypothesis tests. To maintain informativeness, the p-values reported in this paper remain unadjusted, but the evidence should consequently be treated with some caution.

Limitations of the experimental design. A potential shortcoming of the experimental design is that only participants with a socioeconomic status below the scale median in their Prolific profile were recruited for the experiment. Hence, the comparison of participants with relatively high vs. low SES in this sample rather constitutes a comparison of participants with medium/low vs. very low SES in absolute terms. Consequently, the difference between these two groups might not be large enough to generate meaningful psychological and behavioral differences. Creating more extreme distinctions between two SES groups within this sample, such as comparing the lowest vs. highest quartiles of SES, does not change the results qualitatively. Instead, by reducing the analyzed sample size, this method further reduces the statistical power of this study to detect heterogeneous effects.

The sample size in this experiment (N = 1,109) is not small, however, and larger than in many of the original studies on self-affirmation. However, many estimated effect sizes in this paper are quite small. It seems plausible that future studies with different outcome variables that are more prone to explicit negative stereotypes for some groups, different self-affirmation protocols that do not backfire, and a different participant pool, might be able to find larger and more significant heterogeneous effects.

Conclusion. Self-affirmation interventions and other low-cost psychological interventions hold great promise in improving outcomes, especially for disadvantaged groups in society (Cohen and Sherman 2014; Walton 2014; Walton and Wilson 2018; Haushofer and Salicath 2023). However, Rockenbach et al. (2024) demonstrate that such interventions can also backfire and

¹⁹ In the Bonferroni-Holm method, the unadjusted p-values are ranked, the smallest p-value is multiplied with N, the second smallest p-value is multiplied with (N - 1), etc. Adjusted p-values are then compared to the original significance level α .

have unintended negative consequences. The current paper draws attention to individual sociodemographic and psychological characteristics that potentially moderate these effects. The results suggest that there can be heterogeneity in the effects of self-affirmation with respect to gender, socioeconomic status, and baseline self-efficacy. The paper also shows that there can be heterogeneity in ease of retrieval, i.e., in how the difficulty of the self-affirmation protocol influences its effects. Future research should aim at a better assessment of systematic heterogeneity in the effects of self-affirmation with respect to individual characteristics. For example, high-powered studies could investigate gender differences among different self-affirmation protocols and outcomes and try to understand their causes and generalizability.

The results in this paper have to be treated with some caution. The sample size of this study is not large enough to detect relatively small heterogeneous treatment effects with sufficient power, a problem commonly found in studies analyzing heterogeneous effects (Sommet et al. 2023). Hence, future studies that investigate heterogeneity in the effects of self-affirmation could build on the exploratory results of this paper and base their power calculations on the size and direction of the estimates in this experiment.

Finally, the self-affirmation protocol in this experiment did not benefit any subgroup analyzed in this paper. Researchers and practitioners who want to implement psychological interventions in the field should carefully test their protocols to understand and integrate potential heterogeneity, to avoid unintended negative effects, and to create interventions that benefit their recipients.

Appendix A: Additional Tables

TABLE A1. BALANCE TABLE.					
	(1) Full	(2) Easy	(3) Difficult	(4) Control	(5) p-value
Sample: ITT					
Male	0.503	0.503	0.503	0.504	0.999
Self-efficacy, baseline (scale 1-7)	5.097 (1.210)	5.111 (1.229)	5.075 (1.218)	5.104 (1.187)	0.911
Number of previous studies on Prolific	1,338.5 (1,447.5)	1,458.1 (1,542.1)	1,297.0 (1,410.1)	1,260.1 (1,381.2)	0.157
Attention check: failed	0.025 (0.157)	0.027 (0.162)	0.022 (0.146)	0.027 (0.163)	0.853
Observations	1,109	370	370	369	
Sample: Finisher					
Age	37.52 (13.83)	38.75 (14.12)	36.45 (13.62)	37.31 (13.67)	0.100
Net household income (USD)	34,881.7 (21,909.9)	33,137.8 (21,631.4)	36,332.3 (23,335.0)	35,254.2 (20,769.7)	0.169
Net household income per capita (USD)	17,875.9 (14,000.4)	17,231.6 (12,657.4)	17,506.4 (14,435.6)	18,829.6 (14,799.2)	0.282
Socioeconomic status, in questionnaire (scale 1-10)	3.794 (1.355)	3.710 (1.374)	3.859 (1.402)	3.816 (1.292)	0.354
Education: some college Employment: full-time Ethnicity: White Country of birth: USA Nationality: U.S. First language: English	0.308 0.409 0.667 0.909 0.943 0.934	0.358 0.399 0.683 0.906 0.938 0.965	0.288 0.408 0.674 0.925 0.940 0.928	0.277 0.421 0.644 0.898 0.949 0.910	0.051 0.837 0.523 0.454 0.804 0.005
Observations	1,014	341	319	354	

Notes: The table provides means and standard deviations (when meaningful) in parentheses. Statistics on modal characteristics are reported for the categorical variables education, employment, ethnicity, country of birth, nationality and first language. Column (5) reports p-values of F-tests for equality of means in all treatments (i.e., columns 2-4). Note that there are no significant treatment differences in any categorical variable using χ^2 -tests including all categories. Participants indicated net household income in income brackets, I use the bracket mean as their income here. Net household income per capita divides net household income by participants' household size.

78

TABLE A2. SELF-EFFICACY SCALE

- 1. I will be able to achieve most of the goals that I have set for myself.
- 2. When facing difficult tasks, I am certain that I will accomplish them.
- 3. In general, I think that I can obtain outcomes that are important to me.
- 4. I believe I can succeed at most any endeavor to which I set my mind.
- 5. I will be able to successfully overcome many challenges.
- 6. I am confident that I can perform effectively on many different tasks.
- 7. Compared to other people, I can do most tasks very well.
- 8. Even when things are tough, I can perform quite well.

Notes: Participants indicate how strongly they agree or disagree to the statements on a seven-point scale ranging from "strongly disagree" to "strongly agree". The self-efficacy items are equivalent to the "New General Self-Efficacy Scale" in Chen, Gully, and Eden (2001).

HETEROGENOUS	TREATMENT EFFECTS:	ROBUSTNESS CHE	CKS
	(1)	(2)	(3)
	Correct tasks	Correct tasks	Correct tasks
Easy	-2.823	-2.273	-2.672
	(2.215)	(2.176)	(2.426)
Difficult	-9.160***	-9.154***	-10.217***
	(2.269)	(2.243)	-2.672
Male * Easy (ME)	2.242	-1.144	-1.793
	(4.655)	(3.349)	(3.963)
Male * Difficult (MD)	5.590	5.545	5.631
	(3.454)	(3.424)	(4.115)
Male	-1.769	-1.240	-1.801
	(2.399)	(2.393)	(2.838)
H_0 (ME = MD): p-value	0.476	0.048	0.067
Comparison Mean	37.355	37.355	37.863
Sample	ITT	ITT	ITT
Controls	NO	YES	YES
Outlier correction	NO	Winsorized	Robust Reg.
Observations	1,109	1,109	1,109
(Pseudo) R-squared	0.010	0.033	0.033

TABLE A3. GENDER	
HETEROGENOUS TREATMENT EFFECTS: ROBUSTNESS CH	IFCI

Notes: Regression estimates of treatment coefficients in comparison to Control. Base category: female participants. Column (1): OLS regression on the raw data without controls. Column (2): OLS regression with outcome variable top-coded at the 99th percentile. Column (3): Robust regression with MM-estimator and 85%-efficiency. Controls include all variables that are available for the ITT sample: baseline self-efficacy, number of previous studies on prolific and whether the attention check was passed. The regression models differ with respect to dealing with outliers and including controls. The most extreme outlier was a male participant in Easy who correctly solved 628 matrices, corresponding to 17 times the treatment median. Accounting for extreme outliers explains the large change in the estimated coefficient of *Male* * *Easy* from (1) to (2) and (3). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

HETEROGENOUS	TREATMENT EFFECTS:	ROBUSTNESS CHE	CKS
	(1)	(2)	(3)
	Correct tasks	Correct tasks	Correct tasks
Easy	-3.974*	-3.355	-4.525*
	(2.191)	(2.159)	(2.499)
Difficult	-4.662**	-5.034**	-6.153**
	(2.195)	(2.198)	(2.635)
Low SES * Easy (LE)	8.122	3.931	4.773
	(5.033)	(3.336)	(3.871)
Low SES * Difficult (LD)	3.588	3.881	4.520
	(3.536)	(3.491)	(4.225)
Low SES	-1.457	-2.188	-2.874
	(2.428)	(2.404)	(2.829)
H_0 (LE = LD): p-value	0.375	0.989	0.952
Comparison mean	38.614	38.609	39.598
Sample	Finisher	Finisher	Finisher
Controls	NO	YES	YES
Outlier correction	NO	Winsorized	Robust Reg.
Observations	1,014	1,014	1,014
R-squared	0.007	0.028	0.026

TABLE A4. SOCIOECONOMIC STATUS	
HETEROGENOUS TREATMENT EFFECTS: ROBUSTNESS CHE	CK

Notes Regression estimates of treatment coefficients in comparison to Control. Base category: participants with high SES. Column (1): OLS regression on the raw data without controls. Column (2): OLS regression with outcome variable top-coded at the 99th percentile. Column (3): Robust regression with MM-estimator and 85%-efficiency. Controls include gender, age, baseline self-efficacy, number of previous studies on prolific and whether the attention check was passed. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

HETEROGENOUS TREATMENT EFFECTS: ROBUSTNESS CHECKS				
	(1)	(2)	(3)	
	Correct tasks	Correct tasks	Correct tasks	
Easy	-1.495	-1.009	-1.256	
	(2.347)	(2.332)	(2.703)	
Difficult	-7.509***	-7.132***	-8.156***	
	(2.405)	(2.396)	(2.880)	
Low self-efficacy * Easy (LE)	-0.414	-3.665	-4.577	
• • • •	(4.629)	(3.350)	(3.917)	
Low self-efficacy * Difficult (LD)	2.129	1.380	1.085	
- · · · ·	(3.445)	(3.423)	(4.072)	
Low self-efficacy	2.936	3.181	3.991	
	(2.399)	(2.386)	(2.804)	
H_0 (LE = LD): p-value	0.586	0.137	0.162	
Comparison mean	34.984	34.984	34.851	
Sample	ITT	ITT	ITT	
Controls	NO	YES	YES	
Outlier correction	NO	Winsorized	Robust Reg.	
Observations	1,109	1,109	1,109	
R-squared	0.012	0.034	0.034	

TABLE A5. SELF-EFFICACY Hetepogenous Treatment Feeects: Rodustness Check

Notes: Regression estimates of treatment coefficients in comparison to Control. Base category: participants with high baseline self-efficacy perceptions. Column (1): OLS regression on the raw data without controls. Column (2): OLS regression with outcome variable top-coded at the 99th percentile. Column (3): Robust regression with MM-estimator and 85%-efficiency. Controls include gender, number of previous studies on prolific and whether the attention check was passed. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

Appendix B: Instructions

Overview of the Study

The study consists of two parts: The *first part* will involve questions about personal experiences and your self-perception. In the *second part*, you can work on a simple task and earn a substantial bonus depending on your performance.

For your participation, you will receive a fixed payment of 1.70 GBP (about 2.07 US Dollar). The study will take about 10 minutes if you choose not to work on the task in the second part. If you work on the task in the second part, the bonus will substantially increase your earnings depending on your performance. We will explain further details in the course of the study.

Please note that your data will be collected fully anonymously. We will not collect any information that allows us to draw inferences about your real identity. Your data will be used for scientific research only. You may quit the study for whatever reason (e.g., psychological discomfort), but you will only receive your payments if you finish the survey.

Please start by reading the instructions for the first part. You will receive the instructions for the second part once you have completed the first part.

Instructions for the First Part of the Study

The first part of the study begins with questions about your self-perception. Afterwards, you will be asked to describe some personal experiences for five minutes. We recommend using a physical keyboard for typing.

All necessary details will be displayed on the respective screens. Please carefully read and answer all questions.

Screen (Self-efficacy):

On the following two screens, we will give you a list of statements. Read the statements carefully and *think about whether these statements are a good description of yourself.*

Please indicate how strongly you agree/disagree to the following statements.

- Self-efficacy items (see Table A2)
- "It is important that participants in this study pay attention to the questions. Please select "Strongly disagree" for this statement to show that you carefully read the question." (Attention check)

Screen (Instructions Treatment):

Personal Experiences

On the following screens, you will be asked to think about and describe personal experiences. Detailed information will be given on the next screen.

(Treatments Easy, Difficult, and Control:) Please spend about *five minutes to describe your experiences*.

(Treatments Easy+Timer and Difficult+Timer:)

You will have *five minutes to describe your experiences*. You will see a countdown timer on your screen. After the five minutes have elapsed, the system will automatically transition you to the second part of the study.

Screen (Treatment Easy):

Note: Treatment Difficult only differs with respect to the number of experiences. Timer treatments include an additional timer that is displayed in the upper left corner.

Please describe three experiences that made you feel successful and proud.

Please spend about five minutes in total to think and write about your experiences.

Experience 1 (of 3)

Please try to write at least two or three sentences about your experience.

Click the blue "next" button at the bottom of the page to write about another experience on the subsequent page.

If you cannot think of any experiences to describe, please leave the input field blank, select "I cannot think of any experiences" below and click the next button. You will then proceed to the second part of the study.

Note: Once you leave this stage, returning is not permitted.

I cannot think of any experience



Screen (Treatment Control):

Please describe your daily routine on a typical day of the week.

Please spend about five minutes in total to think and write about your daily routine.



Click the blue "next" button at the bottom of the page to continue with the second part of the study.

Note: Once you leave this stage, returning is not permitted.



Screen (Instructions Second Part):

Thank you very much for your answers in the first part of the study.

Instructions for the Second Part of the Study: "Counting Zeros"

In the second part, you have the opportunity to work on a task and earn bonus rewards. The task is called "Counting Zeros".

On the next screen, you will encounter a table filled with zeros and ones. An illustrative example of such a table can be found at the bottom of this page.

The objective of the task is: count all the zeros in the table.

Please provide your answer in the designated input field. Your answer should be numerical - for instance, if your count is 13, as in the example below, please enter "13" instead of "thirteen". Once you have entered your answer, click on the "Submit & Next" button. The subsequent screen will present you with a fresh table of zeros and ones and your task remains the same - count the zeros.

You can solve as many tasks as you want to. *Each correctly solved task will earn you a bonus of 0.04 GBP (0.05 US Dollars).*

The second part of the study is *time-restricted to 10 minutes*.

You will see a countdown timer on your screen. After the 10 minutes have elapsed, the system will automatically transition you to the final questions of this study.

If you want to quit working on the "Counting Zeros" tasks before the 10 minutes have elapsed, you can click the respective button at the end of the page at any time. You will then proceed to the final questions of the study immediately. Please note: Once you leave the stage, returning is not permitted.

The example below shows how the "Counting Zeros" task will appear. (similar to next page)

Screen (Counting Zeros Task):

Time Remaining: 09:56

Count the number of zeros and enter your answer below.



If you want to quit working on the "Counting Zeros" tasks, please click below. You will then proceed to the final questions of the study. You will be paid for all the tasks that you have correctly solved up to this point.

Note: Once you leave this stage, returning is not permitted.

Quit the "Counting Zeros" tasks / Leave this stage

Screen (Manipulation Check / Psychological Questionnaire):

Finally, the statements on this screen refer to the experiences you were asked to describe in the first part of the study.

Please read the statements carefully and think about whether these statements are a good description of how you felt.

Please indicate how strongly you agree/disagree to the following statements.

- When recalling the experiences, I felt very successful.
- When recalling the experiences, I did not feel proud at all.
- It was easy to recall and describe all experiences, i.e., to fill up all the boxes.
- It was challenging to recall and describe all experiences, i.e., to fill up all the boxes.
- Overall, I believe I performed well on the task to recall and describe the experiences.

PUSHING THE BAD AWAY: REVERSE TULLOCK CONTESTS

Joint work with Bettina Rockenbach and Marcin Waligora

Abstract. The literature on rent-seeking primarily focuses on contests for *achieving gains*, although contests for *avoiding losses* are also omnipresent. Examples for such 'reverse' contests are activities to prevent the close-down of a local school or the construction of a waste disposal close-by. While under standard preferences investments in 'reverse' and 'conventional' contests should not be different, loss aversion predicts contests for avoiding losses to be fiercer than conventional ones. In our experimental data, the difference in investments between conventional and reverse Tullock contests is small and statistically insignificant. We discuss several explanations for this remarkable finding.

Published as: Rockenbach, Bettina, Sebastian Schneiders, and Marcin Waligora. 2018. "Pushing the Bad Away: Reverse Tullock Contests." *Journal of the Economic Science Association* 4 (1): 73–85.

As this chapter is based on a published paper, it has been replaced with a bibliographic reference in the published version of this dissertation. The paper can be accessed online via the following link: https://doi.org/10.1007/s40881-018-0052-7

Acknowledgments: Financial support from the German Research Foundation (DFG) is gratefully acknowledged. We like to thank the editor of the Journal of the Economic Science Association and two anonymous referees for helpful suggestions and comments.

References

- Abbink, Klaus, Jordi Brandts, Benedikt Herrmann, and Henrik Orzen. 2010. "Intergroup Conflict and Intra-Group Punishment in an Experimental Contest Game." *American Economic Review* 100 (1): 420–47.
- Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470–92.
- Ahn, T. K., R. Mark Isaac, and Timothy C. Salmon. 2011. "Rent Seeking in Groups." International Journal of Industrial Organization 29 (1): 116–25.
- Albalooshi, Sumaya, Mehrad Moeini-Jazani, Bob M. Fennis, and Luk Warlop. 2020. "Reinstating the Resourceful Self: When and How Self-Affirmations Improve Executive Performance of the Powerless." *Personality and Social Psychology Bulletin* 46 (2): 189–203.
- Amaldoss, Wilfred, and Amnon Rapoport. 2009. "Excessive Expenditure in Two-Stage Contests: Theory and Experimental Evidence." In *Game Theory: Strategies, Equilibria* and Theorems, edited by Ingrid N. Haugen and Anna S. Nilsen, 241–66. New York: Nova Science Publishers.
- Andreoni, James. 1995. "Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments." *The Quarterly Journal of Economics* 110 (1): 1–21.
- Aronson, Joshua, Geoffrey Cohen, and Paul R. Nail. 2019. "Self-Affirmation Theory: An Update and Appraisal." In *Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology*, edited by Eddie Harmon-Jones, 159–74. Washington: American Psychological Association.
- Ashraf, Nava, Gharad Bryan, Alexia Delfino, Emily A. Holmes, Leonardo Iacovone, Christian Johannes Meyer, and Ashley Pople. 2024. "Learning to See the World's Opportunities: Memory, Mental Experiencing, and the Economic Lives of the Vulnerable." *Working Paper*.
- Baik, Kyung Hwan, Subhasish M. Chowdhury, and Abhijit Ramalingam. 2015. "Group Size and Matching Protocol in Contests." *CBESS Discussion Paper 13-11R*, 1–29.
- Banerjee, Abhijit V., Rema Hanna, Gabriel E. Kreindler, and Benjamin A. Olken. 2017.
 "Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs." *The World Bank Research Observer* 32 (2): 155–84.
- Baranov, Victoria, Johannes Haushofer, and Chaning Jang. 2020. "Can Positive Psychology Improve Psychological Well-Being and Economic Decision-Making? Experimental Evidence from Kenya." *Economic Development and Cultural Change* 68 (4): 1345–76.
- Barker, Nathan, Gharad Bryan, Dean Karlan, Angela Ofori-Atta, and Christopher Udry. 2022.
 "Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress." *American Economic Review: Insights* 4 (4): 527–45.

- Berlin, Noémi, and Marie-Pierre Dargnies. 2016. "Gender Differences in Reactions to Feedback and Willingness to Compete." *Journal of Economic Behavior & Organization* 130: 320–36.
- Bernard, Tanguy, Stefan Dercon, Kate Orkin, Giulio Schinaia, and Alemayehu Seyoum Taffesse. 2023. "The Future in Mind: Aspirations and Future-Oriented Behaviour in Rural Ethiopia." *CEPR Discussion Paper No. 18492*, 1-83.
- Bessone, Pedro, Gautam Rao, Frank Schilbach, Heather Schofield, and Mattie Toma. 2021. "The Economic Consequences of Increasing Sleep Among the Urban Poor." *The Quarterly Journal of Economics* 136 (3): 1887–1941.
- Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107 (4): 1165–1206.
- Campos, Francisco, Michael Frese, Markus Goldstein, Leonardo Iacovone, Hillary C. Johnson, David McKenzie, and Mona Mensmann. 2017. "Teaching Personal Initiative Beats Traditional Training in Boosting Small Business in West Africa." *Science* 357 (6357): 1287–90.
- Chen, Gilad, Stanley M. Gully, and Dov Eden. 2001. "Validation of a New General Self-Efficacy Scale." *Organizational Research Methods* 4 (1): 62–83.
- Chowdhury, Subhasish M., Joo Young Jeon, and Abhijit Ramalingam. 2018. "Property Rights and Loss Aversion in Contests." *Economic Inquiry* 56 (3): 1492–1511.
- Chowdhury, Subhasish M., Roman M. Sheremeta, and Theodore L. Turocy. 2014. "Overbidding and Overspreading in Rent-Seeking Experiments: Cost Structure and Prize Allocation Rules." *Games and Economic Behavior* 87: 224–38.
- Cohen, Geoffrey L., Julio Garcia, Nancy Apfel, and Allison Master. 2006. "Reducing the Racial Achievement Gap: A Social-Psychological Intervention." *Science* 313 (5791): 1307–10.
- Cohen, Geoffrey L., Julio Garcia, Valerie Purdie-Vaughns, Nancy Apfel, and Patricia Brzustoski. 2009. "Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap." *Science* 324 (5925): 400–403.
- Cohen, Geoffrey L., and David K. Sherman. 2014. "The Psychology of Change: Self-Affirmation and Social Psychological Intervention." *Annual Review of Psychology* 65 (1): 333–71.
- Congleton, Roger D., Arye L. Hillman, and Kai A. Konrad. 2008. 40 Years of Research on Rent Seeking. Heidelberg: Springer.
- Cornes, Richard, and Roger Hartley. 2003. "Loss Aversion and the Tullock Paradox." University of Nottingham Discussion Papers in Economics No. 03/17, 1–25.

——. 2012. "Loss Aversion in Contests." *The University of Manchester Economics Discussion Paper Series EDP-1204*, 1–27.

- Creswell, J. David, William T. Welch, Shelley E. Taylor, David K. Sherman, Tara L. Gruenewald, and Traci Mann. 2005. "Affirmation of Personal Values Buffers Neuroendocrine and Psychological Stress Responses." *Psychological Science* 16 (11): 846–51.
- Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." Journal of Economic Literature 47 (2): 448–74.
- Dean, Emma Boswell, Frank Schilbach, and Heather Schofield. 2018. "Poverty and Cognitive Function." In *The Economics of Poverty Traps*, edited by Christopher B. Barret, Michael R. Carter, and Jean-Paul Chavas, 57–118. University of Chicago Press.
- Dean, Joshua T. 2024. "Noise, Cognitive Function, and Worker Productivity." *American Economic Journal: Applied Economics* 16 (4): 322–60.
- Dechenaux, Emmanuel, Dan Kovenock, and Roman M. Sheremeta. 2015. "A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments." *Experimental Economics* 18 (4): 609–69.
- DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2): 315–72.
- Dieckmann, Ute, Maarit Thiem, Erik Dirkx, and Jennifer Hays, eds. 2014. *Scraping the Pot: San in Namibia Two Decades after Independence*. Windhoek, Namibia: Land, Environment and Development Project of the Legal Assistance Centre and Desert Research Foundation of Namibia.
- Dutcher, E. Glenn, Loukas Balafoutas, Florian Lindner, Dmitry Ryvkin, and Matthias Sutter. 2015. "Strive to Be First or Avoid Being Last: An Experiment on Relative Performance Incentives." *Games and Economic Behavior* 94: 39–56.
- Escobar-Soler, Carolang, Raúl Berrios, Gabriel Peñaloza-Díaz, Carlos Melis-Rivera, Alejandra Caqueo-Urízar, Felipe Ponce-Correa, and Jerome Flores. 2024. "Effectiveness of Self-Affirmation Interventions in Educational Settings: A Meta-Analysis." *Healthcare* 12 (1): 3.
- Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171–78.
- Ghosal, Sayantan, Smarajit Jana, Anandi Mani, Sandip Mitra, and Sanchari Roy. 2022. "Sex Workers, Stigma, and Self-Image: Evidence from Kolkata Brothels." *The Review of Economics and Statistics* 104 (3): 431–48.
- Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey. 2002. "Quantal Response Equilibrium and Overbidding in Private-Value Auctions." *Journal of Economic Theory* 104 (1): 247–72.
- Goyer, J. Parker, Julio Garcia, Valerie Purdie-Vaughns, Kevin R. Binning, Jonathan E. Cook, Stephanie L. Reeves, Nancy Apfel, Suzanne Taborsky-Barba, David K. Sherman, and Geoffrey L. Cohen. 2017. "Self-Affirmation Facilitates Minority Middle Schoolers' Progress along College Trajectories." *Proceedings of the National Academy of Sciences* 114 (29): 7594–99.

- Greiner, Ben. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association* 1 (1): 114–25.
- Hall, Crystal C., Jiaying Zhao, and Eldar Shafir. 2014. "Self-Affirmation Among the Poor: Cognitive and Behavioral Implications." *Psychological Science* 25 (2): 619–25.
- Haushofer, Johannes, and Ernst Fehr. 2014. "On the Psychology of Poverty." *Science* 344 (6186): 862–67.
- Haushofer, Johannes, Robert Mudida, and Jeremy P. Shapiro. 2020. "The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-Being." NBER Working Paper No. 28106, 1–39.
- Haushofer, Johannes, and Daniel Salicath. 2023. "The Psychology of Poverty: Where Do We Stand?" *Social Philosophy and Policy* 40 (1): 150–84.
- Herrmann, Benedikt, and Henrik Orzen. 2008. "The Appearance of Homo Rivalis: Social Preferences and the Nature of Rent Seeking." *CeDEx Discussion Paper No. 2008–10*, 1–40.
- Hong, Fuhai, Tanjim Hossain, and John A. List. 2015. "Framing Manipulations in Contests: A Natural Field Experiment." *Journal of Economic Behavior & Organization* 118: 372– 82.
- Iriberri, Nagore, and Pedro Rey-Biel. 2017. "Stereotypes Are Only a Threat When Beliefs Are Reinforced: On the Sensitivity of Gender Differences in Performance under Competition to Information Provision." *Journal of Economic Behavior & Organization* 135: 99–111.
- Jann, Ben. 2021. "ROBREG: Stata Module Providing Robust Regression Estimators." https://ideas.repec.org/c/boc/bocode/s458931.html.
- John, Anett, and Kate Orkin. 2022. "Can Simple Psychological Interventions Increase Preventive Health Investment?" *Journal of the European Economic Association* 20 (3): 1001–47.
- Kahneman, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–91.
- Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach. 2024. "Do Financial Concerns Make Workers Less Productive?" *Quarterly Journal of Economics* Advanced access online publication: 1–55.
- Klein, William M. P., Heather K. Blier, and Aura M. Janze. 2001. "Maintaining Positive Self-Evaluations: Reducing Attention to Diagnostic but Unfavorable Social Comparison Information When General Self-Regard Is Salient." *Motivation and Emotion* 25 (1): 23– 40.
- Kong, Xiaojing. 2008. "Loss Aversion and Rent-Seeking: An Experimental Study." *CeDEx Discussion Paper No. 2008–13*, 1–32.

- Kovaleva, Anastassiya. 2012. The IE-4: Construction and Validation of a Short Scale for the Assessment of Locus of Control. Cologne: GESIS Leibniz-Institut für Sozialwissenschaften.
- Kremer, Michael, Gautam Rao, and Frank Schilbach. 2019. "Behavioral Development Economics." In *Handbook of Behavioral Economics: Foundations and Applications 2*, edited by Douglas Bernheim, Stefano DellaVigna, and David Laibson. Vol. 2. Amsterdam: North-Holland.
- Laajaj, Rachid, and Karen Macours. 2021. "Measuring Skills in Developing Countries." Journal of Human Resources 56 (4): 1254–95.
- Larsen, Vance J., Riona Carriaga, Hilary Wething, Jiaying Zhao, and Crystal C. Hall. 2023. "Behavioral Consequences of Income and Expense Shocks." *Journal of Behavioral and Experimental Economics* 106 (102067): 1–10.
- List, John A., Azeem M. Shaikh, and Yang Xu. 2019. "Multiple Hypothesis Testing in Experimental Economics." *Experimental Economics* 22 (4): 773–93.
- Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. "Poverty Impedes Cognitive Function." *Science* 341 (6149): 976–80.
- Martens, Andy, Michael Johns, Jeff Greenberg, and Jeff Schimel. 2006. "Combating Stereotype Threat: The Effect of Self-Affirmation on Women's Intellectual Performance." *Journal of Experimental Social Psychology* 42 (2): 236–43.
- Masiliunas, Aidas, Friederike Mengel, and J. Philipp Reiss. 2014. "Behavioral Variation in Tullock Contests." *KIT Working Paper Series in Economics No. 55*, 1–35.
- McKenzie, David, Aakash Mohpal, and Dean Yang. 2022. "Aspirations and Financial Decisions: Experimental Evidence from the Philippines." *Journal of Development Economics* 156 (102846).
- Miyake, Akira, Lauren E. Kost-Smith, Noah D. Finkelstein, Steven J. Pollock, Geoffrey L. Cohen, and Tiffany A. Ito. 2010. "Reducing the Gender Achievement Gap in College Science: A Classroom Study of Values Affirmation." *Science* 330 (6008): 1234–37.
- Mullainathan, Sendhil, and Eldar Shafir. 2013. *Scarcity: Why Having Too Little Means so Much.* New York: Times Books, Henry Holt and Company.
- Namibia Statistics Agency. 2021. "Namibia Multidimensional Poverty Index (MPI) Report 2021." http://www.npc.gov.na/wp-content/uploads/2022/09/Namibia_ Multidimensional_Poverty_Index_MPI_Report_2021.pdf.
- Orkin, Kate, Rob Garlick, Mahreen Mahmud, Richard Sedlmayr, Johannes Haushofer, and Stefan Dercon. 2023. "Aspiring to a Better Future: Can a Simple Psychological Intervention Reduce Poverty?" *NBER Working Paper No. 31735*, 1–100.
- Palfrey, Thomas R., and Stephanie W. Wang. 2009. "On Eliciting Beliefs in Strategic Games." *Journal of Economic Behavior & Organization* 71 (2): 98–109.

- Parco, James E., Amnon Rapoport, and Wilfred Amaldoss. 2005. "Two-Stage Contests with Budget Constraints: An Experimental Study." *Journal of Mathematical Psychology* 49 (4): 320–38.
- Peiró, José M., Sonia Agut, and Rosa Grau. 2010. "The Relationship between Overeducation and Job Satisfaction among Young Spanish Workers: The Role of Salary, Contract of Employment, and Work Experience." *Journal of Applied Social Psychology* 40 (3): 666–89.
- Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature* 36 (1): 11–46.
 - ——. 2013. "An Approach to Incorporating Psychology into Economics." *American Economic Review: Papers & Proceedings* 103 (3): 617–22.
- Republic of Namibia Central Bureau of Statistics. 2008. "A Review of Poverty and Inequality in Namibia." https://d3rp5jatom3eyn.cloudfront.net/cms/assets/documents/ p19dmr33n31i1b1ko1f0016hg1jdq1.pdf.
- Republic of Namibia National Planning Commission. 2015. "Namibia Poverty Mapping." Windhoek, Namibia. https://www.undp.org/sites/g/files/zskgke326/files/migration/na/ undp_na_povred_npcpovmaprep_2015.pdf.
- Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms." *Science* 370 (6522): 1–12.
- Riley, Emma. 2024. "Role Models in Movies: The Impact of Queen of Katwe on Students' Educational Attainment." *The Review of Economics and Statistics* 106 (2): 334–51.
- Robins, Richard W., Holly M. Hendin, and Kali H. Trzesniewski. 2001. "Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale." *Personality and Social Psychology Bulletin* 27 (2): 151–61.
- Rockenbach, Bettina, Sebastian Schneiders, Sebastian Tonke, Björn Vollan, and Arne R. Weiss. 2024. "Self-Affirmation and Productivity: Backfiring Among Those Who Could Benefit the Most." *Working Paper*.
- Rockenbach, Bettina, Sebastian Schneiders, and Marcin Waligora. 2018. "Pushing the Bad Away: Reverse Tullock Contests." *Journal of the Economic Science Association* 4 (1): 73–85.
- Rockenbach, Bettina, and Marcin Waligora. 2016a. "Beliefs and Behavior in Tullock Contests." *Working Paper*.
- ——. 2016b. "On the Reluctance to Play Best Response in Tullock Contests." *Working Paper*.
- Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan. 2016. "The Psychological Lives of the Poor." American Economic Review: Papers & Proceedings 106 (5): 435– 40.

- Schlag, Karl H., James Tremewan, and Joël J. van der Weele. 2015. "A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs." *Experimental Economics* 18 (3): 457–90.
- Schmitt, Pamela, Robert Shupp, Kurtis Swope, and John Cadigan. 2004. "Multi-Period Rent-Seeking Contests with Carryover: Theory and Experimental Evidence." *Economics of Governance* 5 (3): 187–211.
- Schofield, Timothy P., and Peter Butterworth. 2018. "Are Negative Community Attitudes toward Welfare Recipients Associated with Unemployment? Evidence from an Australian Cross-Sectional Sample and Longitudinal Cohort." Social Psychological and Personality Science 9 (5): 503–15.
- Schwarz, Norbert, Herbert Bless, Fritz Strack, Gisela Klumpp, Helga Rittenauer-Schatka, and Annette Simons. 1991. "Ease of Retrieval as Information: Another Look at the Availability Heuristic." *Journal of Personality and Social Psychology* 61 (2): 195–202.
- Schwarz, Norbert, and Leigh Ann Vaughn. 2002. "The Availability Heuristic Revisited: Ease of Recall and Content of Recall as Distinct Sources of Information." In *Heuristics and Biases*, edited by Thomas Gilovich, Dale Griffin, and Daniel Kahneman, 103–19. Cambridge University Press.
- Selten, Reinhard. 1998. "Axiomatic Characterization of the Quadratic Scoring Rule." *Experimental Economics* 1 (1): 43–61.
- Serra-Garcia, Marta, Karsten T. Hansen, and Uri Gneezy. 2020. "Can Short Psychological Interventions Affect Educational Performance? Revisiting the Effect of Self-Affirmation Interventions." *Psychological Science* 31 (7): 865–72.
- Shapiro, Jenessa R., Amy M. Williams, and Mariam Hambarchyan. 2013. "Are All Interventions Created Equal? A Multi-Threat Approach to Tailoring Stereotype Threat Interventions." *Journal of Personality and Social Psychology* 104 (2): 277–88.
- Sheremeta, Roman M. 2010. "Experimental Comparison of Multi-Stage and One-Stage Contests." *Games and Economic Behavior* 68 (2): 731–47.
- ———. 2013. "Overbidding and Heterogeneous Behavior in Contest Experiments." Journal of Economic Surveys 27 (3): 491–514.
- Sherman, David K., and Geoffrey L. Cohen. 2006. "The Psychology of Self-Defense: Self-Affirmation Theory." *Advances in Experimental Social Psychology* 38:183–242.
- Sherman, David K., Mohini Lokhade, Tim Müller, and Geoffrey M. Cohen. 2021. "Self-Affirmation Interventions." In *Handbook of Wise Interventions: How Social Psychology Can Help People Change*, edited by Gregory M. Walton and Alia J. Crum, 63–99. New York: Guilford Press.
- Smith, Adam. 1792. The Theory of Moral Sentiments; or, An Essay towards an Analysis of the Principes by Which Men Naturally Judge Concerning the Conduct and Character, First of Their Neighbours, and Afterwards of Themselves. To Which Is Added, a Dissertation on the Origin of Languages. 7th ed. Vol. 1. London: A. Strahan, T. Cadell.

- Sommet, Nicolas, David L Weissman, Nicolas Cheutin, and Andrew J Elliot. 2023. "How Many Participants Do I Need to Test an Interaction? Conducting an Appropriate Power Analysis and Achieving Sufficient Power to Detect an Interaction." Advances in Methods and Practices in Psychological Science 6 (3): 1–21.
- Sonnemans, Joep, Arthur Schram, and Theo Offerman. 1998. "Public Good Provision and Public Bad Prevention: The Effect of Framing." *Journal of Economic Behavior & Organization* 34 (1): 143–61.
- Steele, Claude M. 1988. "The Psychology of Self-Affirmation: Sustaining the Integrity of the Self." *Advances in Experimental Social Psychology* 21: 261–302.
- Sullivan, Sian, and Welhemina Suro Ganuses. 2020. "Understanding Damara / ‡Nūkhoen and ||Ubun Indigeneity and Marginalisation in Namibia." In "Neither Here nor There": Indigeneity, Marginalisation and Land Rights in Post-Independence Namibia, edited by Willem Odendaal and Wolfgang Werner. Windhoek, Namibia: Legal Assistance Centre (LAC).
- Suzman, James. 2001. An Assessment of the Status of the San in Namibia. Windhoek, Namibia: Legal Assistance Centre (LAC).
- Szidarovszky, Ferenc, and Koji Okoguchi. 1997. "On the Existence and Uniqueness of Pure Nash Equilibrium in Rent-Seeking Games." *Games and Economic Behavior* 18: 135– 40.
- Thaler, Richard H. 2016. "Behavioral Economics: Past, Present, and Future." *American Economic Review* 106 (7): 1577–1600.
- Tullock, Gordon. 1980. "Efficient Rent-Seeking." In Toward a Theory of the Rent-Seeking Society, edited by James Buchanan, Robert Tollison, and Gordon Tullock, 97–112. Texas University Press.
- Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncentainty* 5: 297–323.
- Walton, Gregory M. 2014. "The New Science of Wise Psychological Interventions." *Current Directions in Psychological Science* 23 (1): 73–82.
- Walton, Gregory M., and Timothy D. Wilson. 2018. "Wise Interventions: Psychological Remedies for Social and Personal Problems." *Psychological Review* 125 (5): 617–55.
- Wilcox, Rand R. 2017. Introduction to Robust Estimation and Hypothesis Testing. 4th ed. Elsevier Academic Press.
- Wood, Joanne V., W.Q. Elaine Perunovic, and John W. Lee. 2009. "Positive Self-Statements: Power for Some, Peril for Others." *Psychological Science* 20 (7): 860–66.
- Yohai, Victor J. 1987. "High Breakdown-Point and High Efficiency Robust Estimates for Regression." *The Annals of Statistics* 15 (2): 642–56.

List of Applied Software

Stata 18.0	Data analysis in all chapters
Qualtrics	Data collection in Chapters 2 and 3
Prolific	Recruitment of participants in Chapters 2 and 3
zTree	Data collection in Chapter 4
ORSEE	Recruitment of participants in Chapter 4
G*Power	Power calculations in Chapter 4

Data Availability

The experimental programs, data, and code used to produce the results in this dissertation are stored at the chair of Prof. Dr. Bettina Rockenbach and are available upon request.