UNIVERSITÉ DU LUXEMBOURG

PhD-FHSE-2025-029
The Faculty of Humanities, Education and Social Sciences

DISSERTATION

Defence held on 25 September 2025 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN ***SCIENCES SOCIALES***

by

Nathalie MÜLLER
Born on 20 January 1992 in Goch (Germany)

Social norms and peer effects in health behaviors

Dissertation defence committee

Dr Marc Suhrcke, dissertation supervisor
Research Program Leader, Luxembourg Institute of Socio-economic Research
Professor, University of York

Dr Francesco Pio Fallucchi
Professor, Università degli studi di Bergamo

Dr Ernesto Reuben
Professor, New York University Abu Dhabi

Dr Silvia Angerer
Professor, UMIT TIROL

Dr Conchita D'Ambrosio, Vice Chairman
Professor, Université du Luxembourg

Affidavit / Statement of originality

I declare that this thesis:

- is the result of my own work. Any contribution from any other party, and any use of generative artificial intelligence technologies have been duly cited and acknowledged;
- is not substantially the same as any other that I have submitted, and;
- is not being concurrently submitted for a degree, diploma or other qualification at the University of Luxembourg or any other University or similar institution except as specified in the text.

With my approval I furthermore confirm the following:

- I have adhered to the rules set out in the University of Luxembourg's Code of Conduct and the Doctoral Education Agreement (DEA)[1], in particular with regard to Research Integrity.
- I have documented all methods, data, and processes truthfully and fully.
- I have mentioned all the significant contributors to the work.
- I am aware that the work may be screened electronically for originality.

I acknowledge that if any issues are raised regarding good research practices based on the review of the thesis, the examination may be postponed pending the outcome of any investigation of such issues. If a degree was conferred, any such subsequently discovered issues may result in the cancellation of the degree.

Approved on 2025-07-28

[1] If applicable (DEA is compulsory since August 2020)

Acknowledgments

I am done. This dissertation marks the end of a four-year journey, during which I have changed more than I could have anticipated. I owe a great deal to the many people who supported me along the way and helped me stay afloat during the more difficult moments.

First and foremost, I thank my doctoral advisor, Marc Suhrcke, for supervising this work. Your consistently constructive yet critical feedback guided me at many points in this dissertation, while your support and trust gave me the freedom to shape it into something of my own. Our path together was not without detours, but looking back, those moments were as formative as the more straightforward ones. I also thank Francesco Fallucchi for his academic supervision, for the opportunities and guidance, and for brightening the path with his humor and energy. Thanks to his leaving Luxembourg almost the instant I arrived, I unexpectedly found myself, two years later, spending four of the most rewarding months of my life in Bergamo.

I am deeply grateful to the colleagues and friends who turned this experience into something far more enjoyable than I had expected. Thank you for your ideas, encouragement, and (for better or worse) your conspiratorial spirit when it came to pushing research questions a little too far. This community made the difficult times more bearable and the good times better still. Beyond academia, I have been lucky to have friends who kept me grounded. Many of you deserve a mention, but far more space (and time) would be needed to do you justice.

I am grateful to my family for their support and understanding, especially in these past months leading up to today. Perhaps no one was more surprised than you when I first announced I was setting out on this path, yet you stood by me through it all. I am particularly thankful to my brother, Kai: you have been an immense support and a loyal late-night companion. Thank you for all those very educational conversations on my nighttime walks, and for so much more.

I also thank the participants who made this research possible. One in particular shared their story with an openness that deeply moved me and reminded me again why I chose this path.

A special thanks goes to Jolt Coffee Roasters and their berry berry good filter coffee, which carried me through the seemingly never-ending summer of thesis writing and allowed this journey to end on a very tasteful note.

And, as always, I saved the best for last.

Finally, to the people who walked beside me: this PhD would have been an entirely different story without you. You gave me a place where I could belong without condition, quirks included. You are at the heart of my love for food, for weeb culture, and for all things cute. You have reminded me, again and again, that the best parts of this journey often happen somewhere between a cup of coffee and the plans we still have not gotten around to. You knew when I needed distraction and when I needed direction. You have sparked ideas I might never have reached alone, and you have grounded me with your honesty and perspective when I needed it most.

Some of you have been with me since day one, and you are still here now. Your friendship, kindness, and unconditional support have meant more to me than I can ever put into words.

Contents

Abstract

What motivates people to make health-related choices that benefit both themselves and society, even when doing so comes at a personal cost? Preventive behaviors such as exercising, getting vaccinated, or seeking timely medical support are essential to individual well-being and to the effective functioning of health systems. Much of the existing research has examined how intrinsic motivations and material incentives influence these choices. This thesis instead investigates how preferences and decisions are shaped by the social environments in which people live, where what others do, expect, and think can be as influential as considerations concerning money, time, or convenience. We know far less about the specific ways these influences operate in health-related decisions and how they interact with individual incentives. Prior evidence suggests that interventions drawing on social norms can be effective, but their effectiveness depends heavily on context – that is, they must be tailored to the right conditions. This thesis addresses that gap by examining three distinct but related channels through which social context influences decisions that have a measurable effect on individual health outcomes.

The first chapter draws on a systematic review and meta-regression of 45 studies to assess peer effects in weight-related behaviors among young people. The findings suggest social norms are the most consistent driver of behavioral adaptation, while other mechanisms, such as learning or peer comparison, appear less often and are rarely tested explicitly across selected studies. Building on this, the second chapter shifts its focus to social comparison processes, and turns to positional concerns, using survey experiments to examine how people trade absolute outcomes for relative advantage in health and economic contexts. The results suggest that social preferences play an important role in driving status motives, more so than socioeconomic or demographic factors. The third chapter examines the question of identity disclosure under stigma, drawing on a randomized online experiment with adults who self-report ADHD. It shows that the willingness to disclose is influenced more by perceptions of fairness and expectations of acceptance than by the framing of the disclosure information, even when disclosure carries not only a social risk, but also risks the loss of pecuniary rewards.

These studies reveal there is no single form of "social influence", but rather a set of several social factors, each with its own way of affecting decisions. They demonstrate that health decisions are made in social environments where norms, status, and image concerns can weigh as heavily as direct and tangible costs or benefits. Recognizing that there are different social mechanisms at work, possibly as influential as economic ones, demonstrates once more that it is crucial for future policy design to achieve more sustainable improvements in public health.

Co-author statement

This dissertation includes both single-authored and co-authored works, with the latter led by me as first author. My co-authors contributed through guidance on study design, statistical analysis, and interpretation of results, as well as providing critical feedback and revisions throughout the writing process. All co-authors reviewed and approved the final versions of the manuscripts. To ensure transparency, I detail below my specific contributions to each chapter.

Chapter 1. Peer effects in weight-related behaviors of young people – A systematic literature review. Co-authored with Francesco Fallucchi (University of Bergamo) and Marc Suhrcke (Luxembourg Institute of Socio-Economic Research and University of York).
I developed the original research question, and conceptualized and designed the systematic literature review. Before conducting the review, I was responsible for authorship and pre-registration of the protocol for this review. I conducted the literature search and coding, performed the meta-regression analysis, produced all visualizations, and drafted the manuscript. My co-authors contributed to the methodological design, and critical revisions of the paper.

Chapter 2. Social comparison and positional preferences in health and economic behavior: Experimental evidence from the US and UK. Co-authored with Francesco Fallucchi (University of Bergamo), Francesco Principe (University of Bergamo) and Marc Suhrcke (Luxembourg Institute of Socio-Economic Research and University of York).
This paper was developed during a research stay at the University of Bergamo (October 2023 – February 2024), which facilitated close collaboration with my co-authors and supported the study's conceptualization and experimental design. I was responsible for the pre-registration, programming of the experiment in Qualtrics, and management of data collection via Prolific, with guidance and active support from Francesco Fallucchi. I also prepared and analyzed the data, produced all figures and tables, and wrote the manuscript. All co-authors contributed to the design, interpretation, and provided critical feedback on the paper.

Chapter 3. The hidden struggle: Navigating disclosure, social backlash, and incentives – experimental evidence from adults with ADHD. This chapter is single-authored and serves as my job market paper. I designed, preregistered and implemented the experiments, obtained ethical approval, collected and analyzed the data, and wrote the manuscript independently.

General introduction

General introduction

Individual decision-making rarely occurs in isolation. Much of what we choose, how to behave, or even what to believe, is determined by social context and critically hinges on what we observe and learn from those around us. When norms are ambiguous or information is scarce, we often look to others for cues about what is appropriate, desired, or expected, even when following those cues comes at a personal cost or counteracts our own best interests. Traditional economic models typically treat decision-making as a function of stable preferences and material incentives. However, a growing body of work shows that social context matters just as much: people compare themselves to others, are averse to inequality, and care about fairness (Fehr and Schmidt, 1999); they also adjust their behavior in response to what they think others expect (Akerlof and Kranton, 2010; Bénabou and Tirole, 2006; Carpenter and Robbett, 2024). These influences extend beyond close relationships. Proximate peers like friends and work colleagues shape behavior, but so do distant and public figures, especially in online environments. They serve as behavioral models and reference points, ultimately triggering behavioral spillovers and adaptations across a wide range of behaviors: eating habits, exercise routines, and sleep patterns shift, aligning with those of peers; elective procedures or branded medications may be chosen not for medical reasons but to signal youth or status; and, on the other end of the spectrum, some people may feel the need to conceal stigmatized traits and identities to avoid judgment and disapproval, even if disclosure would lead to support.

Ultimately, it is perhaps not surprising that social influence is particularly potent in health-related behaviors and outcomes. Whenever we observe such interdependencies, we speak of *peer effects* (Boucher *et al.*, 2024). In domains like nutrition, preventive care, and treatment decisions, individuals often lack clear or reliable information about risks and benefits or long-term consequences. Thus, when people are uncertain about health matters yet face consequential decisions, they naturally look to the behavior and experience of others, seeking guidance. At the same time, many behaviors are observable and leave room for social evaluation. Individuals may conform to what their peers or their communities deem socially acceptable, not only out of uncertainty but also to avoid social disapproval or to present a socially desirable image. Observable actions can serve as signals, and people may choose

behaviors that reinforce their identity or status within a group (Bursztyn and Jensen, 2017; Bursztyn *et al.*, 2025). Such signals have been shown to impact a range of behaviors from exercise and preventive screenings to vaccination and medication uptake. For example, recent studies find that immunization and medication uptake increase when actions are made more publicly visible (for example, via visibly wearing bracelets) and that individuals derive utility from signaling their commitment to health (Jee, Karing and Naguib, 2024; Karing, 2024). Conversely, this could imply that a person might forgo a useful treatment not because it is ineffective, but because of how it might be perceived.

One of the most influential studies on peer effects in health behaviors is the work by Christakis and Fowler (2007), which tracked the spread of obesity over 32 years in a densely connected social network. The authors concluded that these spillovers were driven by a diverse set of mechanisms, and not merely by the social environments their study subjects shared. By demonstrating that social influence, whether through behavioral imitation or social norms, has a measurable impact on health outcomes, their study stimulated numerous studies on the matter. Today, it is irrefutable that, whatever the mechanism, health behaviors have a tendency to spread through social networks in a way that looks a lot like contagion, highlighting the role of peers and the influence of social incentives[1]. And far from slowing down, this interest in studying how people influence one another continues to expand, with research activity increasing in recent years not just to understand individual behavior, but also to design effective policies that aim to shift it. Much of the economics literature has moved in this direction, and offers insights into individual decision-making and its dependence on social context and the behavior of peers (Akerlof and Kranton, 2000; Bénabou and Tirole, 2006; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Fehr and Schmidt, 1999; Sacerdote, 2014), illustrating the importance of social interactions.

This dissertation is motivated by this growing literature on how behavior is driven by social norms, identity signaling, and social comparisons. I study these themes across three empirical papers, each of which uses a different methodological approach (a systematic review and meta-regression, experiments, and a survey design) to address how social context affects health-related decisions. I focus in particular on how people respond to social incentives and

[1]Balsa and Díaz (2018) provide an extensive review of the economics literature from the past two decades, unpacking the magnitude and drivers of peer effects in health behaviors.

constraints, such as concerns about status, image, or stigma, when making choices that bear on their well-being, identity, or economic outcomes. Although these mechanisms and their outcomes often overlap, disentangling them helps clarify what drives behavior and what might make interventions more effective toward more sustainable behavioral change. In the first chapter, I therefore make an effort to understand the underlying mechanisms, by providing a systematic review of peer effects and an investigation of how they operate, highlighting their theoretical distinctions and behavioral implications.

How do peer effects affect and operate in health-related behaviors?

Chapter 1 synthesizes cross-disciplinary evidence from economics, psychology, and health to examine how peers shape weight-related behaviors among young people. Drawing on a systematic literature review of 45 studies and a meta-regression, we assess both the magnitude of peer effects and the mechanisms underlying them. Studies are systematically reviewed and classified by the motives proposed to explain peer effects (e.g., social learning or descriptive norms). The majority of studies indicate social norms as the most important driver of behavioral adaptation among peers, suggesting that young people adjust their behaviors to align with what they perceive as appropriate or common. While few studies attempt to disentangle these motivations directly, and many only touch on them implicitly through study design or framing, alternative explanations such as social comparison or learning are much less frequently examined and often not even discussed. We also find that several of the selected studies do not explicitly develop or test specific motivations of peer influence.

While evidence from the reviewed studies on motives other than social norms was visibly limited, Chapter 2 picks up where Chapter 1 leaves off, shifts the attention to comparisons of health from a broader, more comparative perspective, and investigates positional concerns in the context of various hypothetical scenarios, thus turning to the following research question:

How do concerns about relative position and social comparison influence decision-making in health and economic contexts?

Using an experimental design adapted from Solnick and Hemenway (1998), I investigate how individuals trade off absolute outcomes for better relative standing across health domains, from physical activity to healthcare. In economics, it is well established that people care about relative income or status (in an attempt to "keep up with the Joneses"). For example, Luttmer (2005) finds that an increase in one's neighbor's earnings reduces one's own self-reported happiness, roughly as much as an equivalent drop in one's own income would. Luttmer attributes this to what he calls 'spiteful egalitarianism', that is, people derive disutility from lagging behind their peers. In this paper, I follow more recent literature in referring to this preference as envy (Diaz *et al.*, 2023; Kerschbamer, 2015)[2].

This chapter complements the findings from Chapter 1. Whereas the first chapter emphasizes norm adherence and peer expectations as drivers of behavioral adaptation, Chapter 2 turns to a related, yet distinct concept: the desire to improve one's relative position compared to one's peers rather than simply conforming to the prevalent status quo. Together, these chapters map out different channels through which social context influences individual behavior, from conformity to norms to status concerns. Lastly, Chapter 3 takes this thesis on social norms and peer effects one step further, turning from hypothetical vignettes to the question of how people navigate identity disclosure when their choices are informed not just by social comparison but by reputational concerns and the risk of stigma. By exposing participants to a real rejection risk with economic stakes, the last chapter addresses the following question:

How do individuals with stigmatized identities navigate disclosure under social and reputational incentives?

In Chapter 3 I examine how adults with self-reported ADHD decide whether to reveal their neurodivergent status in exchange for a performance advantage, when doing so might put

[2]Note that this notion of individual utility decreasing as other people's incomes increase is also consistent with what other scholars describe as 'relative income effect (RIE)' (Ifcher *et al.*, 2020), 'spiteful' (Levine, 1998) or 'competitive' (Charness and Rabin, 2002).

them at risk of rejection from a matched partner. Participants were randomly assigned to one of six treatment arms that manipulated both the severity of potential backlash (none, low, or high) and the framing of ADHD (as diagnostic or identity-focused).

While disclosure rates remained relatively stable across experimental conditions, the results show that two factors did indeed predict disclosure. First, participants' fairness perceptions mattered: that is, those who viewed the receipt of accommodations or conditional benefits as fair were significantly more likely to disclose their ADHD status, regardless of framing. Second, disclosure was influenced by participants' *second-order beliefs* – that is, their expectations about how others perceive neurodivergence. Those who anticipated greater acceptance were more likely to disclose, particularly under the high-stakes condition where rejection entailed the potential for monetary loss. Interestingly, participants with ADHD anticipated greater acceptance than their non-neurodivergent matches: they expected less discrimination, which might have explained why disclosure remained high even when rejection was costly. When people believe others will treat them fairly, they are more willing to open up about stigmatized aspects of who they are.

This dissertation consists of three self-contained essays. Each can be read on its own and provides different perspectives on the possible drivers of social influence. Each chapter concludes with its own bibliography and appendices containing supplementary material.

Bibliography

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *The Quarterly Journal of Economics*, 115(3): 715–753.

Akerlof, George A, and Rachel E Kranton. 2010. "Identity economics: How our identities shape our work, wages, and well-being." In *Identity Economics*. Princeton University Press.

Balsa, Ana Inés, and Carlos Díaz. 2018. "Social interactions in health behaviors and conditions." *Documentos de trabajo del Departamento de Economía*.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and prosocial behavior." *American Economic Review*, 96(5): 1652–1678.

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

Boucher, Vincent, Michelle Rendall, Philip Ushchev, and Yves Zenou. 2024. "Toward a general theory of peer effects." *Econometrica*, 92(2): 543–565.

Bursztyn, Leonardo, and Robert Jensen. 2017. "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure." *Annual Review of Economics*, 9(1): 131–153.

Bursztyn, Leonardo, Ingar K Haaland, Nicolas Röver, and Christopher Roth. 2025. "The Social Desirability Atlas." *National Bureau of Economic Research*.

Carpenter, Jeffrey, and Andrea Robbett. 2024. "Measuring socially appropriate social preferences." *Games and Economic Behavior*, 147: 517–532.

Charness, Gary, and Matthew Rabin. 2002. "Understanding social preferences with simple tests." *The Quarterly Journal of Economics*, 117(3): 817–869.

Christakis, Nicholas A, and James H Fowler. 2007. "The spread of obesity in a large social network over 32 years." *The New England Journal of Medicine*, 357(4): 370–379.

Diaz, Lina, Daniel Houser, John Ifcher, and Homa Zarghamee. 2023. "Estimating social preferences using stated satisfaction: Novel support for inequity aversion." *European Economic Review*, 155: 104436.

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

Ifcher, John, Homa Zarghamee, Dan Houser, and Lina Diaz. 2020. "The relative income effect: an experiment." *Experimental Economics*, 23(4): 1205–1234.

Jee, Edward, Anne Karing, and Karim Naguib. 2024. "Optimal policy with social image concerns: Experimental evidence from deworming." Working paper.

Karing, Anne. 2024. "Social signaling and childhood immunization: A field experiment in Sierra Leone." *The Quarterly Journal of Economics*, 139(4): 2083–2133.

Kerschbamer, Rudolf. 2015. "The geometry of distributional preferences and a non-parametric identification approach: The Equality Equivalence Test." *European Economic Review*, 76: 85–103.

Levine, David K. 1998. "Modeling altruism and spitefulness in experiments." *Review of Economic Dynamics*, 1(3): 593–622.

Luttmer, Erzo FP. 2005. "Neighbors as negatives: Relative earnings and well-being." *The Quarterly Journal of Economics*, 120(3): 963–1002.

Sacerdote, Bruce. 2014. "Experimental and quasi-experimental analysis of peer effects: two steps forward?" *Annual Review of Economics*, 6(1): 253–272.

Solnick, Sara J, and David Hemenway. 1998. "Is more always better? A survey on positional concerns." *Journal of Economic Behavior & Organization*, 37(3): 373–383.

Peer effects in weight-related behaviors of young people – a systematic literature review

Peer effects in weight-related behaviors of young people – A systematic literature review

1.1 Introduction

Poor health behaviors and the persistently increasing prevalence of obesity across all age groups and world regions have become a burden on health systems worldwide. According to the World Health Organization (2022), a substantial prevalence of overweight and obesity was observed among European children and adolescents, with one in five (21.4%) US children facing obesity challenges in 2016. Moreover, adults experienced even higher prevalence rates, reaching 23.3% in the European Region and 36.2% in the United States. A steady increase in these figures implies major future direct and indirect costs for the individuals affected, as well as for society as a whole. OECD (2019) estimates that OECD countries will, on average, spend 8.4% of their total health care budgets to treat health conditions associated with excess body weight. Obesity is intricately linked to metabolic syndrome, a cluster of modifiable risk factors associated with cardiovascular disease and type 2 diabetes (Després and Lemieux, 2006). Various factors, spanning from behavioral to economic determinants, have emerged as prominent drivers of these health conditions, heightening the risk for severe chronic diseases and posing a significant public health challenge (Rosin, 2008). Despite its growing recognition across diverse academic fields, the association of sleep with physiological and psychological mechanisms driving weight gain and metabolic syndrome is often overlooked in comparison to more visible drivers. While the impact of poor dietary choices and physical inactivity on body weight is well documented, an evolving body of research sheds light on the substantial influence of sleep behaviors on metabolic health and other health-related outcomes (Dulloo, Miles-Chan and Montani, 2017; Grimaldi *et al.*, 2023; Oftedal, Vandelanotte and Duncan,

This chapter is based on a paper published as: Müller, N., Fallucchi, F., & Suhrcke, M. (2024). Peer effects in weight-related behaviours of young people: A systematic literature review. Economics & Human Biology, 53, 101354. DOI: `https://doi.org/10.1016/j.ehb.2024.101354`.
Minor changes in language, formatting, notation, and structure were made to adapt the manuscript to the style of this thesis. Additionally, some revisions to the content were made to the summary tables, including updates to mechanism classifications and clarifications of their empirical testing in the literature.

2019).

Recognizing the steady increases in body size among youth[1] and in an attempt to address this public health challenge, national public health agencies as well as international organizations have actively engaged in the development of action plans and policies to promote healthier lifestyles (Office of Disease Prevention and Health Promotion, 2022; World Health Organization, 2016). Yet, despite these efforts, achieving sustainable changes in people's behavior at scale remains difficult (Swinburn *et al.*, 2019), and there continues to be strong public policy and research interest in the factors contributing to the trend of high and rising obesity levels in young people. In addition to economic, sociodemographic, and convenience factors, social interaction and peer effects stand out as potential explanations for the onset of obesity (Cawley, 2015). Formally, the latter is now described as the tendency of individuals to behave in a certain way based on how common that behavior is among the people they are interacting with (Manski, 1993). Christakis and Fowler (2007) demonstrated in their seminal study the impact of social networks in shaping individual obesity levels, highlighting the importance of social factors in the study of weight-related outcomes. A rapidly expanding body of empirical research followed, reporting positive associations between peer behavior and individual body weight outcomes (Auld, 2011; Cohen-Cole and Fletcher, 2008; Mora and Gil, 2013; Nie, Sousa-Poza and He, 2015; Trogdon, Nonnemaker and Pais, 2008).

While the core idea of peer effects has remained consistent throughout the years, the understanding and study of peer effects have evolved, driven by methodological challenges to provide credible estimates of the magnitude and nature of peer effects. Manski (1993) showed that in standard social interaction models, the pure effect of one individual on another is indistinguishable from contextual or correlated effects. These factors can complicate the identification of the endogenous peer effect, where an individual's outcome is directly affected by the behavior of their peers. One main challenge in measuring peer effects emerges from structural simultaneity inherent in the reciprocal influences of peers, also commonly known as the reflection problem: while peers exert influence on individual outcomes, simultaneous spillovers from individuals onto their respective peer group can occur. Researchers have

[1]As the chapter unfolds, I refer interchangeably to this cohort as "youth" and "young people" to describe individuals aged 15 to 24, aligning with the definitions by United Nations General Assembly (2002) in the General Assembly Resolutions, A/RES/56/117.

adapted various strategies to address this identification issue, ranging from the use of lagged peer outcomes as covariates when examining health behaviors (Ali, Amialchuk and Heiland, 2011; Clark and Lohéac, 2007), to the estimation of instrumental variable regressions (Auld, 2011; Mora and Gil, 2013; Trogdon, Nonnemaker and Pais, 2008), to natural experiments and randomization to identify the causal effect of social interactions (Condliffe, Isgin and Fitzgerald, 2017; Golberstein, Eisenberg and Downs, 2016; Yakusheva, Kapinos and Weiss, 2011). When carefully applied, these models may elucidate the relationship between individual health-related behaviors and those of their peers.

Despite these challenges in the study of peer effects, researchers have made steady progress toward a better understanding of the influence of social interactions. Previous systematic reviews and meta-analyses of empirical research exploring the relationship between social networks and weight-related outcomes have generally supported the hypothesis of peer influence and behavioral adaptation. These studies have shown that weight-related behaviors and outcomes during certain life stages may be associated with the attitudes, behaviors, and interactions of their peers. This can occur through homophily, where individuals with similar attributes and behaviors self-select into friendships, or through social influence, where peers have an impact on each other's decisions and behaviors over time (Chung, Ersig and McCarthy, 2017; Cunningham *et al.*, 2012; Montgomery *et al.*, 2020). While the number of primary and review studies on peer effects keeps growing, important gaps remain. First, existing research strongly focuses on children and adolescents up to the age of 18, while dedicating limited attention to young people beyond 18. Transitioning from adolescence to adulthood marks a decisive developmental period characterized by significant cognitive, behavioral, and social changes that may critically shape lifelong health habits (Elkins, Kassenboehmer and Schurer, 2017; McDade *et al.*, 2011). Second, there is little doubt about the potency of peer influence on individual outcomes, yet credible evidence of the underlying mechanisms is limited. This is partly due to the complexity of studying social networks and the difficulty in isolating the effect of peers from other confounding factors, as described above. Until now, research has proposed different mechanisms underlying peer effects, including conformity (Bernheim, 1994; Jones, 1984), social learning (Bandura, 1977; Bikhchandani, Hirshleifer and Welch, 1992), and social utility (Bursztyn *et al.*, 2014), to name but a few. Any of these mechanisms might be

utilized as a reference by which peers affect individual behavioral outcomes, leading to the adoption of both health-enhancing and health-harming behaviors. Importantly, the generated spillovers in health outcomes from one individual to another create multiplier effects, resulting in larger aggregate effects (Glaeser, Sacerdote and Scheinkman, 2003). Above all, expanding our knowledge about the drivers of peer effects in weight-related behaviors is an important first step to better inform policy, as these may serve as potential entry points for policy intervention. If policymakers were provided with better knowledge of how and through which mechanisms the diffusion of behaviors works, these insights could be used as leverage to improve young people's health trajectories. Hence, the need for (and the benefit from) seeking to disentangle the different models and mechanisms underlying peer effects in young people's weight-related behaviors.

Considering these developments and the current imperative toward a better understanding of the driving factors of behavioral adaptation, this review's key objectives are threefold: to provide a transparent synthesis and critical appraisal of current evidence on social influence in weight-related behaviors, to examine the potential drivers of peer effects, and to derive and discuss policy implications, offering insights for future research and decision-making. While peer effects and their drivers are a central focus of this study, it is important to note that the underlying processes do not operate in isolation; rather, they involve complex dynamics with other individual and demographic factors. For details on the interplay of diet, physical (in)activity, and sleep behaviors with factors not covered in the primary analysis, we provide a conceptual framework (Figure A1.1 in Appendix Section A1.1). This review further enhances its synthesis by incorporating a meta-regression analysis on a subset of the reviewed studies to examine the contribution of study-level characteristics to the heterogeneity among reported research findings. In conclusion, this synthesis emphasizes the need for further research into the underlying mechanisms of peer effects to better inform policymakers in designing effective policies for curbing unhealthy weight-related behaviors in young people.

1.2 Methodology

This systematic review was conducted and reported in adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page *et al.*, 2021). The protocol for this review is preregistered in PROSPERO under registration number CRD42022370974. In the interest of transparency, it is important to note that the meta-regression analysis was not preregistered. Our initial study design did not allow for an a priori specification of the meta-regression parameters. Hence, the decision to perform the meta-regression was based on a subset of results from quantitatively comparable studies. As an additional robustness check, we adopted a methodology similar to that used by Card, Kluve and Weber (2010). For this approach, we categorized the outcomes of all studies based on whether the observed peer effect was statistically significant and positive, or statistically insignificant.

1.2.1 Systematic literature search

We conducted a systematic search for studies that empirically assessed the association between peer effects and young people's weight-related behaviors. In February and March 2022, we searched the Web of Science, EconLit, MEDLINE, and SocIndex databases for studies published from January 2011 through February 2022, targeting peer-reviewed full-text articles in English. Synthesizing evidence from this decade allows us to cover the most recent developments in a dynamic field that undergoes changes over time, including evolving contexts (e.g., the rise of social media) and methodological advancements (e.g., use of wearables for data collection). The search strategy was developed to capture relevant papers covering the following key topics: weight-related health behaviors including diet, activity levels, and sleep, and peer effects among youth. For each database, we included keywords from the main topics, *peer effects* AND *weight-related behaviors* AND *youth*, applied to titles, keywords, and abstracts. The filters applied for the search were year and language of publication. Additionally, we performed manual searches of reference lists and citations. We excluded review articles, conference papers, or commentaries, as well as gray literature such as theses and dissertations.

1.2.2 Study eligibility criteria

Studies were eligible if they examined peer effects on health behaviors, focusing specifically on diet, physical (in)activity, and sleep outcomes within a sample of young people (from 15 to 24 years old). Other behaviors that potentially affect body weight and metabolism, such as stress, alcohol, and tobacco consumption, were excluded. These behaviors have been extensively covered in previous reviews and meta-analyses (Henneberger, Mushonga and Preston, 2021; Leung, Toumbourou and Hemphill, 2014; Wardle *et al.*, 2011). Further, this review was restricted to quantitative studies that employed experimental (e.g., field and laboratory studies) and non-experimental research methods (e.g., cross-sectional and longitudinal designs). Studies addressing social influence on psychopathological behaviors (e.g., eating disorders or compulsive exercise) were outside the scope of this study and were therefore excluded, as were studies that did not provide a direct measure of peers' weight-related behaviors. Note that influence may be exercised through other forms of social interaction, including direct verbal encouragement or displays of contempt, shaping individuals' behaviors through external pressures. While these aspects are important, we prioritize the direct impact of peers' observable health behaviors. To maintain a focused and comprehensive analysis of the resulting peer effects and their underlying mechanisms, we excluded studies evaluating influence through social pressure or coercion, as well as studies measuring social support by combining an array of different social support types (e.g., co-participation in exercise and giving evaluative feedback) into a single global score.

1.2.3 Selection and coding

The study selection process is illustrated in Figure 1.1. The first step in this process was to identify all records through the database search complying with our search strategy, followed by exporting and merging the datasets. After eliminating duplicates, an initial screening of titles and abstracts for obvious irrelevance was performed, which led to the exclusion of the majority of records. Subsequently, full-text articles of the preliminarily selected studies were obtained to assess their eligibility based on the predefined inclusion and exclusion criteria. For all excluded articles, we provided a rationale explaining why they were discarded.

We developed a standardized table to extract the following data from the selected studies: authors, journal and publication year, aims of the study, study characteristics (study design, data, method), sample characteristics (sample size, sociodemographic characteristics), and primary outcomes (peer group, health-related behavioral outcome measures, explanatory and mediating variables, results).

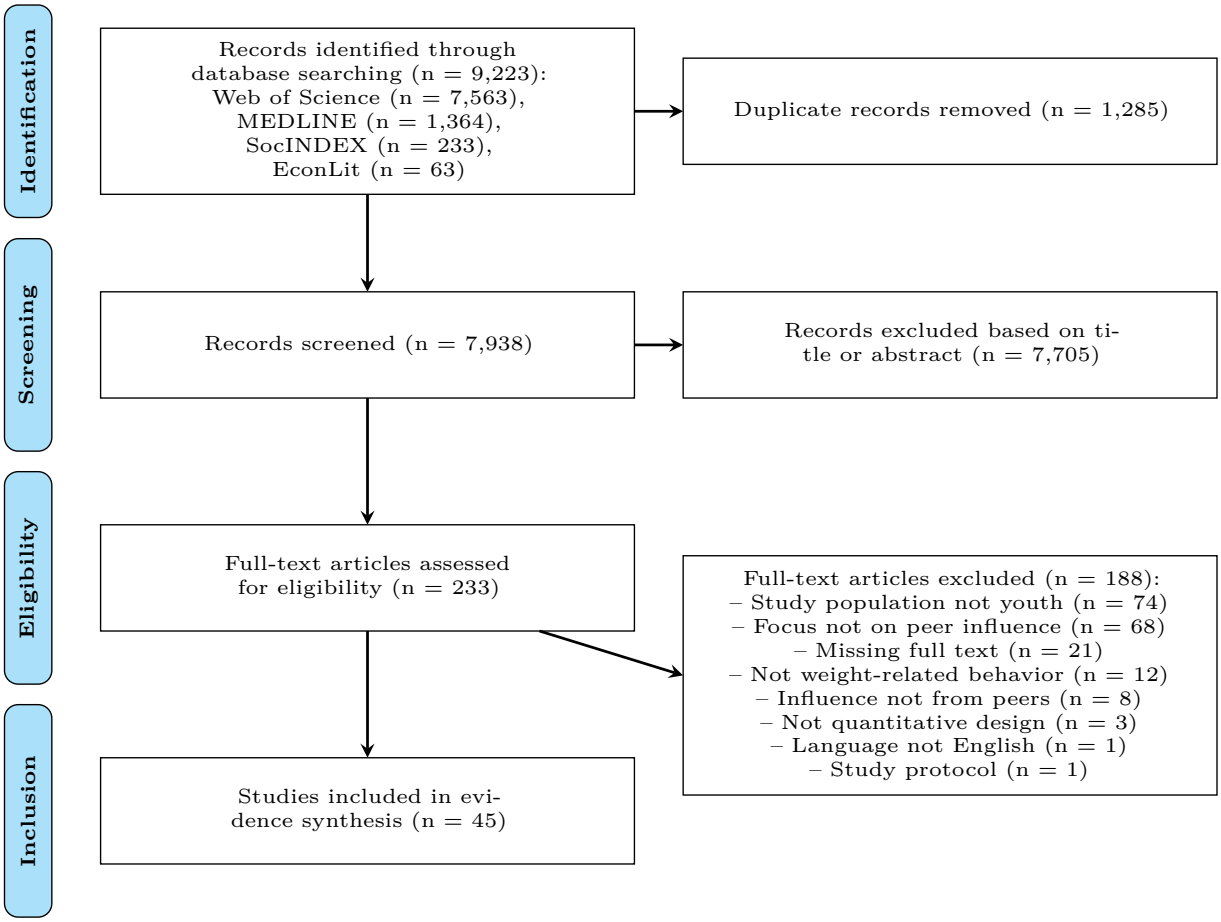


Figure 1.1: Study selection flowchart (PRISMA 2020)

1.2.4 Methodological quality assessment

The methodological quality of the selected studies was assessed using Joanna Briggs Institute (JBI) critical appraisal checklists for analytical cross-sectional and (quasi-)experimental studies (Moola *et al.*, 2020; Tufanaru *et al.*, 2020). We adapted these checklists to allow for consistent study assessment while accommodating the wide range of quantitative study designs. Appendices A1.2, A1.3, and A1.4 provide details about the checklist criteria and

summaries of results for cross-sectional, longitudinal, and experimental studies, respectively. The checklists were used independently by the three researchers, and any ambiguities were resolved through discussion. The quality review comprised a rating of the sample selection, comparability, study design and methodology applied, as well as a rating of the validity and reliability of peer variables and health-related behavioral outcome measures, whereby higher ratings indicated higher study quality. The results from the quality assessment did not determine the final study inclusion.

Note that we did not include the availability of data and code for statistical analyses among the criteria to assess the quality of the studies. The reason is that the great majority of the studies did not provide data, and only one of those did provide the code to replicate the results. We also checked the current state of the replication policies in place in the journals hosting the selected articles and, as of November 2023, none of them have strict requirements about data availability.

1.2.5 Meta-regression procedures

Statistical analyses were performed using the `metafor` package (Viechtbauer, 2010). Effect sizes from 19 eligible studies were pooled and estimated using the restricted maximum likelihood (REML) method. These effect sizes included regression coefficients and were used to evaluate the association between study-specific factors and the reported peer effects. While standardized regression coefficients are rarely employed in economic research (Stanley and Doucouliagos, 2012), there has been a growing focus on their application in meta-analytic syntheses of research in recent years. In the current study, for original studies that only reported unstandardized regression coefficients or did not present standard errors alongside the reported coefficients, conversion and imputation methods based on Nieminen (2022) were implemented to calculate the effect sizes. As an additional robustness check, effects from the full set of study outcomes were categorized into significantly positive and statistically insignificant outcomes to examine the likelihood of reporting positive and significant peer effects across studies and methodologies. Finally, publication bias was assessed by applying Egger's regression test (Egger *et al.*, 1997), a simplified selection model (McShane, Böckenholt and Hansen, 2016), as well as by visually inspecting the symmetry of funnel plots (provided

in the Appendix, Figure A1.4). While these methods do not directly examine publication bias, they aid in identifying asymmetry in the funnel plot, which is potentially attributable to publication bias.

1.3 Results

The initial database search yielded 9,223 articles. After screening titles and abstracts, 233 articles remained for full-text examination. This final record selection process required that either the title or abstract indicate an association between young people's social interactions and their impact on others' weight-related behaviors. Ultimately, we obtained a list of 45 studies deemed eligible for the evidence synthesis (see Figure 1.1 for further details of the search and study selection process). Tables 1.1 – 1.3 list the papers that fit our eligibility criteria for analyzing social influences in different health behaviors and provide a comprehensive overview of the respective study characteristics. In what follows, we summarize the study results and present separate evidence for dietary behaviors, physical (in)activity, and sleep-related decisions. While all studies addressed whether and how others influence individual health-related behaviors and decisions, only a few gave particular attention to the underlying mechanisms; furthermore, studies varied greatly in their methodological approaches to assessing peer influence.

1.3.1 Descriptive characteristics

The database search yielded 13 experimental, 12 longitudinal, and 20 cross-sectional studies. The majority of the studies (n = 25) were based in the United States; 14 studies were conducted in Europe, 3 in Asia, 3 in Brazil, and 1 in Australia. Sample sizes varied from 51 to 8,000, comprising young people aged on average between 15 and 24 years. Twenty-one studies investigated eating behaviors only, such as fruit and vegetable (FV) intake, high-energy-dense (HED) foods, or sugar-sweetened beverages (SSB), with the second-largest group of papers studying physical activity (PA) (n = 17). Three studies examined peer effects on sleep behaviors only, such as bedtime decisions, sleep quality, and quantity. Four studies focused on a set of different weight-related behaviors, studying physical activity combined with diet

(n = 3) or sleep duration (n = 1). The majority of papers used self-reports to measure health outcomes and peer-related variables.

1.3.2 Risk of bias and study quality

Overall, most of the identified studies met at least half of the study assessment criteria set by the JBI critical appraisal checklists for analytical cross-sectional and (quasi-)experimental studies. Tables A1.2 – A1.4 in the Appendix provide more detailed information about the ratings of the studies. Studies were considered to be of good quality when they met more than 80% of the assessment criteria, and studies were ranked as low quality when they met less than 60% of the assessment criteria. Most of the studies used self-reports to assess the outcome of interest, which may have compromised the validity and reliability of these measures. Yet some studies did not adjust for potential confounders in their data analysis or showed a lack of description of effect size in statistical tests. Some experimental studies were also found to be at risk of selection bias. In these studies, study participants in the comparison groups displayed differences in mean characteristics, meaning that treatment and control groups may not have been comparable in certain aspects.

1.3.3 Peers' weight-related behaviors

1.3.3.1 Diet

Twenty-four articles assessed the impact of social influence on dietary behaviors, with a number of them reporting significant peer effects on food intake. Five studies dealt with peer effects on young people's healthy eating behaviors only (Graham *et al.*, 2013; Meng *et al.*, 2017; Nix and Wengreen, 2017; Wengreen, Nix and Madden, 2017; Zhylyevskyy *et al.*, 2013), seven papers focused on unhealthy food (Cruwys *et al.*, 2012; Fortin and Yazbeck, 2015; Hirata *et al.*, 2015; Jones and Robinson, 2017; Robinson, Benwell and Higgs, 2013; Robinson, Otten and Hermans, 2016) and sugar-sweetened beverage consumption (Jones and Robinson, 2017; Melbye and Helland, 2018; Robinson, Otten and Hermans, 2016), and twelve papers examined both healthy and unhealthy dietary behaviors. Among these studies, daily fruit and vegetable servings and the frequency of high-energy-dense food intake have been

employed as preferred outcome measures. Overall, fourteen studies investigated individual behaviors in response to peer fruit and vegetable consumption, with the majority of them suggesting positive associations. Experimental research supported these findings, estimating that participants' daily fruit and vegetable intake increased by approximately 0.2 to 0.9 cups in response to perceived peer behaviors (Liu and Higgs, 2019; Nix and Wengreen, 2017; Wengreen, Nix and Madden, 2017). Yet, amidst the generally positive associations found, some authors yielded different conclusions, reporting no significant associations between peers' intake amount and frequency and an individual's own fruit and vegetable consumption (Ali, Amialchuk and Heiland, 2011; Pelletier, Graham and Laska, 2014; Robinson *et al.*, 2013; Yuan, Lv and Vanderweele, 2013; Zhylyevskyy *et al.*, 2013).

Eleven studies used young people's consumption of high-calorie food and beverages to examine unhealthy dietary behaviors in non-experimental settings (Ali, Amialchuk and Heiland, 2011; Fortin and Yazbeck, 2015; Gesualdo and Pinquart, 2021; Hawkins, Farrow and Thomas, 2020; Jones and Robinson, 2017; Lally, Bartle and Wardle, 2011; Melbye and Helland, 2018; Pelletier, Graham and Laska, 2014; Perkins, Perkins and Craig, 2018; Robinson, Otten and Hermans, 2016; Yuan, Lv and Vanderweele, 2013). In these studies, results were generally more unanimous, and overall positive associations were found. For example, Ali, Amialchuk and Heiland (2011) and Fortin and Yazbeck (2015) estimated that an adolescent's weekly frequenting of fast food restaurants increased when their friends went more often, a finding that was corroborated by Pelletier, Graham and Laska (2014). However, some evidence indicated that differences in age groups may play a role in unhealthy consumption behaviors: while daily intake of sugar-containing drinks by high school students was still strongly predicted by their friends' and grademates' consumption (Melbye and Helland, 2018; Perkins, Perkins and Craig, 2018), this effect vanished in populations of university students (Hawkins, Farrow and Thomas, 2020; Robinson, Otten and Hermans, 2016).

1.3.3.2 Physical activity

Twenty-one studies investigated the associations between peer behaviors and individual physical activity outcomes. Most papers on physical activity aimed at estimating the impact of peers' activity behaviors on the individual frequency of engaging in some type of exercise

Table 1.1: Characteristics of selected studies on dietary behaviors

Study	Country	N	Population	Method	Results	Mechanisms	Quality
Ali, Amialchuk and Heiland (2011)	USA	2760	HS	MLR; Probit, L	+ve association between friends' fast food consumption and own behavior: $\beta = 0.178^{***}$. NS effects for breakfast, FV servings, and calorie-dense snacks.	PN†; SN†	good
Cruwys *et al.* (2012)	AUS	119	UG	ANCOVA, E	+ve: participants ate more popcorn when they believed a prior participant (in-group member) had eaten all theirs vs. none; MDiff = 8.3g. Effect moderated by shared social identity.	DN (Mod); SI	good
Fortin and Yazbeck (2015)	USA	2355	HS	GSAR, L	+ve: small endogenous peer effect in friends' fast food consumption: $\beta = 0.129^{*}$. Amplified via a social multiplier of 1.15.	SN†	good
Gesualdo and Pinquart (2021)	DEU	208	UNI	MLR, CS	+ve association between peers' eating behavior and own behavior: $\beta = 0.21^{*}$. NS for partner's eating.	-	medium
Graham *et al.* (2013)	USA	1201	COL	MLR, CS	+ve association between friends' behavior and own FV consumption: $\beta = 0.084^{***}$.	-	medium
Hawkins, Farrow and Thomas (2020)	GBR	369	UNI	MLR, CS	+ve associations between amount and frequency of Facebook users' FV and participants own consumption. NS for HED and SSB intake.	DN†	good
Hirata *et al.* (2015)	BRA DEU	83 100	UG	ANCOVA, E	+ve: subjects exposed to high intake norm consumed more chocolates compared to those exposed to low intake norm. NS between-country difference in informational eating norm effects.	DN (Mod)	good
Jones and Robinson (2017)	GBR	340	UNI	MLR, L	+ve: higher perceived peer frequency of cake/pastry consumption predicted increased personal intake over time ($\beta = 0.377^{**}$). NS for SSB or descriptive peer norms.	DN	low

Table continues on next page

Table 1.1: Characteristics of selected studies on dietary behaviors (continued)

Study	Country	N	Population	Method	Results	Mechanisms	Quality
Kimura *et al.* (2021)	JPN	51	UG	t-tests (unpaired), E	+ve: co-eating with a friend increased intake of unfamiliar snacks. Participants in pair condition consumed more and were more likely to try all unfamiliar snacks.	Mod†; SC† (affiliation)	medium
König *et al.* (2017)	DEU	402	UNI	MLM, E	+ve: perceived snacking behavior of popular peers predicted own healthy ($\beta = 0.25^{***}$) and unhealthy ($\beta = 0.24^{***}$) snacking. Moderated by peer group identification.	SImg; SC (identification); DN	good
Lally, Bartle and Wardle (2011)	GBR	264	HS	MLR, CS	+ve associations between perceived DN and own FV intake: $\beta = 0.50^{**}$, SSB: $\beta = 0.44^{**}$, and unhealthy snacks: $\beta = 0.49^{**}$. NS for injunctive norms.	DN; Norm misperception†	good
Liu and Higgs (2019)	GBR	84 90	UNI	MLR, E	+ve: participants ate more vegetables ($\beta = 36.84^{*}$) or cookies ($\beta = 18.63^{**}$) when they believed others had consumed more; ate fewer cookies when low norm presented. NS moderation by group identification.	DN; Mod; SC (identification)	good
Melbye and Helland (2018)	NOR	694	HS	Mlogit, CS	+ve: perceived descriptive friend norms increased odds of SSB consumption with school lunch: OR = 1.33^{*}.	DN	low
Meng *et al.* (2017)	USA	73	COL	ITT; LMM, E	+ve: participants who self-tracked FV intake collectively consumed more FV (M = 3.37, SD = 2.01) than those who tracked alone (M = 1.37, SD = 1.44); MDiff = 2 servings. NS: demographic similarity had no effect.	SC; SL	medium
Nix and Wengreen (2017)	USA	167	UG	ANOVA, E	+ve: participants receiving a descriptive norm message indicating they were in the lowest 20th percentile increased self-reported FV intake by 0.5 cups and peer intake perception by 1 cup. Likely driven by social approval bias.	DN; SApp; Norm misperception†	good

Table continues on next page

Table 1.1: Characteristics of selected studies on dietary behaviors (continued)

Study	Country	N	Population	Method	Results	Mechanisms	Quality
Pelletier, Graham and Laska (2014)	USA	996	UNI	MLR, CS	+ve: perceived peers' fast food consumption is positively associated with own consumption, if peer is a friend: : $\beta = 0.22^{***}$ or significant other: $\beta = 0.25^{**}$. Perceived friends' FV intake is positively associated with own consumption: $\beta = 0.13^{***}$. NS association between own consumption others' FV intake or peers' SSB consumption.	DN; SProx	medium
Perkins, Perkins and Craig (2018)	USA	5841	HS	MLM, CS	+ve: perceived peer FV and SSB consumption norms were strongly associated with personal consumption among both genders ($\beta = 0.56^{***}$ for SSB, males; $\beta = 0.52^{***}$ for SSB, females). High prevalence of unhealthy norm misperceptions documented.	DN; Norm misperception; SN†	good
Robinson and Higgs (2013)	GBR	100	UNI	ANOVA, E	NS effect in peers' HED food choice on own selection. Participants chose a meal higher in energy density, when they ate with an unhealthy peer. NS difference for total energy amount of food. Participants observing others making an unhealthy food choice, chose less healthy food.	DN	good
Robinson *et al.* (2013)	GBR	129	YA	ANCOVA, E	+ve: participants consumed less high calorie snacks if they believed that students in general ate less junk food: MDiff = -11.5 (p = 0.046). NS for FV consumption and total energy intake.	DN	medium
Robinson, Benwell and Higgs (2013)	GBR	64	UG	ANOVA, E	+ve: participants who believed the norm was to eat many cookies increased their intake (MDiff = 1.46*), while those exposed to a low intake norm ate less (MDiff = –1.3*)	DN; SApp†	good
Robinson, Otten and Hermans (2016)	GBR	1056	UNI	MLM, CS	+ve: perceived peer eating norms predicted own frequency of consumption of SSB: $S\beta = 0.08^{**}$ and SP: $S\beta = 0.33^{***}$.	DN	medium

Table continues on next page

Table 1.1: Characteristics of selected studies on dietary behaviors (continued)

Study	Country	N	Population	Method	Results	Mechanisms	Quality
Wengreen, Nix and Madden (2017)	USA	251	COL	MLR, E	+ve: participants receiving normative information increased their daily FV intake by approx. 0.89 cups. Skin carotenoid levels also increased significantly, but self-reported intake did not.	DN†; SApp†	good
Yuan, Lv and Vanderweele (2013)	CHN	419	UNI	MLR, CS	+ve association of roommates' unhealthy dietary intake on own eating behaviors. NS effect for FV intake.	SL†	good
Zhylyevskyy *et al.* (2013)	USA	502	YA	Ordered probit, CS	NS effects in fruit and vegetable consumption between youths and their best (same gender) friends.	-	good

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: **Population**: COL = college students; HS = high school students; UG = undergraduate students; UNI = university students; YA = young adults. **Method** (statistical models and study design): CS = cross-sectional; E = experimental; GSAR = generalized spatial autoregressive model; ITT = intention-to-treat analysis; L = longitudinal; LMM = linear mixed-effects model; MLM = multilevel regression; Mlogit = multiple logistic regression; MLR = multiple linear regression. **Results**: FV = fruits and vegetables; HED = high-energy-dense snack; MDiff = mean difference; NS = not significant ($p > 0.05$); $S\beta$ = standardized regression coefficient; SSB = sugar-sweetened beverages; +ve = positive, statistically significant peer effect. **Mechanisms**: Conf = conformity; DN = descriptive norms; IF = information feedback; Mod = modeling; SApp = social approval; SC = social comparison; SI = social identity; SImg = social image; SL = social learning; SN = social norms (general); SProx = social proximity.

† Mechanism inferred from the study's framing or cited theories, not directly tested. Absence of † indicates mechanisms were empirically tested. Terms in parentheses (e.g., identification, affiliation) specify conditions under which the mechanism has been studied.

per week, using continuous or discrete measures. Only Barclay, Edling and Rydgren (2013) used a binary variable to measure individual regular exercise behavior. However, measures of physical activity behavior varied widely between the studies, hence limiting comparability. While the majority of studies relied on self-reports from questionnaires, three articles deployed wearable technology to gather objective data about participants' active behaviors (Li *et al.*, 2016*a*; Morrissey *et al.*, 2015; Wang, Lizardo and Hachen, 2021), and one paper (Condliffe, Isgin and Fitzgerald, 2017) used time stamp information collected through student ID cards to track their average weekly gym visits.

Collectively, studies mostly indicated significant positive associations between youth and their peers' physical activity, even across a diverse set of global contexts, including Brazil (Cheng, Mendonça and Farias Júnior, 2014; Mendonça and Farias Júnior, 2015), China (Yuan, Lv and Vanderweele, 2013), Germany (Gesualdo and Pinquart, 2021), Korea (Lee and Lee, 2020), Sweden (Barclay, Edling and Rydgren, 2013), and the USA (Ali, Amialchuk and Heiland, 2011; Condliffe, Isgin and Fitzgerald, 2017; Graham *et al.*, 2014; Li *et al.*, 2014, 2016*b*; Long, Barrett and Lockhart, 2017; Morrissey *et al.*, 2015; Shoham *et al.*, 2012; Simpkins *et al.*, 2013; Wang, Lizardo and Hachen, 2021; Yakusheva, Kapinos and Weiss, 2011). In terms of sedentary behaviors, some studies reported that individual inactive behaviors such as screen time and sitting were positively correlated with an increase in peer screen time (Aalsma *et al.*, 2012; Lopes, Gabbard and Rodrigues, 2013; Shoham *et al.*, 2012).

1.3.3.3 Sleep

Our search produced five studies dealing with peer effects in sleep behaviors (Aalsma *et al.*, 2012; Ali, Amialchuk and Heiland, 2011; Li *et al.*, 2019; Liu, Patacchini and Rainone, 2017; Wang *et al.*, 2021). All studies were conducted in the United States, and most of them utilized self-reports from the National Longitudinal Study of Adolescent to Adult Health (Add Health)[2] from the 1990s to assess peer-related data. Two of these studies measured sleep by asking participants how many hours of sleep they obtained per night (Aalsma *et al.*, 2012; Li *et al.*, 2019), and Ali, Amialchuk and Heiland (2011) coded sleep as a binary variable

[2]Add Health is a longitudinal study following a US sample from early adolescence into adulthood. With data collection having started in 1994–95, it provides comprehensive data on individual characteristics over the life course, including social, behavioral, and environmental factors. See Harris (2013) for details.

Table 1.2: Characteristics of selected studies on physical activity

Study	Country	N	Population	Method	Results	Mechanisms	Quality
Aalsma *et al.* (2012)	USA	160	ADO	MLM, L	NS effect of romantic partners' health-protective behaviors (PA). +ve association between partner's behavior and own sedentary behavior (watching TV): $\beta = 3.02^{**}$.	-	low
Ali, Amialchuk and Heiland (2011)	USA	2760	HS	MLR, L	+ve association between friends' and own weight-related behaviors: exercising (> 3 times/week), $\beta = 0.079^{**}$; participating in active sports, $\beta = 0.184^{***}$. NS effect for sedentary behavior (watching TV).	social norms†	good
Barclay, Edling and Rydgren (2013)	SWE	5695	YA	Logit, CS	+ve: peer regular exercise is associated with an increased predicted probability of own regular exercise for both males and females. Interaction effects with relationship strength and gender homogeneity further elevate the likelihood of engaging in health behaviors.	-	medium
Barnett *et al.* (2014)	USA	129	UG	NAM, CS	NS effects in MVPA levels among residence hall members.	-	good
Cheng, Mendonça and Farias Júnior (2014)	BRA	2361	HS	SEM, CS	+ve associations between (participant-reported) friend PA and (self-reported) MVPA: $S\beta = 0.11^{***}$ (males), $S\beta = 0.07^{*}$ (females); effect mediated by SSupp.	SSupp†	good
Condliffe, Isgin and Fitzgerald (2017)	USA	181	UG	OLS, E	+ve: teams attended the gym more frequently than other groups during the experiment: $\beta = 0.795^{**}$ (teams without feedback) and $\beta = 0.719^{*}$ (with feedback). NS effect for teams post-experiment. Individual + feedback group showed increased visits in weeks 2–3: $\beta = 1.031^{*}$; remained significant after experiment: $\beta = 0.269^{*}$.	IF; SC†	good
Gesualdo and Pinquart (2021)	DEU	208	UNI	MLR, CS	+ve correlation between close social ties' PA and own current PA: $S\beta = 0.34^{*}$ for partners, $S\beta = 0.31^{**}$ for peers.	-	medium
Graham *et al.* (2014)	USA	356	HS	MLR, CS	+ve association between friends' active behaviors and own baseline PA: $\beta = 0.69^{***}$ and MVPA: $\beta = 0.56^{***}$.	SSupp†	medium

Table continues on next page

Table 1.2: Characteristics of selected studies on physical activity (continued)

Study	Country	N	Population	Method	Results	Mechanisms	Quality
Graham *et al.* (2014)	USA	356	HS	MLR, L	NS effects in PA between adolescents and their friends over time.	SSupp†	medium
Lee and Lee (2020)	KOR	740	HS, UNI	MLR, CS	+ve correlation of friends' exercise participation level with own behavior: $\beta = 0.395^{***}$.	SSupp†	low
Li *et al.* (2014)	USA	2439	HS	SEM, CS	+ve association between closest friend's MVPA and adolescent MVPA: $S\beta = 0.21^{***}$. The association was situated within a broader model that included internal motivation and MVPA planning as mediators.	SSupp†	low
Li *et al.* (2016*a*)	USA	561	HS	MLM, L	NS association between closest friends' MVPA and adolescents' own MVPA.	-	good
Li *et al.* (2016*b*)	USA	2659	HS	SEM, L	+ve association between closest friends' active behaviors and own activity: adjusted OR = 1.11^{**} for MVPA and adjusted OR = 1.17^{***} for VPA.	-	medium
Long, Barrett and Lockhart (2017)	USA	1796	HS	SAOM, L	+ve: adolescents adjusted their physical activity behavior to their friends' behavior: $\beta = 0.903^{***}$ in school A and $\beta = 0.729^{**}$ in school B.	-	good
Lopes, Gabbard and Rodrigues (2013)	PRT	268	HS	MLM, CS	+ve moderate correlation between best friend's weight-related behaviors and own activity (VPA, MPA, and sedentary behavior). NS effect for walking.	SSupp†	low
Mendonça and Farias Júnior (2015)	BRA	2859	HS	Logit, CS	+ve: friend co-participation in PA ($\geq$ 300 minutes/week) increased odds of activity: OR = 2.51^{***} for males and OR = 2.48^{***} for older adolescents (17–19 years). NS for females and 14–16-year-olds.	SSupp†	good
Morrissey *et al.* (2015)	USA	401	ADO	MLR, L	+ve association of friends' co-participation on MVPA: $\beta = 0.309^{***}$ and on weekday afternoon MVPA: $\beta = 0.377^{***}$.	SSupp†	good

Table continues on next page

Table 1.2: Characteristics of selected studies on physical activity (continued)

Study	Country	N	Population	Method	Results	Mechanisms	Quality
Shoham *et al.* (2012)	USA	1775	HS	SAOM, L	+ve: egos (high on screen time) are likely to remain high or increase their screen time if their friends are also a high screen time type. Egos (playing an active sport weekly) had a 75% predicted probability of decreasing their playing sports if alter did not play any sports.	-	good
Simpkins *et al.* (2013)	USA	1896	HS	SAOM, L	+ve association between friend PA and own behavior: $\beta = 0.45^{*}$.	DN†	good
Wang, Lizardo and Hachen (2021)	USA	619	UG	MLM, L	+ve associations between peers' active behaviors and own daily activity: $\beta = 0.06^{***}$, steps: $\beta = 0.07^{***}$, active minutes: $\beta = 0.07^{***}$, and activity calories: $\beta = 0.07^{***}$.	DN	good
Yakusheva, Kapinos and Weiss (2011)	USA	144	UNI	OLS, E	+ve and marginally significant association between roommate's pre-college exercise behavior and own outside exercise per week: $\beta = 0.13$ ($p < 0.1$). NS effect for gym usage.	-	good
Yuan, Lv and Vanderweele (2013)	CHN	419	UNI	MLR, CS	+ve peer effect of roommates on bicycle use ($\beta = 4.00^{**}$) and marginal effect on moderate-intensity exercise ($\beta = 0.22$, $p = 0.078$). NS for vigorous-intensity activity.	SL†	good

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: Graham *et al.* (2014) appears twice as they report both cross-sectional and longitudinal analyses of physical activity associations among adolescent girls and their friends. **Population**: ADO = adolescents; HS = high school students; UG = undergraduate students; UNI = university students; YA = young adults. **Method** (statistical models and study design): CS = cross-sectional; E = experimental; L = longitudinal; MLM = multilevel regression; MLR = multiple linear regression; NAM = network autocorrelation model; SAOM = stochastic actor-oriented model; SEM = structural equation modeling. **Results**: MDiff = mean difference; MVPA = moderate to vigorous physical activity; NS = not significant ($p > 0.05$); OR = odds ratio; PA = physical activity; $S\beta$ = standardized regression coefficient; VPA = vigorous physical activity; +ve = positive, statistically significant peer effect. **Mechanisms**: DN = descriptive norms; IF = information feedback; Mod = modeling; SC = social comparison; SL = social learning; SN = social norms (general); SSupp = social support (e.g., co-participation).

† Mechanism inferred from the study's framing or cited theories, not directly tested. Absence of † indicates mechanisms were empirically tested. Terms in parentheses (e.g., identification, affiliation) specify conditions under which the mechanism has been studied.

indicating whether the adolescent usually slept six hours at most. None of these papers found any significant influence of romantic partners' or friends' behaviors on an individual's predisposition to sleep duration.

Liu, Patacchini and Rainone (2017) explored an alternative outcome measure to analyze peer effects in sleep. Rather than sleep duration, they studied the influence peers exert on adolescent bedtime decisions, also using Add Health data. They estimated a nonlinear least-squares model to account for measurement issues originating from missing observations in the original dataset. While controlling for demographic characteristics (age, gender, and race) and family background (e.g., parents' education and occupation), their most extensive model indicated a large, positive, and significant relationship. With an endogenous peer effect of 0.602, adolescents were found to go to bed approximately 36 minutes later if their friends delayed their own bedtime by one hour.

Wang *et al.* (2021) was the only study using objective sleep and network data obtained via wearables. After using an extensive set of time-variant variables (e.g., weather indicators) and time-invariant variables (e.g., gender, race, and ethnicity or personality traits) as controls, the authors reported that peer behaviors influence students' daily time in bed, sleep duration, as well as daily bed- and rising time: a one-unit increase in a peer's sleep duration was associated with a 0.11 unit increase in one's own duration of sleep, which translates to 6.5 additional minutes. Wang *et al.* (2021) found no association with sleep quality measures (e.g., hourly awakening frequency or the number of sleep episodes). Other sleep quality measures (insomnia symptoms and sleep insufficiency) were investigated by Li *et al.* (2019). While the authors did not report any significant association between friends' insomnia symptoms and individuals' difficulties falling asleep, they found an elevated risk of sleep insufficiency of 41% for those whose friends reported not getting enough rest per night.

1.3.4 Experimental and non-experimental evidence

1.3.4.1 Field and laboratory experiments

Many of the identified studies were motivated by the increasing prevalence of health-harming behaviors and the subsequent spread of modifiable risk factors and outcomes such as obesity

Table 1.3: Characteristics of selected studies on sleep behaviors

Study	Country	N	Population	Method	Results	Mechanisms	Quality
Aalsma *et al.* (2012)	USA	160	ADO	MLM, L	NS effects in self-reported hours of sleep between adolescents and their romantic partners.	-	low
Ali, Amialchuk and Heiland (2011)	USA	2760	HS	MLR; Probit, L	NS association between friends' sleep duration and adolescents' own sleep behavior.	SN†	good
Li *et al.* (2019)	USA	2550	HS	MLR; Poisson, CS	NS effects in friends' sleep duration and own sleep duration. +ve: friends' sleep insufficiency positively affects own sleep insufficiency: $\beta = 1.41^{**}$.	-	good
Liu, Patacchini and Rainone (2017)	USA	8000	HS	NLS, CS	+ve: friends going to bed one hour later on average delays own bedtime by 36 minutes: $\beta = 0.602^{**}$.	DN†; Conf†	good
Wang *et al.* (2021)	USA	619	UG	MLM, L	+ve associations between peers' time in bed and own total sleep time: $\beta = 0.11^{***}$; bedtime: $\beta = 0.08^{***}$; rising time: $\beta = 0.09^{***}$. NS effects for sleep efficiency, number of sleep episodes, sleep onset latency, and frequency of awakenings.	-	good

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: **Population**: ADO = adolescents; HS = high school students; UG = undergraduate students. **Method** (statistical models and study design): CS = cross-sectional; L = longitudinal; MLM = multilevel regression; MLR = multiple linear regression; NLS = nonlinear least squares regression. **Results**: NS = not significant ($p > 0.05$); +ve = positive, statistically significant peer effect. **Mechanisms**: Conf = conformity; DN = descriptive norms; SN = social norms (general).

† Mechanism inferred from the study's framing or cited theories, not directly tested.

across populations (e.g., Fortin and Yazbeck, 2015; Yakusheva, Kapinos and Weiss, 2011; Yuan, Lv and Vanderweele, 2013). One reason for the observation of *socially contagious* health-related outcomes could be that individuals select their peers based on certain characteristics, which poses a challenge in distinguishing peer effects from non-random selection effects (Bramoullé, Djebbari and Fortin, 2020). These correlated effects are most credibly addressed by applying (quasi-)random assignments of study subjects to peer groups, a method commonly used in experimental designs.

Overall, our search identified 13 experimental studies, with the majority focusing on peer effects in dietary behaviors, and only two studies focusing on exercise (Condliffe, Isgin and Fitzgerald, 2017; Yakusheva, Kapinos and Weiss, 2011). Yakusheva, Kapinos and Weiss (2011) convincingly addressed the identification problem by exploiting randomized roommate assignments to study weight change and related behaviors of female university students living on campus. By regressing individual behaviors during the freshman year on own and roommate's behavior prior to entering college, they were able to address structural simultaneity between individual outcomes within peer groups. To control for unobserved heterogeneity, dormitory fixed effects were included. They found evidence of small but positive peer effects in weekly exercise and a significant increase in the use of weight-loss supplements. Alternatively, Condliffe, Isgin and Fitzgerald (2017) conducted a field experiment to identify peer effects in gym attendance among a sample of predominantly female undergraduate students. Students were randomly assigned to one of five experimental groups to explore the impact of information and group incentives, both separately and in combination. The authors ran OLS regressions of participants' average weekly gym visits, conditional on their assigned treatment condition and demographic control variables. In line with previous peer effects studies on exercise behaviors, the paper revealed significant associations: being assigned to a partner increased early within-experiment gym visits. Teams that received no information on peers' gym attendance increased their participation by 0.795 weekly visits, and teams with information feedback by 0.719 weekly visits. By contrast, the authors found that students without a partner but with information on their peers' outcomes showed no statistically significant effect on gym visits in the first week. However, they exhibited an increase of 1.031 weekly visits in the third week. Further, this positive association remained significant in the

weeks after the experiment.

Some experiments relied on self-reported fruit and vegetable (Meng *et al.*, 2017; Nix and Wengreen, 2017) or unhealthy snack consumption (König *et al.*, 2017). Subjectively reported peer behavior has been argued to overcome some of the challenges arising from the common use of average peer behavior to measure peer effects (Kawaguchi, 2004). On the other hand, subjective reports may be prone to errors and biases, such as recall errors or social desirability bias (Bertrand and Mullainathan, 2001), and these issues may become particularly important in small sample sizes. To address these concerns, Wengreen, Nix and Madden (2017) used a skin biomarker for fruit and vegetable intake to estimate the effect of normative information on food intake. They informed their study participants that 80% of their peers in college had higher skin carotenoid scores than they did. Over the course of eight weeks, this led to a rise in skin carotenoid scores, indicating an increase of 0.89 cups in daily fruit and vegetable consumption.

Yet, it is likely that results on peer effects derived from experimentally manipulated peer groups, such as random roommate assignments, may be fundamentally different from those effects that come from close relationships among people. For example, a recent experimental study (Kimura *et al.*, 2021) estimated the influence of co-eating with friends on unfamiliar snack intake in Japanese undergraduate students. The authors reported that friend pairs ate more and for longer during the snack-tasting sessions compared to those eating alone, while pointing out that conformity norms may work differently among pairs composed of strangers. Contrastingly, Robinson and Higgs (2013) examined food choice in the presence of an unfamiliar person. Study participants who were assigned to the unhealthy treatment condition (i.e., they observed another person choosing high-energy-dense food items for lunch) were less likely to select low-energy-dense foods (e.g., carrot sticks). Yet, Robinson and Higgs did not find significant differences in kilocalories consumed, which they considered indicative of less pronounced peer effects in food choice settings. These results might be taken with care, though; the sample sizes in both studies were among the smallest in the set of articles identified, with a sample size of 51 (Kimura *et al.*, 2021) and 100 (Robinson and Higgs, 2013), respectively. We are unable to derive solid conclusions about any fundamental difference between peer effects from close and distal peers based only on this evidence.

1.3.4.2 Observational research

We identified 20 studies with cross-sectional and 13 studies with longitudinal study designs. Notably, the majority of longitudinal studies were conducted in the United States, with nearly half of them utilizing Add Health data. Only one study was conducted in the United Kingdom (Jones and Robinson, 2017). Both cross-sectional and longitudinal studies often relied on the average of peers' reported frequency of engaging in weight-related behaviors as the main predictor (e.g., Ali, Amialchuk and Heiland, 2011; Fortin and Yazbeck, 2015; Shoham *et al.*, 2012; Wang, Lizardo and Hachen, 2021; Wang *et al.*, 2021). In contrast to the experimental literature that mainly focused on eating behaviors, the observational studies investigated a wide range of outcomes, including fruit and vegetable intake (Ali, Amialchuk and Heiland, 2011), fast food consumption (Ali, Amialchuk and Heiland, 2011; Fortin and Yazbeck, 2015), soda, sweets, and pastry consumption (Jones and Robinson, 2017), active behaviors per day (e.g., Graham *et al.*, 2014; Wang, Lizardo and Hachen, 2021) and per week (e.g., Aalsma *et al.*, 2012; Li *et al.*, 2016*a*,*b*; Long, Barrett and Lockhart, 2017; Shoham *et al.*, 2012; Simpkins *et al.*, 2013), as well as a variety of sleep behaviors (Ali, Amialchuk and Heiland, 2011; Wang *et al.*, 2021). While a considerable number of studies employed linear regression models to analyze their data, some papers deviated from this conventional approach in the literature, opting for alternative regression methods and analytical frameworks. Longitudinal studies, for example, utilized multilevel modeling to analyze nested data structures in the survey responses to quantify peer effects over time (Aalsma *et al.*, 2012; Li *et al.*, 2016*a*; Wang, Lizardo and Hachen, 2021; Wang *et al.*, 2021), as well as stochastic actor-based models to exploit social network data from the Add Health data set (Long, Barrett and Lockhart, 2017; Shoham *et al.*, 2012; Simpkins *et al.*, 2013).

The general notion of positive peer effects in youth behaviors was supported by 74% of the cross-sectional studies and the majority of longitudinal studies, particularly those studying peer effects on physical activity. There is limited observational evidence on the impact of peers on sleep behaviors, and data on how peers affect sleep patterns remains scarce.

1.3.5 Meta-regression analysis: exploring moderators and sources of variability

The presence of heterogeneity in effect sizes between reviewed studies, as evident from the previous sections, warrants further investigation. Meta-regression is one method to examine the role of contextual factors at the study level that potentially contribute to the observed dispersion (Borenstein *et al.*, 2009). In spite of the variability in measures used to quantify weight-related outcomes, 19 studies were analyzed to examine the impact of a small set of contextual characteristics (such as sample size, population, and quality rating) to explain the variability in regression coefficients. The results are reported in Table 1.4. The fitted meta-regression models included categorical and continuous variables and were run independently for two subgroups: dietary behaviors (1) and physical (in)activity (2). While positive associations between sample size and reported effect size were found in dietary studies, along with a negative impact from studies conducted in the United States, these effects appear to be primarily driven by one specific study (refer to Table A1.6 for more details). In contrast, studies on physical activity with larger sample sizes tend to report smaller effect sizes, and an inverse relationship is observed between quality rating and effect size in these studies. This suggests that higher-quality studies are more likely to report smaller effects – a potential indicator of publication bias.

As a further robustness check, we examined the impact of additional study characteristics on the likelihood of positive and significant peer effects being reported, employing a dummy dependent variable in our analysis. The analysis was conducted across three model specifications (3)–(5), for which the outcomes of all studies were categorized according to whether they reported a positive and significant association between peer behaviors and individual health outcomes. Of the 111 observations, 74% reported positive associations. In all model specifications, the likelihood of observing significant and positive peer effects increases with larger sample sizes. One possible explanation is that larger sample sizes generally yield smaller standard errors, and thereby more precise estimates, while smaller studies may show greater variability in effect sizes, including more extreme values. Yet, the magnitude of these effects remains relatively small. Additionally, studies employing university student samples were more likely to report

positive associations. In our most comprehensive model (5), the dummy variable for studies on activity-related behaviors is statistically significant and increases the likelihood of studies reporting positive associations compared to studies on diet. Conversely, we do not find a significant difference in the likelihood of reporting positive results between studies employing experimental designs and those using non-experimental methodologies.

Table 1.4: Meta-regression analysis: effect of sample and study characteristics on the magnitude of effect size (1)–(2) and the likelihood of studies reporting positive and significant peer effects (3)–(5)

	Dependent variable				
	Diet	PA	Peer effect significant (dummy)		
	(1)	(2)	(3)	(4)	(5)
Sample characteristics					
Sample size n	0.000*	−0.000**	0.000*	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Country: US	−0.177*	−0.005	−0.183+	−0.154	−0.137
	(0.081)	(0.053)	(0.094)	(0.096)	(0.096)
University sample	0.063	−0.131+	0.147	0.245*	0.234*
	(0.111)	(0.060)	(0.108)	(0.110)	(0.109)
Quality rating	−0.001	−0.006**	0.000	0.002	−0.000
	(0.004)	(0.001)	(0.003)	(0.003)	(0.004)
Study characteristics					
Health behavior: Physical activity				0.181+	0.204*
				(0.095)	(0.095)
Health behavior: Sleep				−0.304*	−0.237
				(0.142)	(0.148)
Study design: Experiment					0.178
					(0.116)
Observations	21	15	111	111	111

Standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. + indicates marginal significance at the 10% level. PA = Physical activity.

1.3.6 The nature of peer effects

The study of social interactions and peer effects has become a subject of interest to scholars in economics, health, psychology, and other social sciences for many years. As of today, there exists a considerable body of literature linking peer influence to health-related behaviors.

The prevalence of young people's unhealthy lifestyles and the extent to which peer influence may be involved in shaping others' behaviors has attracted great interest among researchers, shifting the focus from merely identifying the occurrence of peer effects to examining the different processes by which they operate. Thus, one focus of this review has been on the underlying mechanisms of peer effects. We reviewed studies that adduced several economic and psychological models to explain or elicit peer effects, including conformity, social norms and normative information feedback, social comparison, image-related concerns, as well as social identity. In spite of the advances made in the peer effects literature, the motives behind an individual's behavioral adaptation and convergence to others' behaviors still resemble a black box, as do the conditions under which we observe these effects.

The dominant mechanism discussed in most of the identified studies to explain peer effects is social norms. Social norms can help achieve collectively desirable outcomes when formal institutions fail (Nyborg *et al.*, 2016). They are implicit rules defining societies, affecting people's everyday decisions and economic behavior in various ways. Several studies show how social norms impact individual and wider socioeconomic outcomes. Nyborg and Rege (2003) found that norm changes in public smoking behavior following law amendments in Norway explained the formation of considerate smoking norms even in non-targeted areas like private homes. Young (2015) reviewed studies on norm dynamics in health- and non-health-related settings and concluded that under certain conditions, policy can effectively create or modify norms and thus individual behavior. While the formation of norms driving peer effects may happen gradually, a variety of other motives can be involved, including a person's desire to conform and avoid social disapproval.

Other competing explanations may be that people simply learn from the choices others make or that conformity results from people's image concerns and aspirations toward social status (Andreoni and Bernheim, 2009; Bernheim, 1994). A related, yet distinct, concept includes competitive preferences and the social comparison motive. If an individual cares about social comparison, conformity can arise because their utility depends on how they compare to their peers (Zafar, 2011). Social comparison may give rise to behavioral adaptation if individuals use their peers' choices as a reference point for their own behavior. For example, Condliffe, Isgin and Fitzgerald (2017) provided informative feedback on peers' exercising

behaviors. They showed that gym attendance of those students who learned about their peers' performance (i.e., the number of peers who had met a weekly goal) increased compared to participants without such information. Condliffe, Isgin and Fitzgerald attributed this to a competitive effect from entering a lottery, where participants used their peers' performance as a reference point. Although the information was anonymous and increases were observed for both individuals with information and teams with information compared to their equivalents without information, social image or status concerns may also have been at play. In a web-based experiment, Meng *et al.* (2017) examined social comparison processes in healthy food consumption. Participants were assigned to groups sharing a collective goal of fruit and vegetable intake, where they received experimentally manipulated information about their group members' identities and consumption behavior. While participants who were assigned to a group aiming at a collective goal reported an average of 3.37 daily servings of fruit and vegetable intake, participants not assigned to a condition where they could compare themselves to other group members' outcomes recorded only 1.37 servings.

At least ten experimental studies investigated the impact of social norms on a person's dietary behavior. All papers focused on the effect of norm-based interventions on subjects' food intake, either by analyzing fruit and vegetable consumption (Liu and Higgs, 2019; Meng *et al.*, 2017; Wengreen, Nix and Madden, 2017) or high-energy-dense and sweet food intake (Hirata *et al.*, 2015; Robinson, Benwell and Higgs, 2013). However, none of these studies has been able to credibly differentiate between the possible processes at work that may all lead to the same empirical outcome, namely, conformity to peer behavior. While these studies exposed their participants to information about their peers' behavior, it is possible that this information helped participants learn about the correct behavior (e.g., increasing fruit and vegetable intake or reducing high-calorie snack consumption). It is also possible that the normative information about others' outcomes was instead used as a reference point for their own decisions. As a result, behavioral adaptation is consistent with a variety of explanations, from people's willingness to conform to social norms to comparison and learning mechanisms.

The reviewed evidence suggests that behavioral spillovers via social influence are more likely to occur when a norm originates from an in-group than when it comes from an out-group with which an individual does not identify, indicating a social identity motive underlying these peer

effects. For example, in a laboratory experiment with female university students, Cruwys *et al.* (2012) reported a significant interaction between social identity and food intake: participants ate more when an individual who allegedly belonged to the same university – and thus the same social group (in-group) – set a high food intake norm, whereas participants reduced their consumption significantly when they encountered an in-group member in the low norm condition. When participants encountered confederates they believed were not affiliated with the same university, no such effects were found. König *et al.* (2017) exploited a within-subjects design to experimentally examine peer effects in fruit and sugary snack consumption. Using a predominantly female student sample, they found that a one-unit increase in healthy snacking frequency ascribed to popular peers increased own weekly healthy snacking by 0.253. The authors reported that this effect appeared to be stronger for those participants who identified with the student population of their university. For unhealthy snacking, such as candy bars, cookies, and wine gums, König *et al.* (2017) found no moderating effect of identification with a reference group. Drawing on these findings, Liu and Higgs (2019) ran two experiments using a remote-confederate paradigm with female students to test whether subjects adjusted their eating behaviors in response to normative information about healthy and unhealthy snack consumption of their peers. Controlling for habitual snack food intake in the first of their studies, they estimated a positive peer effect of alleged same-university students on subjects' cookie intake in the high norm condition compared to the control group. In addition, they found a negative effect when the perceived cookie intake of peers was low compared to those who did not receive any normative information. However, in line with König *et al.* (2017), they could not confirm their hypothesis that the strength of identification served as a significant moderator of peer effects.

Overall, our results suggest that peer effects in weight-related behaviors are best explained by social norms. Although the reviewed studies do find evidence for the empirical importance of other mechanisms, only a few studies attempted to examine how social comparison, social learning, and image-related concerns, among others, explain behavioral adaptation to peer behavior. While norm-based interventions provide reasonable means to directly study the effect of peer norms on individual behavior, their study designs generally do not take into account other motives that are also consistent with the observed patterns of conformity. An

overview of the mechanisms discussed is provided in Appendix A1.4 (Table A1.5).

1.4 Discussion

1.4.1 Principal findings

Through a systematic literature review, we identified 45 studies from the past decade that focused on the importance of close and distal others in shaping one's own health-related behaviors during youth. Herein, we turned our attention to evaluating the evidence not only on the presence or absence of peer effects, but also on their underlying mechanisms.

First, our review identified considerable methodological heterogeneity across studies. For a start, the utilization of natural experiments and randomization (as for example in Yakusheva, Kapinos and Weiss (2011) and Condliffe, Isgin and Fitzgerald (2017), respectively) has been a reliable tool in increasing our understanding of the causal influence of peer effects. While experimental study designs provide important means to investigate peer influence, these approaches also come with their challenges specific to the study of social interactions: randomization cannot generally overcome the reflection problem, as has been pointed out earlier by Hsieh and Van Kippersluis (2018). Studies using observational data have adopted other strategies. Typically, the impact of peer effects has been estimated via regression methods, incorporating the average behavior of peers as an independent variable, while adjusting for both the individual's background and the mean background characteristics of the peers. However, the relation between peer and individual behavior is often assumed to be linear-in-means. Manski (1993) noted that these models also suffer from a reflection problem, which leads to considerable challenges in the identification of the endogenous peer effect. More precisely, researchers are unable to identify whether the observed effects actually stem from the group itself or whether it is the individual's behavior as part of the group that matters. Over the years, researchers have relied on a set of different methods to estimate social interactions and draw conclusions about the particular role peers play in individual health-related decision-making processes. Balsa and Díaz (2019) briefly summarized different approaches that have been utilized to address the reflection problem in the literature on social

interaction models, and Pratschke and Abbiati (2023) provide a concise summary of the main empirical research methods commonly applied in the study of peer effects using observational data, emphasizing the advances made through multilevel models, social network analysis, and spatial autoregressive modeling, while addressing the common identification problems.

Second, our review pointed out a range of potential mechanisms driving peer effects in health behaviors. In our main analysis, we observed mostly positive associations between other people's behaviors and one's own health-related outcomes. Based on our synthesis, we find that drivers of peer effects are often modeled and expressed in the form of one particular mechanism, neglecting the role of complementarities, reciprocity, risk sharing, and other motives. The major problem with this lack of knowledge is that we would derive different policy implications depending on which mechanisms are more pronounced in a specific context (Nakamura, Suhrcke and Zizzo, 2017). We elaborate more on this in Section 1.4.2.

Third, we identified a number of limitations and gaps in the existing evidence base that should be noted. As illustrated in previous sections, the included studies exhibited vast heterogeneity in the measures used to quantify outcomes, which poses challenges in terms of drawing consistent conclusions about peer effects on weight-related behaviors. Some studies distinguished between exercise and sports and their respective intensities, while others did not specify the type of physical activity; there were also differences in frequency measures, for example, fruit and vegetable consumption during the past seven or 30 days, or physical activity in minutes per day or times per week. An additional factor is the use of mostly self-reported measures to assess the outcome of interest, which may have compromised the validity and reliability of these metrics.

We identified a lack of research on peer effects and weight-related behaviors in non-educational contexts. More than half (56%) of the empirical evidence originated from the United States and, perhaps unsurprisingly, most studies were set in high schools and universities of industrialized countries. Thus, the analysis is especially important for high-income countries but less generalizable for other regions and environments. Evidence on the association of peer behaviors and diet, exercise, or sleep in other settings such as clubs or workplaces is also underrepresented. Consequently, other types of relationships besides friendships or fellow students have received little attention, including siblings, romantic partnerships, and club

mates. Additionally, only a few studies examined peer influence in online settings and networks such as Facebook and Instagram, despite the already significant and growing role of online peer communities. Our review also identified another limitation in the reviewed evidence base (possibly the most important one) which is the time of data collection. Although our systematic literature search explicitly addressed studies from 2011 to 2022 for the most recent evidence, many of the identified studies employed data that was collected only before 2010. A mere eight studies from the final selection used data collected between 2017 and 2020, while 42% of the included papers employed data from 1994 to 2010. In view of the rise of social networking sites during the past decade in the 2010s (Ortiz-Ospina, 2019), everyday communication and peer interaction also expanded considerably to social media platforms such as YouTube and Instagram[3]. However, although previous research found that social network site usage was significantly and negatively associated with adolescents' health behaviors, including healthy eating and sleep duration (Serenko, Turel and Bohonis, 2021), the channels through which adolescent behavior is affected remain unknown, and research in this area continues to be scarce.

1.4.2 Mechanisms and policy implications

For the purpose of curbing unhealthy weight-related behaviors in young people, policymakers could leverage the insights gained from current evidence on peer influence to design effective policies. Based on a brief summary of how different motivations related to social interactions could be utilized in health policy settings (Nakamura, Suhrcke and Zizzo, 2017), this section revolves around the identified mechanisms in Section 1.3.6 and their respective policy implications.

It is well established that other people's actions convey informational cues, especially in situations of uncertainty. Individuals are able to acquire knowledge about others' behaviors and the outcomes they experience through observation (Bikhchandani, Hirshleifer and Welch, 1998). This in turn may be used to form preferences, update beliefs, and inform subsequent decisions and actions. Along these lines, Cutler and Glaeser (2010) emphasized that even

[3]For instance, a great majority of US-American youth reported that they had ever used YouTube (90% of 18- to 24-year-olds) or Instagram (75% of 18- to 24-year-olds) in 2019 (Ortiz-Ospina, 2019).

social spillovers in harmful behaviors such as tobacco or alcohol consumption may result from learning about the putative benefits of these activities. From another point of view, people also get to know about prevalent norms through observational or social learning. This indicates that there are significant health-related policy implications of this motive. Educational policies could be designed to improve youth health literacy and provide information about peer behaviors and their consequences. Programs could also help to correct misperceptions through the dissemination of knowledge about the real prevalence of certain activities. For example, Amialchuk, Ajilore and Egan (2019) reported that misperceptions about social norms exhibit a strong effect on young students' alcohol use and related behaviors. Utilizing social norms to indicate what most people do as well as the disclosure of health consequences of specific actions could be used to harness social learning as an underlying mechanism of peer effects.

In contrast, the social comparison motive adopts the idea that individuals emulate their peers' behaviors because their utility depends on how they compare to others. Peers and their behaviors serve as a reference group: an individual's utility increases as they align with other people's behavioral outcomes; by contrast, deviation from the norm may be associated with utility decreases. Arduini, Iorio and Patacchini (2019) found evidence that female students exposed to an environment with thinner peers are more likely to adopt harmful weight-related behaviors, including induced vomiting, laxative or diet pill abuse, and excessive exercise behaviors. They attributed this outcome to social comparisons with peers that had an appearance perceived as more desirable relative to themselves. Another approach was taken by Mathieu-Bolh (2020), who built a theoretical model of intertemporal food consumption, where individuals are motivated to adapt to an endogenous social weight norm because of a desire to conform. Mathieu-Bolh posited that the perception of a healthy weight depends on the degree to which individuals compare themselves to their peers. According to these results, interventions other than those targeting norm changes could show more promise if decision-makers intend to leverage the power of peer effects. Thus, if social comparison underlies conformity as a primary driver in health-related decision-making, Nakamura, Suhrcke and Zizzo (2017) suggested considering the creation of incentives to opt for healthier options or, conversely, disincentivizing unhealthy behaviors through the enforcement of punishment.

Arduini, Iorio and Patacchini (2019) went one step further and emphasized the benefits of educational programs supporting young people in capacity building to manage negative effects of comparison processes and to develop a positive self-image.

A different, but closely related, motivation underlies peer effects if individuals care not only about conforming to a reference group, but also about complying with norms to gain esteem and social prestige. Individuals who are motivated by social status or image-related concerns conform because their status depends on adhering to observed social norms, and deviation is viewed as a loss of prestige (Akerlof, 1980; Bernheim, 1994). While social comparison can play a role even in settings where identities are anonymous, previous studies implied that behavior must be observable by others in order for a status-related mechanism to become effective. This had previously been demonstrated for productivity in the workplace (Mas and Moretti, 2009) and charitable giving (Zafar, 2011), among other contexts. Policymakers could leverage this mechanism by targeting current beliefs and expectations about what others expect one to do. Exploiting media campaigns may be an obvious choice to take full advantage of their generally wide coverage and accessibility as well as their ability to shape norms, for example through the regulation of advertising and marketing practices. Additionally, social media may prove to be an even more influential lever as they provide a more tangible environment and more reciprocal interactions compared to traditional media (Fardouly, Pinkus and Vartanian, 2017).

Moreover, there is evidence that social identity may also play an important role in generating peer effects in health behaviors. For example, Gioia (2017) noted that there is an association between the extent of identification with a certain social group and the magnitude of peer effects. As the sense of belonging to a certain group increases, the more influential peers from the same group become. This would imply that policies targeting only a few, but the most relevant, peers could generate more effective outcomes. However, despite the solid progress scholars have made toward a better understanding of the nature of peer effects, it still proves challenging to choose the most relevant peers and to understand how people interact in certain domains.

Finally, we found that most studies generally attributed behavioral changes to the influence of social norms, especially those investigating dietary behaviors. While social norms are clearly

a potential candidate to explain such behavior, none of the selected studies distinguished between normative and empirical expectations (Bicchieri and Xiao, 2009), or ruled out alternative mechanisms like imitation or status concerns. Following the seminal work by Krupka and Weber (2013), previous research in behavioral economics has successfully managed to disentangle the role of social norms from social preferences in prosocial behavior (Gächter, Nosenzo and Sefton, 2013; Kimbrough and Vostroknutov, 2016). Yet, the challenge would be to apply this knowledge to better understand the different mechanisms behind weight-related behavior in the field.

1.4.3 Limitations of the review

Our systematic review synthesized the association between peer effects and weight-related behaviors, as well as the underlying mechanisms leading to behavioral adaptation among young people. While this review may be able to provide some first tentative guidance for decision-makers to leverage the insights gained from current evidence on peer influence, the number of studies actually investigating the driving factors was limited. Another limitation of this review is that we limited our search of electronic databases to peer-reviewed articles published in English, excluding gray literature, unpublished studies, and non-English publications. It is possible that some relevant works have been excluded based on the applied criteria, which may directly affect the validity of our study and give rise to publication bias. Asymmetrical funnel plots, Egger's regression test for funnel plot asymmetry, and a simple selection model indicate potential publication bias, specifically regarding studies reporting null or negative effects with high standard errors. Further, the methodological heterogeneity across the reviewed studies precluded a comprehensive quantitative meta-analysis of the observed patterns. Instead, we applied a meta-regression analysis using a subset of the reviewed studies to examine the effect of study-specific characteristics on the magnitude of effect sizes. To ensure the robustness of our findings, we further compared our results with models assessing the impact of study characteristics on the sign and significance of the reported peer effects.

While our review combined a large body of evidence from high-income countries with studies from emerging countries such as Brazil, China, and Korea to deepen our understanding of the nature of peer effects, the scarcity of studies from populations outside the United States or

Europe may have had an impact on the external validity of our findings, considering different cultural norms and social dynamics across countries. Our meta-regression indicated that studies from the United States, in particular, did not emerge as a significant predictor of the magnitude, sign, or significance of peer effects in health behaviors. In line with this finding, previous studies have reported similar magnitudes of peer effects across different cultures, such as the United States and China (Nie, Sousa-Poza and He, 2015), as well as cross-cultural similarities in susceptibility to peer effects (Hirata *et al.*, 2015). However, although certain aspects of peer effects seem to transcend cultural boundaries, it is important to note that comprehensive generalizability requires further research. Moreover, most of the included studies investigated peer effects on dietary and physical activity behaviors, whereas only five studies on peer effects in sleep behaviors met our inclusion criteria. Four out of the five studies on how peers may affect individuals' sleep-related behaviors make use of the Add Health data set from the US, with data collected in the mid-1990s. While Add Health has provided important means to inform behavioral and social science, any inference from the findings should be made with caution.

1.5 Conclusion

In this systematic literature review, we synthesized studies adducing several economic and psychological models to explain peer effects, including conformity, social norms and normative information, social comparison, social image, social support, as well as social identity. The evidence on behavioral drivers through which young people's weight-related outcomes are affected by their peers does not deliver a clear understanding of which mechanisms matter the most for this population. Youth may experience heterogeneous effects depending on their own developmental stage, but also conditional on the composition of their peer group and interpersonal dynamics. Although our study design did not allow for a comprehensive meta-analysis, we performed a meta-regression based on a subset of results from quantitatively comparable studies.

The necessity and benefits of seeking to disentangle the different mechanisms underlying peer effects have been discussed by Nakamura, Suhrcke and Zizzo (2017) in a broader context.

Building on and extending their model, we provided further insights into how current empirical findings might inform the design of policy interventions in health-related contexts. Ultimately, while some policy implications can already be inferred from existing evidence, a deeper and more fine-grained understanding of the exact mechanisms underlying peer effects in the present context would allow for more targeted and effective policymaking. Identifying the distinct channels through which peer effects operate in this context could help translate empirical findings into concrete strategies for promoting healthier trajectories among young people. For instance, correcting misperceptions and harnessing the influence of normative information have been shown to positively impact weight-related behaviors. However, this may only be an effective way if individuals actually engage with their environment and learn from their peers, allowing them to retrieve relevant information to update their beliefs and make their own informed choices. Considering that motives differ empirically, as do their implications, the scarcity of studies credibly disentangling the underlying mechanisms of peer effects further accentuates the rationale for more robust evidence.

Bibliography

Aalsma, Matthew C, Melissa Y Carpentier, Faouzi Azzouz, and J Dennis Fortenberry. 2012. "Longitudinal effects of health-harming and health-protective behaviors within adolescent romantic dyads." *Social Science & Medicine*, 74(9): 1444–1451.

Akerlof, George A. 1980. "A theory of social custom, of which unemployment may be one consequence." *The Quarterly Journal of Economics*, 94(4): 749–775.

Ali, Mir M, Aliaksandr Amialchuk, and Frank W Heiland. 2011. "Weight-related behavior among adolescents: the role of peer effects." *PloS one*, 6(6): e21179.

Amialchuk, Aliaksandr, Olugbenga Ajilore, and Kevin Egan. 2019. "The influence of misperceptions about social norms on substance use among school-aged adolescents." *Health economics*, 28(6): 736–747.

Andreoni, James, and B Douglas Bernheim. 2009. "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects." *Econometrica*, 77(5): 1607–1636.

Arduini, Tiziano, Daniela Iorio, and Eleonora Patacchini. 2019. "Weight, reference points, and the onset of eating disorders." *Journal of health economics*, 65: 170–188.

Auld, M Christopher. 2011. "Effect of large-scale social interactions on body weight." *Journal of Health Economics*, 30(2): 303–316.

Balsa, Ana, and Carlos Díaz. 2019. "Social Interactions in Health Behaviors and Conditions." *Oxford Research Encyclopedia of Economics and Finance*.

Bandura, Albert. 1977. *Social learning theory.* Vol. 1, Englewood cliffs Prentice Hall.

Barclay, Kieron J., Christofer Edling, and Jens Rydgren. 2013. "Peer clustering of exercise and eating behaviours among young adults in Sweden: A cross-sectional study of egocentric network data." *BMC Public Health*, 13.

Barnett, Nancy P, Miles Q Ott, Michelle L Rogers, Michelle Loxley, Crystal Linkletter, and Melissa A Clark. 2014. "Peer associations for substance use and exercise in a college student social network." *Health Psychology*, 33(10): 1134.

Bernheim, B Douglas. 1994. "A theory of conformity." *Journal of political Economy*, 102(5): 841–877.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do people mean what they say? Implications for subjective survey data." *American Economic Review*, 91(2): 67–72.

Bicchieri, Cristina, and Erte Xiao. 2009. "Do the right thing: but only if others do so." *Journal of Behavioral Decision Making*, 22(2): 191–208.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of political Economy*, 100(5): 992–1026.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1998. "Learning from the behavior of others: Conformity, fads, and informational cascades." *Journal of economic perspectives*, 12(3): 151–170.

Borenstein, Michael, Larry V Hedges, Julian PT Higgins, and Hannah R Rothstein. 2009. *Introduction to Meta-Analysis.* John Wiley & Sons.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2020. "Peer effects in networks: A survey." *Annual Review of Economics*, 12: 603–629.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman. 2014. "Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions." *Econometrica*, 82(4): 1273–1301.

Card, David, Jochen Kluve, and Andrea Weber. 2010. "Active labour market policy evaluations: A meta-analysis." *The economic journal*, 120(548): F452–F477.

Cawley, John. 2015. "An economy of scales: A selective review of obesity's economic causes, consequences, and solutions." *Journal of health economics*, 43: 244–268.

Cheng, Luanna Alexandra, Gerfeson Mendonça, and José Cazuza De Farias Júnior. 2014. "Physical activity in adolescents: Analysis of the social influence of parents and friends." *Jornal de Pediatria*, 90(1): 35–41.

Christakis, Nicholas A, and James H Fowler. 2007. "The spread of obesity in a large social network over 32 years." *The New England Journal of Medicine*, 357(4): 370–379.

Chung, Sophia Jihey, Anne L Ersig, and Ann Marie McCarthy. 2017. "The influence of peers on diet and exercise among adolescents: a systematic review." *Journal of pediatric nursing*, 36: 44–56.

Clark, Andrew E, and Youenn Lohéac. 2007. ""It wasn't me, it was them!" Social influence in risky behavior by adolescents." *Journal of health economics*, 26(4): 763–784.

Cohen-Cole, Ethan, and Jason M Fletcher. 2008. "Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic." *Journal of health economics*, 27(5): 1382–1387.

Condliffe, Simon, Ebru Isgin, and Brynne Fitzgerald. 2017. "Get thee to the gym! A field experiment on improving exercise habits." *Journal of Behavioral and Experimental Economics*, 70: 23–32.

Cruwys, Tegan, Michael J. Platow, Sarah A. Angullia, Jia Min Chang, Sema E. Diler, Joanne L. Kirchner, Charlotte E. Lentfer, Ying Jun Lim, Aleisha Quarisa, Veronica W.L. Tor, and Amanda L. Wadley. 2012. "Modeling of food intake is moderated by salient psychological group membership." *Appetite*, 58(2): 754–757.

Cunningham, Solveig A, Elizabeth Vaquera, Claire C Maturo, and KM Venkat Narayan. 2012. "Is there evidence that friends influence body weight? A systematic review of empirical research." *Social science & medicine*, 75(7): 1175–1183.

Cutler, David M, and Edward L Glaeser. 2010. "Social interactions and smoking." In *Research findings in the economics of aging*. 123–141. University of Chicago Press.

Després, Jean-Pierre, and Isabelle Lemieux. 2006. "Abdominal obesity and metabolic syndrome." *Nature*, 444(7121): 881–887.

Dulloo, Abdul G, Jennifer L Miles-Chan, and J-P Montani. 2017. "Nutrition, movement and sleep behaviours: their interactions in pathways to obesity and cardiometabolic diseases." *Obesity Reviews*, 18: 3–6.

Egger, Matthias, George Davey Smith, Martin Schneider, and Christoph Minder. 1997. "Bias in meta-analysis detected by a simple, graphical test." *Bmj*, 315(7109): 629–634.

Elkins, Rosemary K, Sonja C Kassenboehmer, and Stefanie Schurer. 2017. "The stability of personality traits in adolescence and young adulthood." *Journal of Economic Psychology*, 60: 37–52.

Fardouly, Jasmine, Rebecca T Pinkus, and Lenny R Vartanian. 2017. "The impact of appearance comparisons made through social media, traditional media, and in person in women's everyday lives." *Body image*, 20: 31–39.

Fortin, Bernard, and Myra Yazbeck. 2015. "Peer effects, fast food consumption and adolescent weight gain." *Journal of Health Economics*, 42: 125–138.

Gächter, Simon, Daniele Nosenzo, and Martin Sefton. 2013. "Peer effects in pro-social behavior: Social norms or social preferences?" *Journal of the European Economic Association*, 11(3): 548–573.

Gesualdo, Chrys, and Martin Pinquart. 2021. "Health behaviors of German university freshmen during COVID-19 in association with health behaviors of close social ties, living arrangement, and time spent with peers." *Health Psychology and Behavioral Medicine*, 9(1): 582–599.

Gioia, Francesca. 2017. "Peer effects on risk behaviour: the importance of group identity." *Experimental economics*, 20(1): 100–129.

Glaeser, Edward L, Bruce I Sacerdote, and Jose A Scheinkman. 2003. "The social multiplier." *Journal of the European Economic Association*, 1(2-3): 345–353.

Golberstein, Ezra, Daniel Eisenberg, and Marilyn F Downs. 2016. "Spillover effects in health service use: Evidence from mental health care using first-year college housing assignments." *Health economics*, 25(1): 40–55.

Graham, Dan J., Jennifer E. Pelletier, Dianne Neumark-Sztainer, Katherine Lust, and Melissa N. Laska. 2013. "Perceived Social-Ecological Factors Associated with Fruit and Vegetable Purchasing, Preparation, and Consumption among Young Adults." *Journal of the Academy of Nutrition and Dietetics*, 113(10): 1366–1374.

Graham, Dan J., Katherine W. Bauer, Sarah Friend, Daheia J. Barr-Anderson, and Dianne Nuemark-Sztainer. 2014. "Personal, behavioral, and socioenvironmental correlates of physical activity among adolescent girls: Cross-sectional and longitudinal associations." *Journal of Physical Activity and Health*, 11(1): 51–61.

Grimaldi, Martina, Valeria Bacaro, Vincenzo Natale, Lorenzo Tonetti, and Elisabetta Crocetti. 2023. "The Longitudinal Interplay between Sleep, Anthropometric Indices, Eating Behaviors, and Nutritional Aspects: A Systematic Review and Meta-Analysis." *Nutrients*, 15(14): 3179.

Harris, Kathleen Mullan. 2013. "The add health study: Design and accomplishments." *Chapel Hill: Carolina Population Center, University of North Carolina at Chapel Hill*, 1: 1–22.

Hawkins, Lily K., Claire Farrow, and Jason M. Thomas. 2020. "Do perceived norms of social media users' eating habits and preferences predict our own food consumption and BMI?" *Appetite*, 149(January): 104611.

Henneberger, Angela K, Dawnsha R Mushonga, and Alison M Preston. 2021. "Peer influence and adolescent substance use: A systematic review of dynamic social network research." *Adolescent Research Review*, 6: 57–73.

Hirata, Elizabeth, Ulrich Kühnen, Roel C.J. Hermans, and Sonia Lippke. 2015. "Modelling of food intake in Brazil and Germany: Examining the effects of self-construals." *Eating Behaviors*, 19: 127–132.

Hsieh, Chih-Sheng, and Hans Van Kippersluis. 2018. "Smoking initiation: Peers and personality." *Quantitative Economics*, 9(2): 825–863.

Jones, Andrew, and Eric Robinson. 2017. "The longitudinal associations between perceived descriptive peer norms and eating and drinking behavior: An initial examination in young adults." *Frontiers in Psychology*, 8(JAN): 1–7.

Jones, Stephen. 1984. *The Economics of Conformism.* Oxford: Blackwell.

Kawaguchi, Daiji. 2004. "Peer effects on substance use among American teenagers." *Journal of Population Economics*, 17(2): 351–367.

Kimbrough, Erik O, and Alexander Vostroknutov. 2016. "Norms make preferences social." *Journal of the European Economic Association*, 14(3): 608–638.

Kimura, Atsushi, Hiroko Tokunaga, Hiroki Sasaki, Masaki Shuzo, Naoki Mukawa, and Yuji Wada. 2021. "Effect of co-eating on unfamiliar food intake among Japanese young adults." *Food Quality and Preference*, 89(October 2020): 104135.

König, Laura M., Helge Giese, F. Marijn Stok, and Britta Renner. 2017. "The social image of food: Associations between popularity and eating behavior." *Appetite*, 114: 248–258.

Krupka, Erin L, and Roberto A Weber. 2013. "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association*, 11(3): 495–524.

Lally, Phillippa, Naomi Bartle, and Jane Wardle. 2011. "Social norms and diet in adolescents." *Appetite*, 57(3): 623–627.

Lee, Keunchul, and Kanghun Lee. 2020. "Relationship of friend/parent exercise participation levels and adolescents' exercise intention/behavior as moderated by action control." *Perceptual and Motor Skills*, 127(2): 347–366.

Leung, Rachel K, John W Toumbourou, and Sheryl A Hemphill. 2014. "The effect of peer influence and selection processes on adolescent alcohol use: a systematic review of longitudinal studies." *Health psychology review*, 8(4): 426–457.

Li, Kaigang, Danping Liu, Denise Haynie, Benjamin Gee, Ashok Chaurasia, Dong Chul Seo, Ronald J. Iannotti, and Bruce G. Simons-Morton. 2016*a*. "Individual, social, and environmental influences on the transitions in physical activity among emerging adults." *BMC Public Health*, 16(1): 1–12.

Li, Kaigang, Denise Haynie, Leah Lipsky, Ronald J. Iannotti, Charlotte Pratt, and Bruce Simons-Morton. 2016*b*. "Changes in moderate-to-vigorous physical activity among older adolescents." *Pediatrics*, 138(4).

Li, Kaigang, Ronald J. Iannotti, Denise L. Haynie, Jessamyn G. Perlus, and Bruce G. Simons-Morton. 2014. "Motivation and planning as mediators of the relation between social support and physical activity among U.S. adolescents: A nationally representative study." *International Journal of Behavioral Nutrition and Physical Activity*, 11(1): 1–9.

Liu, Jinyu, and Suzanne Higgs. 2019. "Social modeling of food intake: No evidence for moderation by identification with the norm referent group." *Frontiers in Psychology*, 10(FEB): 1–9.

Liu, Xiaodong, Eleonora Patacchini, and Edoardo Rainone. 2017. "Peer effects in bedtime decisions among adolescents: a social network model with sampled data." *Econometrics Journal*, 20(3): S103–S125.

Li, Xiaoyu, Ichiro Kawachi, Orfeu M Buxton, Sebastien Haneuse, and Jukka-Pekka Onnela. 2019. "Social network analysis of group position, popularity, and sleep behaviors among US adolescents." *Social Science & Medicine*, 232: 417–426.

Long, Emily, Tyson S. Barrett, and Ginger Lockhart. 2017. "Network-behavior dynamics of adolescent friendships, alcohol use, and physical activity." *Health Psychology*, 36(6): 577–586.

Lopes, Vítor P, Carl Gabbard, and Luis P Rodrigues. 2013. "Physical activity in adolescents: Examining influence of the best friend dyad." *Journal of Adolescent Health*, 52(6): 752–756.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *The review of economic studies*, 60(3): 531–542.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at work." *American Economic Review*, 99(1): 112–45.

Mathieu-Bolh, Nathalie. 2020. "Could obesity be contagious? Social influence, food consumption behavior, and body weight outcomes." *Macroeconomic Dynamics*, 24(8): 1924–1959.

McDade, Thomas W, Laura Chyu, Greg J Duncan, Lindsay T Hoyt, Leah D Doane, and Emma K Adam. 2011. "Adolescents' expectations for the future predict health behaviors in early adulthood." *Social science & medicine*, 73(3): 391–398.

McShane, Blakeley B, Ulf Böckenholt, and Karsten T Hansen. 2016. "Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes." *Perspectives on Psychological Science*, 11(5): 730–749.

Melbye, Elisabeth Lind, and Merete Hagen Helland. 2018. "Soft drinks for lunch? Self-control, intentions and social influences." *British Food Journal*.

Mendonça, Gerfeson, and José Cazuza de Farias Júnior. 2015. "Physical activity and social support in adolescents: analysis of different types and sources of social support." *Journal of sports sciences*, 33(18): 1942–1951.

Meng, Jingbo, Wei Peng, Soo Yun Shin, and Minwoong Chung. 2017. "Online self-tracking groups to increase fruit and vegetable intake: A small-scale study on mechanisms of group effect on behavior change." *Journal of Medical Internet Research*, 19(3): 1–15.

Montgomery, Shannon C, Michael Donnelly, Prachi Bhatnagar, Angela Carlin, Frank Kee, and Ruth F Hunter. 2020. "Peer social network processes and adolescent health behaviors: A systematic review." *Preventive medicine*, 130: 105900.

Moola, S, Z Munn, C Tufanaru, E Aromataris, K Sears, R Sfetcu, M Currie, R Qureshi, P Mattis, K Lisy, and P-F Mu. 2020. *Chapter 7: Systematic reviews of etiology and risk.* The Joanna Briggs Institute.

Mora, Toni, and Joan Gil. 2013. "Peer effects in adolescent BMI: evidence from Spain." *Health Economics*, 22(5): 501–516.

Morrissey, Joanna L., Kathleen F. Janz, Elena M. Letuchy, Shelby L. Francis, and Steven M. Levy. 2015. "The effect of family and friend support on physical activity through adolescence: A longitudinal study." *International Journal of Behavioral Nutrition and Physical Activity*, 12(1): 1–9.

Nakamura, Ryota, Marc Suhrcke, and Daniel John Zizzo. 2017. "A triple test for behavioral economics models and public health policy." *Theory and Decision*, 83(4): 513–533.

Nieminen, Pentti. 2022. "Application of standardized regression coefficient in meta-analysis." *BioMedInformatics*, 2(3): 434–458.

Nie, Peng, Alfonso Sousa-Poza, and Xiaobo He. 2015. "Peer effects on childhood and adolescent obesity in China." *China Economic Review*, 35: 47–69.

Nix, Elizabeth, and Heidi J. Wengreen. 2017. "Social approval bias in self-reported fruit and vegetable intake after presentation of a normative message in college students." *Appetite*, 116: 552–558.

Nyborg, Karine, and Mari Rege. 2003. "On social norms: the evolution of considerate smoking behavior." *Journal of Economic Behavior & Organization*, 52(3): 323–340.

Nyborg, Karine, John M Anderies, Astrid Dannenberg, Therese Lindahl, Caroline Schill, Maja Schlüter, W Neil Adger, Kenneth J Arrow, Scott Barrett, Stephen Carpenter, and Aart de Zeeuw. 2016. "Social norms as solutions." *Science*, 354(6308): 42–43.

OECD. 2019. *The Heavy Burden of Obesity.*

Office of Disease Prevention and Health Promotion. 2022. "Healthy People 2030." Accessed: October 25, 2022.

Oftedal, Stina, Corneel Vandelanotte, and Mitch J Duncan. 2019. "Patterns of diet, physical activity, sitting and sleep are associated with socio-demographic, behavioural, and health-risk indicators in adults." *International journal of environmental research and public health*, 16(13): 2375.

Ortiz-Ospina, Esteban. 2019. "The rise of social media." Accessed: October 17, 2022.

Page, Matthew J, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, and David Moher. 2021. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." *Systematic reviews*, 10(1): 1–11.

Pelletier, Jennifer E., Dan J. Graham, and Melissa N. Laska. 2014. "Social norms and dietary behaviors among young adults." *American Journal of Health Behavior*, 38(1): 144–152.

Perkins, Jessica M, H Wesley Perkins, and David W Craig. 2018. "Misperceived norms and personal sugar-sweetened beverage consumption and fruit and vegetable intake among students in the United States." *Appetite*, 129: 82–93.

Pratschke, Jonathan, and Giovanni Abbiati. 2023. ""Like with like" or "do like?" Modeling peer effects in the classroom." *Social Science Quarterly*, 104(3): 265–280.

Robinson, Eric, and Suzanne Higgs. 2013. "Food choices in the presence of 'healthy' and 'unhealthy' eating partners." *British Journal of Nutrition*, 109(4): 765–771.

Robinson, Eric, Ellis Harris, Jason Thomas, Paul Aveyard, and Suzanne Higgs. 2013. "Reducing high calorie snack food in young adults: A role for social norms and health based messages." *International Journal of Behavioral Nutrition and Physical Activity*, 10: 1–8.

Robinson, Eric, Helen Benwell, and Suzanne Higgs. 2013. "Food intake norms increase and decrease snack food intake in a remote confederate study." *Appetite*, 65: 20–24.

Robinson, Eric, Roy Otten, and Roel CJ Hermans. 2016. "Descriptive peer norms, self-control and dietary behaviour in young adults." *Psychology & health*, 31(1): 9–20.

Rosin, Odelia. 2008. "The economic causes of obesity: a survey." *Journal of Economic Surveys*, 22(4): 617–647.

Serenko, Alexander, Ofir Turel, and Hafsa Bohonis. 2021. "The impact of social networking sites use on health-related outcomes among UK adolescents." *Computers in Human Behavior Reports*, 3: 100058.

Shoham, David A., Liping Tong, Peter J. Lamberson, Amy H. Auchincloss, Jun Zhang, Lara Dugas, Jay S. Kaufman, Richard S. Cooper, and Amy Luke. 2012. "An actor-based model of social network influence on adolescent body size, screen time, and playing sports." *PLoS ONE*, 7(6).

Simpkins, Sandra D., David R. Schaefer, Chara D. Price, and Andrea E. Vest. 2013. "Adolescent friendships, BMI, and physical activity: Untangling selection and influence through longitudinal social network analysis." *Journal of Research on Adolescence*, 23(3): 537–549.

Stanley, Tom D, and Hristos Doucouliagos. 2012. *Meta-regression analysis in economics and business.* Routledge.

Swinburn, Boyd A, Vivica I Kraak, Steven Allender, Vincent J Atkins, Phillip I Baker, Jessica R Bogard, Hannah Brinsden, Alejandro Calvillo, Olivier De Schutter, Raji Devarajan, and William H Dietz. 2019. "The global syndemic of obesity, undernutrition, and climate change: the Lancet Commission report." *The Lancet*, 393(10173): 791–846.

Trogdon, Justin G, James Nonnemaker, and Joanne Pais. 2008. "Peer effects in adolescent overweight." *Journal of health economics*, 27(5): 1388–1399.

Tufanaru, C, Z Munn, E Aromataris, J Campbell, and L Hopp. 2020. *Chapter 3: Systematic reviews of effectiveness.* The Joanna Briggs Institute.

United Nations General Assembly. 2002. "General Assembly Resolutions, A/RES/56/117."

Viechtbauer, Wolfgang. 2010. "Conducting meta-analyses in R with the metafor package." *Journal of Statistical Software*, 36(3): 1–48.

Wang, Cheng, Omar Lizardo, and David S. Hachen. 2021. "Using Fitbit data to examine factors that affect daily activity levels of college students." *PLoS ONE*, 16(1 January): 1–20.

Wang, Cheng, Stephen Mattingly, Jessica Payne, Omar Lizardo, and David S Hachen. 2021. "The impact of social networks on sleep among a cohort of college students." *SSM - Population Health*, 16: 100937.

Wardle, Jane, Yoichi Chida, E Leigh Gibson, Katriina L Whitaker, and Andrew Steptoe. 2011. "Stress and adiposity: a meta-analysis of longitudinal studies." *Obesity*, 19(4): 771–778.

Wengreen, Heidi J., Elizabeth Nix, and Gregory J. Madden. 2017. "The effect of social norms messaging regarding skin carotenoid concentrations among college students." *Appetite*, 116: 39–44.

World Health Organization. 2016. "Physical activity strategy for the WHO European Region 2016–2025."

World Health Organization. 2022. "World health statistics 2022: monitoring health for the SDGs, sustainable development goals."

Yakusheva, Olga, Kandice Kapinos, and Marianne Weiss. 2011. "Peer effects and the Freshman 15: Evidence from a natural experiment." *Economics and Human Biology*, 9(2): 119–132.

Young, H. Peyton. 2015. "The Evolution of Social Norms." *Annual Review of Economics*, 7(1): 359–387.

Yuan, Changzheng, Jun Lv, and Tyler J Vanderweele. 2013. "An Assessment of Health Behavior Peer Effects in Peking University Dormitories : A Randomized Cluster- Assignment Design for Interference." *PLoS ONE*, 8(9): 6.

Zafar, Basit. 2011. "An experimental investigation of why individuals conform." *European Economic Review*, 55(6): 774–798.

Zhylyevskyy, Oleksandr, Helen H. Jensen, Steven B. Garasky, Carolyn E. Cutrona, and Frederick X. Gibbons. 2013. "Effects of Family, Friends, and Relative Prices on Fruit and Vegetable Consumption by African Americans." *Southern Economic Journal*, 80(October 2012): 226–251.

Appendix

A1.1 Conceptual framework

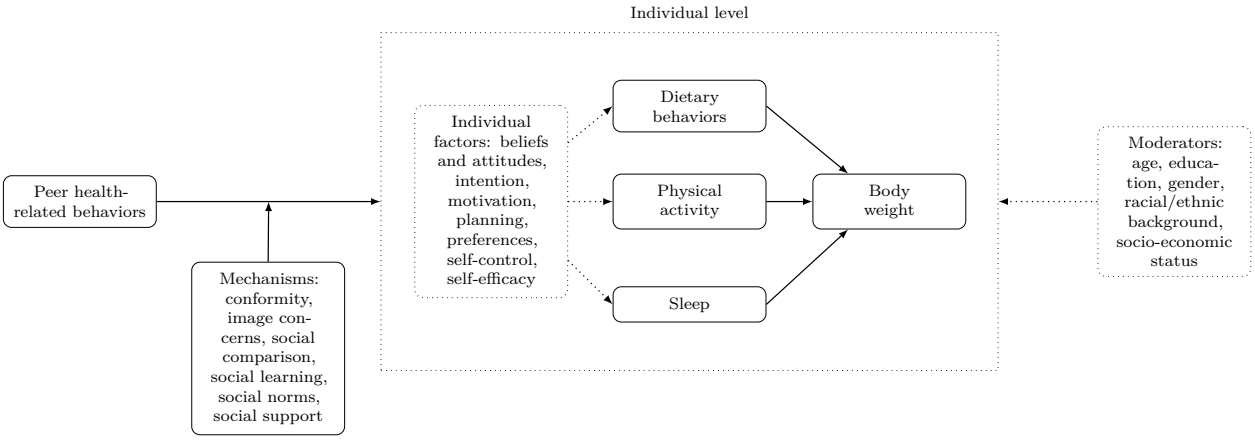


Figure A1.1: Conceptual framework for peer effects on weight-related behaviors in young people

Notes: The conceptual model builds on the work of Dulloo, Miles-Chan and Montani (2017); Grimaldi *et al.* (2023), and Oftedal, Vandelanotte and Duncan (2019) as initial foundations and was further expanded through the comprehensive literature review conducted for this study. The figure illustrates the directional influence and interaction of environmental (peer) and individual factors shaping weight-related behaviors. While this systematic review focuses on peer behaviors and their underlying mechanisms, it is important to acknowledge the presence of additional influential factors, such as individual and demographic moderating factors (represented by dashed lines), which fall outside the scope of this study.

A1.2 Search strategy and search terms

Table A1.1: Search terms used in the literature review

Database, Search Date	Key Concepts			Results
	Peer effects	Weight-related behavior	Population (youth)	
Web of Science 9/2/2022	TS=(peer* OR "peer effect*" OR "peer influence*" OR "social influence*" OR "social network*" OR friend* OR "friend* effect*" OR "friend* network*" OR "interpersonal relation*")	TS=("health behavio?r*" OR "health outcome*" OR diet OR eating OR "food choice" OR "food intake" OR "health* diet" OR "health* eating" OR exercis* OR gym OR "physical* activ*" OR sport* OR sleep OR bedtime)	TS=(adolescent* OR teen* OR youth OR student* OR "young people" OR "young adult*" OR "emerging adult*")	7.563
EconLit 3/3/2022	AB ((peer* OR "peer effect*" OR "peer influence*" OR "social influence*" OR "social network*" OR friend* OR "friend* effect*" OR "friend* network*" OR "interpersonal relation*")) OR TI ((peer* OR "peer effect*" OR "peer influence*" OR "social influence*" OR "social network*" OR friend* OR "friend* effect*" OR "friend* network*" OR "interpersonal relation*"))	AB (("health behavio?r*" OR "health outcome*" OR diet OR eating OR "food choice" OR "food intake" OR "health* diet" OR "health* eating" OR exercis* OR gym OR "physical* activ*" OR sport* OR sleep OR bedtime)) OR TI (("health behavio?r*" OR "health outcome*" OR diet OR eating OR "food choice" OR "food intake" OR "health* diet" OR "health* eating" OR exercis* OR gym OR "physical* activ*" OR sport* OR sleep OR bedtime))	AB ((adolescent* OR teen* OR youth OR student* OR "young people" OR "young adult*" OR "emerging adult*")) OR TI ((adolescent* OR teen* OR youth OR student* OR "young people" OR "young adult*" OR "emerging adult*"))	63
MEDLINE 5/3/2022	AB ((peer* OR "peer effect*" OR "peer influence*" OR "social influence*" OR "social network*" OR friend* OR "friend* effect*" OR "friend* network*" OR "interpersonal relation*")) OR TI ((peer* OR "peer effect*" OR "peer influence*" OR "social influence*" OR "social network*" OR friend* OR "friend* effect*" OR "friend* network*" OR "interpersonal relation*"))	AB (("health behavio?r*" OR "health outcome*" OR diet OR eating OR "food choice" OR "food intake" OR "health* diet" OR "health* eating" OR exercis* OR gym OR "physical* activ*" OR sport* OR sleep OR bedtime)) OR TI (("health behavio?r*" OR "health outcome*" OR diet OR eating OR "food choice" OR "food intake" OR "health* diet" OR "health* eating" OR exercis* OR gym OR "physical* activ*" OR sport* OR sleep OR bedtime))	AB ((adolescent* OR teen* OR youth OR student* OR "young people" OR "young adult*" OR "emerging adult*")) OR TI ((adolescent* OR teen* OR youth OR student* OR "young people" OR "young adult*" OR "emerging adult*"))	1.364
SocINDEX 2/3/2022	AB ((peer* OR "peer effect*" OR "peer influence*" OR "social influence*" OR "social network*" OR friend* OR "friend* effect*" OR "friend* network*" OR "interpersonal relation*")) OR TI ((peer* OR "peer effect*" OR "peer influence*" OR "social influence*" OR "social network*" OR friend* OR "friend* effect*" OR "friend* network*" OR "interpersonal relation*"))	AB (("health behavio?r*" OR "health outcome*" OR diet OR eating OR "food choice" OR "food intake" OR "health* diet" OR "health* eating" OR exercis* OR gym OR "physical* activ*" OR sport* OR sleep OR bedtime)) OR TI (("health behavio?r*" OR "health outcome*" OR diet OR eating OR "food choice" OR "food intake" OR "health* diet" OR "health* eating" OR exercis* OR gym OR "physical* activ*" OR sport* OR sleep OR bedtime))	AB ((adolescent* OR teen* OR youth OR student* OR "young people" OR "young adult*" OR "emerging adult*")) OR TI ((adolescent* OR teen* OR youth OR student* OR "young people" OR "young adult*" OR "emerging adult*"))	233

Abbreviations: AB = Abstract; TI = Title; TS = Topic (i.e., title, abstract, and keywords).

A1.3 Methodological quality using Joanna Briggs Institute (JBI) checklists for quantitative studies

A1.3.1 JBI criteria for cross-sectional and longitudinal studies

Q1 Were the criteria for inclusion in the sample clearly defined?

Q2 Were the study subjects and the setting described in detail?

Q3 Was the exposure measured in a valid and reliable way?

Q4 Were confounding factors identified?

Q5 Were strategies to deal with confounding factors stated?

Q6 Were the outcomes measured in a valid and reliable way?

Q7 Was appropriate statistical analysis used?

Table A1.2: Appraisal of cross-sectional studies using JBI criteria

	JBI Critical Appraisal Items								
Article	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total	%
Barclay, Edling and Rydgren 2013, *BMC Publ Health*	1	1	1	1	1	0	0	5	71
Barnett et al. 2014, *Health Psychol*	1	1	1	1	1	0	1	6	86
Belanger and Patrick 2018, *J Phys Activ Health*	1	1	1	1	0	0	1	5	71
Cheng, Mendonça and de Farias Júnior 2014, *J Pediatr*	1	1	1	1	1	0	1	6	86
Gesualdo and Pinquart 2021, *Health Psychol Behav Med*	1	1	0	1	1	0	1	5	71
Graham et al. 2013, *J Acad Nutr Diet*	1	1	0	1	1	0	1	5	71
Graham et al. 2014, *J Phys Activ Health*	1	1	0	1	1	0	1	5	71
Hawkins, Farrow and Thoma 2020, *Appetite*	1	1	1	1	1	0	1	6	86
Lally, Bartle and Wardle 2011, *Appetite*	1	1	1	1	1	0	1	6	86
Lee and Lee 2020, *Percept Mot Skills*	1	1	0	0	0	0	1	3	43
Li et al. 2014, *Int J Behav Nutr Phys Activ*	0	1	0	1	1	0	1	4	57
Li et al. 2019, *Soc Sci Med*	1	1	1	1	1	0	1	6	86
Liu, Patacchini and Rainone 2017, *Econometrics J*	1	1	1	1	1	0	1	6	86
Lopes, Gabbard and Rodrigues 2013, *J Adolesc Health*	0	1	1	1	0	0	0	3	43
Melbye and Helland 2018, *Br Food J*	1	0	0	1	1	0	1	4	57
Mendonça and de Farias Júnior 2015, *J Sports Sci*	1	1	1	1	1	0	1	6	86
Pelletier, Graham and Laska 2014, *Am J Health Behav*	1	1	0	1	1	0	1	5	71
Perkins, Perkins and Craig 2018, *Appetite*	1	1	1	1	1	0	1	6	86
Robinson, Otten and Hermans 2016, *Psychol Health*	0	1	1	1	1	0	1	5	71
Yuan, Lv and Vanderweele 2013, *PLoS ONE*	1	1	1	1	1	0	1	6	86
Zhylyevskyy et al. 2013, *South Econ J*	1	1	1	1	1	0	1	6	86

Scoring: 1 = criterion met; 0 = not met. Total scores are out of a maximum of 7. Quality rating based on percentage of criteria met: $\leq$ 60% = low; 61–80% = medium; $>$ 80% = good.

Table A1.3: Appraisal of longitudinal studies using JBI criteria

	JBI Critical Appraisal Items								
Article	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total	%
Aalsma et al. 2012, *Soc Sci Med*	1	1	1	0	0	0	1	4	57
Ali, Amialchuk and Heiland 2011, *PLoS ONE*	1	1	1	1	1	0	1	6	86
Fortin and Yazbeck 2015, *J Health Econ*	1	1	1	1	1	0	1	6	86
Jones and Robinson 2017, *Front Psychol*	1	0	0	1	1	0	1	4	57
Li et al. 2016a, *Pediatrics*	1	1	1	1	1	1	1	7	100
Li et al. 2016b, *BMC Publ Health*	1	1	0	1	1	0	1	5	71
Long, Barrett and Lockhart 2017, *Health Psychol*	1	1	1	1	1	0	1	6	86
Morrissey et al. 2015, *Int J Behav Nutr Phys Activ*	1	1	1	1	0	1	1	6	86
Shoham et al. 2012, *PLoS ONE*	1	1	1	1	1	0	1	6	86
Simpkins et al. 2013, *J Res Adolesc*	1	1	1	1	1	0	1	6	86
Wang et al. 2021, *SSM Popul Health*	1	1	1	1	1	1	1	7	100
Wang, Lizardo and Hachen 2021, *PLoS ONE*	1	1	1	1	1	1	1	7	100

Scoring: 1 = criterion met; 0 = not met. Total scores are out of a maximum of 7. Quality rating based on percentage of criteria met: $\leq$ 60% = low; 61–80% = medium; $>$ 80% = good.

A1.3.2 JBI criteria for (quasi-)experimental studies

Q1 Is it clear in the study what the 'cause' is and what the 'effect' is?

Q2 Were the participants included in any comparisons similar?

Q3 Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest?

Q4 Was there a control group?

Q5 Were there multiple measurements of the outcome both pre- and post-intervention/exposure?

Q6 Was follow-up complete, and if not, were differences between groups in terms of their follow-up adequately described and analyzed?

Q7 Were the outcomes of participants included in any comparisons measured in the same way?

Q8 Were outcomes measured in a reliable way?

Q9 Was appropriate statistical analysis used?

Table A1.4: Appraisal of (quasi-)experimental studies using JBI criteria

	JBI Critical Appraisal Items											
Article	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Max	Total	%
Condliffe, Işgın and Fitzgerald 2017, *J Behav Exp Econ*	1	1	1	1	1	1	1	1	1	9	9	100
Cruwys et al. 2012, *Appetite*	1	1	1	1	NA	NA	1	1	1	7	7	100
Hirata et al. 2015, *Eat Behav*	1	NR	1	NA	1	1	1	1	1	8	7	88
Kimura et al. 2021, *Food Quality and Preference*	1	NR	1	1	0	NA	1	1	1	8	6	75
König et al. 2017, *Appetite*	1	0	1	NA	NA	1	1	1	1	7	6	86
Liu and Higgs 2019, *Front Psychol*	1	1	1	1	0	NA	1	1	1	8	7	88
Meng et al. 2017, *J Med Internet Res*	1	0	1	1	1	1	1	0	1	9	7	78
Nix and Wengreen 2017, *Appetite*	1	1	1	1	1	1	1	0	1	9	8	89
Robinson and Higgs 2013, *Br J Nutr*	1	1	1	1	NA	NA	1	1	1	7	7	100
Robinson et al. 2013, *Int J Behav Nutr Phys Activ*	1	0	1	1	NA	NA	1	0	1	7	5	71
Robinson, Benwell and Higgs 2013, *Appetite*	1	0	1	1	NA	NA	1	1	1	7	6	86
Wengreen, Nix and Madden 2017, *Appetite*	1	1	1	1	1	NA	1	1	1	8	8	100
Yakusheva, Kapinos and Weiss 2011, *Econ Hum Biol*	1	1	1	NA	1	1	1	1	1	8	8	100

Scoring: 1 = criterion met; 0 = not met; NA = Not applicable; NR = Not reported. Quality rating based on percentage of criteria met: $\leq$ 60% = low; 61–80% = medium; $>$ 80% = good.

A1.4 Mechanisms underlying peer effects in included studies

Table A1.5: Proposed peer influence mechanisms in reviewed studies

Study	Proposed mechanisms[a,b]	Location in article	Empirically tested[c]
Ali, Amialchuk and Heiland (2011)	social norms	discussion	N
Cheng, Mendonça and Farias Júnior (2014)	social support	methods/model	N
Condliffe, Isgin and Fitzgerald (2017)	information feedback; conformity	results	**Y**
Cruwys *et al.* (2012)	social norms; social identity	introduction, results	**Y**
Graham *et al.* (2014)	social support	results	N
Hawkins, Farrow and Thomas (2020)	descriptive norms	introduction, discussion	**Y**
Hirata *et al.* (2015)	descriptive norms	results	**Y**
Jones and Robinson (2017)	social norms	introduction, discussion	**Y**
Kimura *et al.* (2021)	modeling; social comparison	results	N
König *et al.* (2017)	social image; social identity	theory/conceptual framing, results	**Y**
Lally, Bartle and Wardle (2011)	descriptive norms	results	**Y**
Lee and Lee (2020)	social support	discussion	N
Li *et al.* (2014)	social support	theory/conceptual framing, discussion	N
Liu and Higgs (2019)	descriptive norms; social identity	theory/conceptual framing, results	**Y**
Liu, Patacchini and Rainone (2017)	social norms; conformity	appendix, theory/conceptual framing	$\mathbf{Y^m}$
Lopes, Gabbard and Rodrigues (2013)	social support	discussion	N
Melbye and Helland (2018)	social norms	introduction, results	**Y**
Mendonça and Farias Júnior (2015)	social support	results	N
Meng *et al.* (2017)	social comparison; social learning; social norms	results	**Y**
Morrissey *et al.* (2015)	social support	results	N
Nix and Wengreen (2017)	descriptive norms; information feedback; social approval	results	**Y**
Pelletier, Graham and Laska (2014)	descriptive norms; social proximity	results	**Y**
Perkins, Perkins and Craig (2018)	social norms	results	**Y**
Robinson and Higgs (2013)	descriptive norms	results	**Y**
Robinson *et al.* (2013)	descriptive norms	results	**Y**
Robinson, Benwell and Higgs (2013)	descriptive norms; social approval	results	**Y**
Robinson, Otten and Hermans (2016)	descriptive norms	discussion	**Y**
Simpkins *et al.* (2013)	social norms	discussion	N
Wang, Lizardo and Hachen (2021)	descriptive norms	discussion	N
Wengreen, Nix and Madden (2017)	descriptive norms; social approval	discussion	N
Yakusheva, Kapinos and Weiss (2011)	social learning	discussion	N
Yuan, Lv and Vanderweele (2013)	social learning	methods/model, discussion	N
Zhylyevskyy *et al.* (2013)	social norms	introduction, theory/conceptual framing	$\mathbf{Y^m}$

[a] Mechanism coding follows the terminology and description used in each study.
[b] *Social norms* denotes norm-based mechanisms where the type of norm, i.e., *descriptive* (what others do) or *injunctive* (what others approve), is unspecified. *Descriptive norms* refers to influence via peers' actual or perceived behaviors.
[c] Empirically tested: **Y** = Mechanism directly analyzed using empirical data; $\mathbf{Y^m}$ = Mechanism estimated via structural or model-based inference, relying on theoretical assumptions rather than exogenous variation (e.g., random assignment, instrumental variables). N = Mechanism proposed or discussed, not empirically tested.

Several studies are excluded from Table A1.5 because they do not explicitly develop or test specific mechanisms of peer influence, specifically: Aalsma *et al.* (2012); Barclay, Edling and Rydgren (2013); Barnett *et al.* (2014); Fortin and Yazbeck (2015); Gesualdo and Pinquart (2021); Graham *et al.* (2013); Li *et al.* (2016*a*,*b*, 2019); Long, Barrett and Lockhart (2017); Shoham *et al.* (2012), and Wang *et al.* (2021).

A1.5 Additional meta-regression analysis results

Table A1.6: Meta-regression of study characteristics on effect sizes in dietary behavior

	DV: Effect size	
Moderators	(1)	(2)
Sample characteristics		
Sample size n	0.000^{*}	-0.000^{+}
	(0.000)	(0.000)
Country = US	-0.177^{*}	0.050
	(0.081)	(0.074)
University sample	0.063	−0.170
	(0.110)	(0.109)
Quality rating	−0.001	0.000
	(0.004)	(0.003)
Author = Perkins et al. (2018)		0.901^{**}
		(0.226)
Constant	0.130	0.367
	(0.363)	(0.256)
Observations	21	21

Standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. $^{+}$ indicates marginal significance at the 10% level.

Table A1.7: Pooled probit model with clustered standard errors: effect of sample and study characteristics on the likelihood of reporting positive and significant peer effects.

	DV: Positive and significant effect (binary)		
Moderators	(1)	(2)	(3)
Sample characteristics			
Sample size n	0.000^{*}	0.000^{**}	0.000^{***}
	(0.000)	(0.000)	(0.000)
Country = US	-0.609^{*}	-0.651^{*}	-0.600
	(0.321)	(0.377)	(0.369)
University sample	0.382	0.706^{*}	0.658
	(0.389)	(0.400)	(0.444)
QA rating	0.003	0.008	0.000
	(0.010)	(0.011)	(0.012)
Study characteristics			
Health behavior = PA		0.654^{*}	0.727^{*}
		(0.360)	(0.380)
Health behavior = Sleep		-0.962^{**}	-0.753^{*}
		(0.364)	(0.440)
Study design = Experiment			0.796
			(0.510)
Constant	0.334	-0.443	-0.118
	(0.659)	(0.775)	(0.856)
Observations	111	111	111

Standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. $^{+}$ indicates marginal significance at the 10% level.

Forest plots by subgroup: Individual studies on dietary behavior

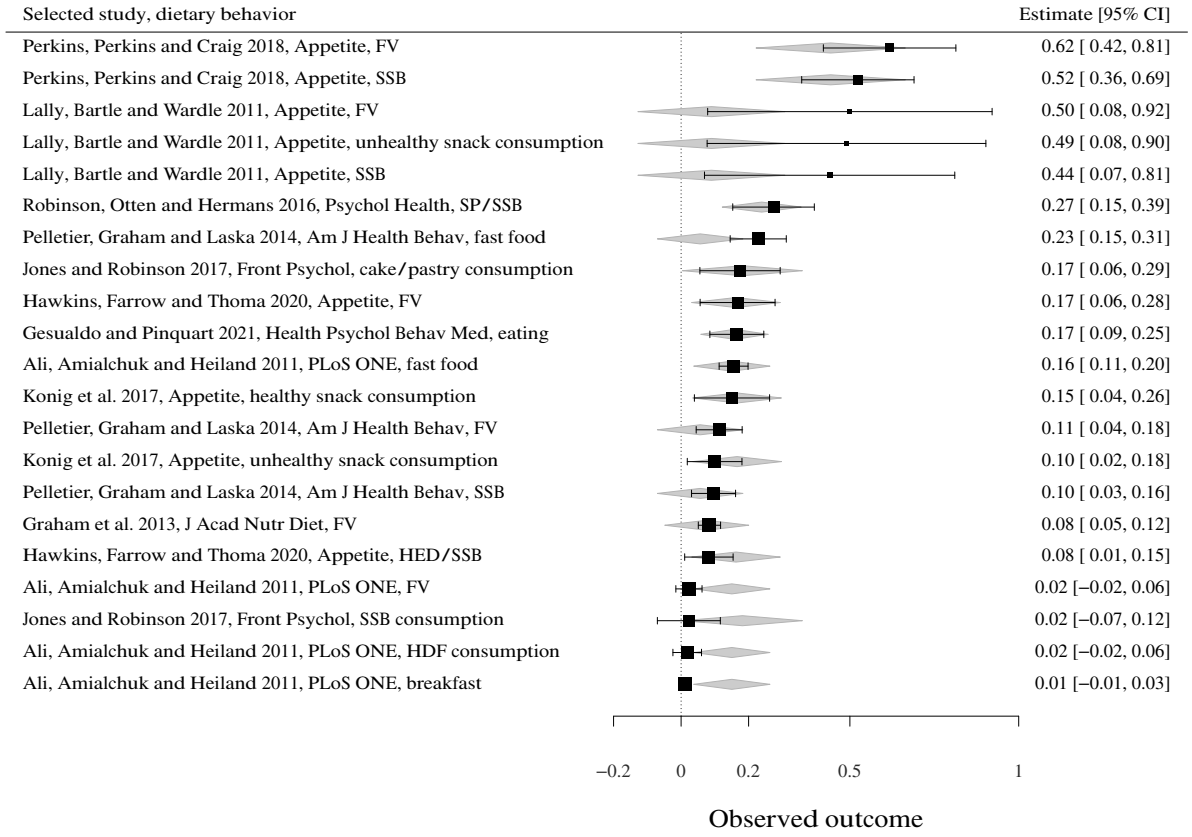


Figure A1.2: Forest plot of the standardized regression coefficients and global effects for dietary behaviors

Notes: Squares represent effect sizes, their sizes varying based on their applied regression weight. 95% Confidence intervals (CI) are depicted as horizontal lines. Diamonds represent the overall effect of all moderators included in the model.

Forest plots by subgroup: Individual studies on physical (in)activity

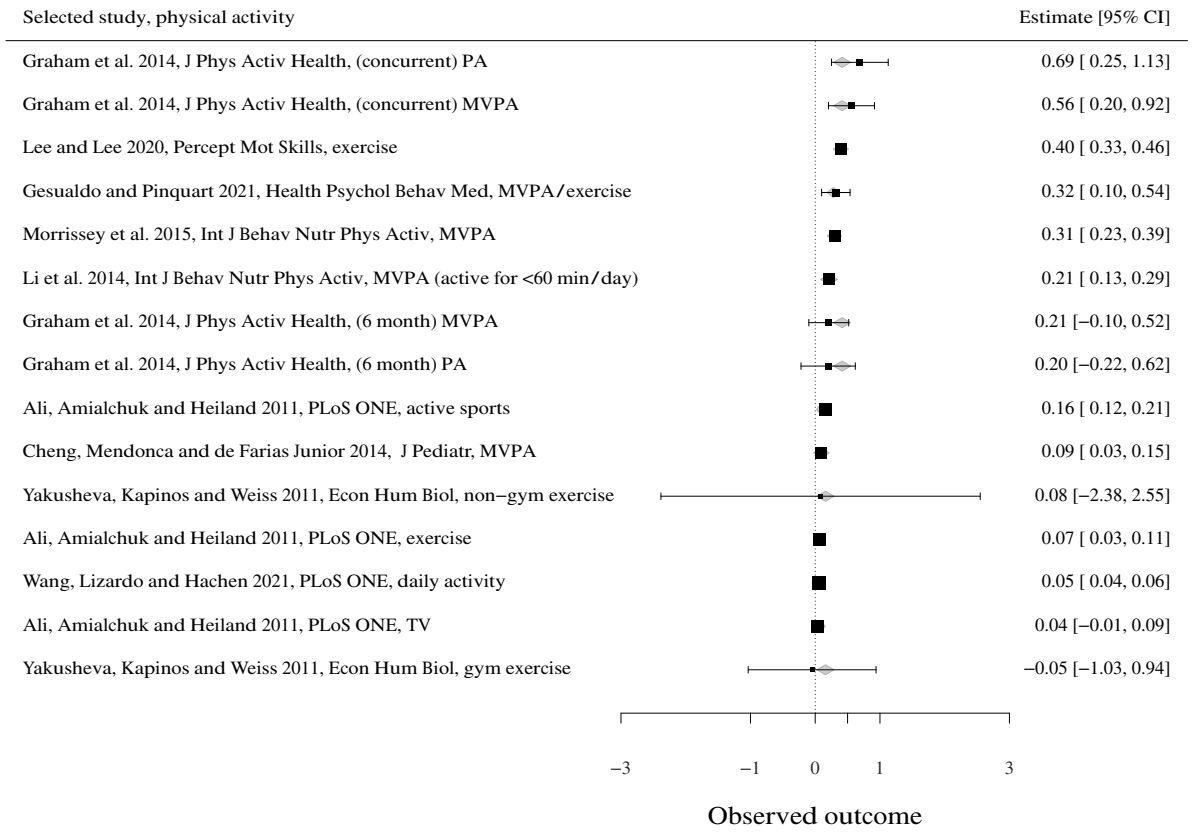
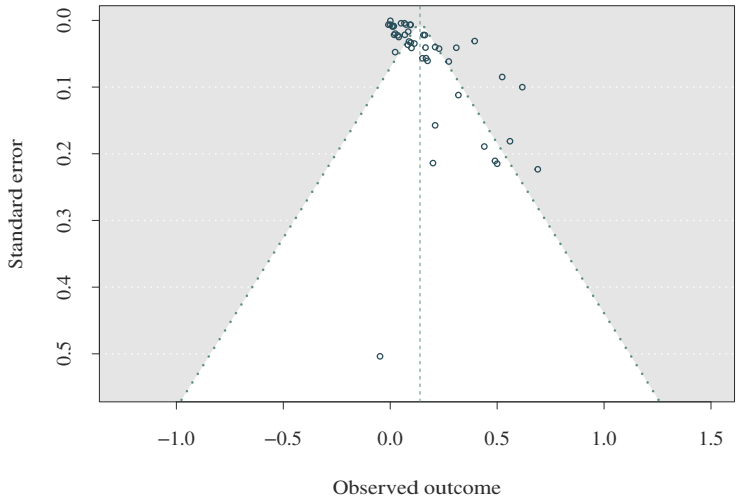


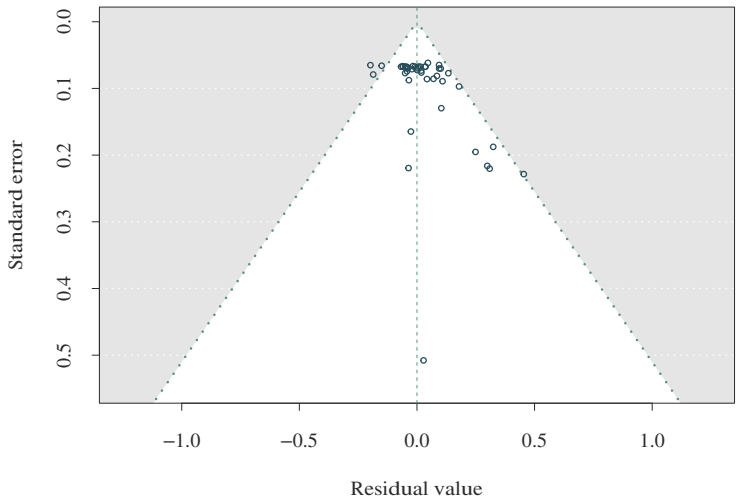
Figure A1.3: Forest plot of the standardized regression coefficients and global effects for physical (in)activity

Notes: Squares represent effect sizes, their sizes varying based on their applied regression weight. 95% Confidence intervals (CI) are depicted as horizontal lines. Diamonds represent the overall effect of all moderators included in the model.

Funnel plots on the effect sizes and residuals



(a) Effect sizes



(b) Residuals

Figure A1.4: Funnel plots of (a) effect sizes and (b) residuals, including all 19 eligible studies

Notes: Panel (a) shows funnel plots of observed effect sizes; (b) shows residuals from the meta-regression model with moderators. The dashed vertical line in (a) marks the pooled effect size; in (b), zero. Absence of null or negative effects suggests possible publication bias. Egger's regression test on the residuals (b) indicates statistically significant funnel plot asymmetry. A simple selection model further indicated substantially lower publication probabilities for studies reporting non-significant results. One extreme outlier (SE = 1.258; Yakusheva, Kapinos and Weiss, 2011) was excluded from the plot for clarity.

Social comparison and positional preferences in health and economic behavior: experimental evidence from the US and UK

Social comparison and positional preferences in health and economic behavior: Experimental evidence from the US and UK

2.1 Introduction

Individuals care deeply about where they stand in society and frequently engage in comparisons with those they deem relevant. The idea that utility depends not only on absolute levels of consumption and well-being, but also on relative comparisons with others, has long been recognized in economics. Veblen's seminal work on conspicuous consumption was among the first to bring attention to how individuals signal social status through the consumption of certain goods, introducing the concept of *positional goods* – goods that are valued more for the status they confer than for their intrinsic utility. Subsequent research has linked positional concerns to a range of behavioral and policy-relevant outcomes, including consumption externalities, tax preferences, and subjective well-being (Clark and d'Ambrosio, 2015; Frank, 2008). Despite their importance, positional concerns have often been overlooked in policy planning, especially in domains traditionally seen as non-competitive, such as health. Health has typically been regarded as less prone to positional concerns, partly because it is commonly perceived as universally accessible and as a fundamental human right, which presumably diminishes incentives for competitive or status-driven behaviors. However, from a welfare perspective, such oversight may be costly: if individuals derive utility from being relatively healthy or having better access to healthcare than others, such preferences may lead to inefficient allocations or unintended behavioral spillovers (Hemenway, 2021).

Theoretical models suggest that accounting for positional concerns can improve predictions and guide more effective policies. Status-seeking motivations influence decisions ranging from labor supply to investment in education or health (Postlewaite, 1998). These concerns may also justify higher marginal income tax rates or affect the willingness to pay for public goods when relative standing influences utility (Aronsson and Johansson-Stenman, 2008; Frank, 1985). In health, positional preferences can drive decisions such as dietary choices or

healthcare utilization, leading to disparities in health outcomes (Hemenway, 2021; Mathieu-Bolh and Wendner, 2020). These decisions not only affect individual well-being but also the efficiency and equity of health and healthcare provision. If status-seeking leads individuals to underinvest in less observable but highly beneficial health behaviors, or to overconsume visible but low-value services, the result can be resource misallocation and strain on public health systems (Hemenway, 2021).

This chapter contributes to the literature on positional concerns by focusing on domains typically regarded as less status-driven, particularly health and healthcare. While positional preferences are well-documented for income, leisure, and housing (Alpizar, Carlsson and Johansson-Stenman, 2005; Carlsson, Johansson-Stenman and Martinsson, 2007; Frank, 1985; Grolleau, Ibanez and Mzoughi, 2012), their role in health-related behaviors and outcomes remains comparatively underexplored. Although traditional views suggest that health is less subject to positional comparison (Grolleau and Saïd, 2008; Solnick and Hemenway, 2005; Wouters *et al.*, 2015), this contrasts with evidence showing that individuals' health perceptions and behaviors are indeed shaped by comparative contexts (Carrieri, 2012; Blanchflower, Van Landeghem and Oswald, 2009). This further aligns with the notion of *conspicuous health behaviors*, which refer to health-related actions, often intentionally displayed, to signal social status, and thus clearly suggesting that positional motives extend into health domains (Mujcic and Frijters, 2015).

Yet, when status concerns drive health investments, more cost-effective and equitable health strategies may be neglected. Status-driven preferences can inflate demand for visible but less impactful services (e.g., elective procedures, brand-name medications), while underfunding preventive measures such as exercise, vaccinations, or healthy diets. These behaviors can 'hollow out' public health systems and misallocate resources away from those who would benefit most (Mumtaz *et al.*, 2013). For example, Huțu *et al.* (2024) document how moral hazard in insured populations increases spending on surgical procedures and branded pharmaceuticals, crowding out investment in basic but high-impact public health services. From a welfare perspective, these findings highlight the importance of understanding how positional preferences affect individual behavior and, in turn, systemic outcomes.

Building on this literature, this chapter examines when and why individuals value being

better off than others, even if it comes at the cost of being worse off in absolute terms. Our analysis is guided by a set of hypotheses drawn from prior work on positional concerns and social comparison, which we outline in the next section (Section 2.2.2). These include both preregistered[1] and secondary hypotheses focusing on how visibility, reference groups, and societal inequality affect positional preferences. We test these hypotheses using a randomized survey experiment designed to elicit trade-offs between absolute and relative outcomes across economic (e.g., income, vacation time) and health-related domains (e.g., life expectancy, healthcare access). In addition, we pay particular attention to how individual choices respond to our manipulations, and further examine respondents' underlying preferences.

This chapter complements Chapter 1, which reviewed how social norms and peer effects influence health-related behaviors among young people. While Chapter 1 emphasized conformity and identity considerations as important drivers of health behaviors, this chapter turns to positional concerns, where individuals care explicitly about how their outcomes compare against others rather than simply conforming to what others do to fit in. By integrating this mechanism, this dissertation helps advance our understanding of how relative comparisons shape these preferences and influence decisions in both social and institutional contexts.

To investigate these mechanisms, we ran a cross-national survey experiment in the UK and US, building on the survey design by Solnick and Hemenway (1998). Respondents chose between two hypothetical societies that differed in absolute and relative outcomes across several domains. Stating a preference for a relatively better outcome over an absolutely better one (e.g., choosing to be richer relative to others in a poorer society) is interpreted as evidence of positional concerns. In the present study, we made different reference groups salient (the general population or friends and acquaintances) and looked at both economic and health-related choices, varying how visible and measurable the outcomes were.

The sample was drawn from the online panel provider Prolific and was gender-balanced[2]. To complement the main survey, we implemented (i) a slider-based measure of marginal positionality, (ii) an incentivized Krupka-Weber task to elicit second-order normative beliefs,

[1]Hypotheses H1–H3 were preregistered prior to data collection (Müller *et al.*, 2024); Hypotheses H4 and H5 were developed subsequently and are interpreted as exploratory.

[2]Note that they are not nationally representative, especially in terms of education levels. Approximately 58% of study participants reported that they had completed at least a university bachelor's degree, which is higher than the national averages in both the US (50.0%) and the UK (51.3%) (Statista, 2025).

and (iii) a stated-satisfaction approach, adapted from Diaz *et al.* (2023), to classify respondents' social preference types.

Several findings emerge from our survey experiment. First, positional concerns are more pronounced for visible goods, particularly those in the health domain, such as physical activity. This supports prior work suggesting that visibility enhances status-signaling potential (Alpizar, Carlsson and Johansson-Stenman, 2005; Hillesheim and Mechtel, 2013), although not all studies have found this relationship (Leites, Rivero and Salas, 2024). Our results strengthen the view that perceived observability plays a role in the formation of positional preferences. Second, reference groups matter. Study participants display stronger positionality when comparisons are framed with respect to the general population than to a closer reference group, such as friends or acquaintances. Contrary to findings from previous literature that closer social ties are the central reference group for social comparison (Clark and Senik, 2010), participants were more positional when comparing themselves to the general population than to close peers, suggesting instead that general societal comparisons may be more salient in certain domains, particularly health. Third, cross-country differences lend some support to the hypothesis that perceived inequality may heighten positional concerns. A comparison of identical hypothetical choices in the US and UK samples demonstrates that those living in the US – and thus, in a more unequal health system – displayed stronger positional preferences than UK respondents, particularly with regard to healthcare access. While country of residence is only a proxy for exposure to inequality, this result tentatively aligns with the prediction that greater societal disparities activate greater status sensitivity (Schneider *et al.*, 2021; Bowles and Park, 2005; Velandia-Morales, Rodríguez-Bailón and Martínez, 2022). Finally, individual-level preferences such as envy and inequality aversion were robust predictors of positional choice across and within the studied domains. These results suggest that positional concerns are not merely context-driven, but also affected by certain social preference types.

These findings have implications for the design of health interventions and redistribution policies. If individuals care not only about being healthy but about being *healthier* than others, then public health messaging, incentive design, and resource allocation may need to account for the effects of status motives and how social comparison drives behavior. Incorporating positional motives into policy planning offers a promising avenue, as it acknowledges that

individuals' preferences for status and relative positioning can drive the overconsumption of conspicuous services and contribute to resource misallocation, ultimately affecting societal welfare. If any of the aforementioned drivers prove significant, new policy strategies could be devised to better address these inefficiencies, and redirect attention toward basic, yet high-impact interventions. Such strategies could enhance economic efficiency, promote equity, and improve overall public health outcomes. To guide the empirical analysis, we formulate a set of preregistered and exploratory hypotheses in the next section.

The remainder of this chapter is organized as follows. In section 2.2 we first introduce our primary and secondary hypotheses, modeling framework, and sample, and finally present the five experimental blocks along with the reference-group manipulation. Section 2.3 presents the empirical findings. Section 2.4 discusses implications for policy and future research.

2.2 Methods and study design

2.2.1 Preregistered hypotheses (H1–H3)

Hypotheses **H1** to **H3** were formulated prior to data collection as part of the study's preregistration. The remaining hypotheses, Hypotheses **H4** and **H5**, were developed subsequently and are therefore interpreted as exploratory.

Cross-country variation and inequality. Naturally, positional concerns emerge in contexts where individuals observe disparities and compare their own position with that of others. Hillesheim and Mechtel (2013) show that people care even more about their relative position when it is linked to competitive disadvantages for others (e.g., in terms of intelligence or education). It thus seems plausible that living in a more unequal society may intensify concerns about one's position – especially in environments where the cost of falling behind is greater (Frank, 2008). Social norms and contexts further shape these relative concerns, particularly in societies characterized by greater economic disparities (Bowles and Park, 2005; Charles, Hurst and Roussanov, 2009), which are further exacerbated by social comparison processes (Velandia-Morales, Rodríguez-Bailón and Martínez, 2022). Whether this also translates to health and healthcare remains an open question, given that health is

often understood as a basic right rather than a positional good. However, a recent choice experiment conducted in the US and UK, two countries chosen for their marked differences in healthcare infrastructure and equity, reports that individuals not only care about their own and others' waiting times but also about their relative position in the queue. Our first hypothesis is therefore:

H1) Positional concerns are more pronounced in societies characterized by higher social, economic, and health inequalities, with US respondents expected to show stronger preferences for relative advantage than those in the UK.

This hypothesis also builds on methodological insights from Alvarez-Cuadrado and Long (2012), whose study highlights how social comparisons and status seeking contribute to economic disparities. Our study tests the reverse relationship: what motivates concerns about one's relative standing.

Reference groups. Prior research has shown that individuals evaluate their own well-being by comparing it to the health of others, leading to a "keeping up with the Joneses" effect even in health-related domains (Blanchflower, Van Landeghem and Oswald, 2009; Carrieri, 2012; Mujcic and Frijters, 2015). Positional concerns thus often arise in social contexts where individuals continually assess their standing relative to others (Gugushvili, 2021). These comparisons typically depend on the perceived proximity and relevance of the reference group, so that individuals are more likely to compare themselves to close others, like friends or colleagues, than to more general or distal populations (Clark and Senik, 2010). This leads to our second hypothesis:

H2) Positional concerns are stronger when comparisons are framed with respect to friends and acquaintances than to the general population.

Social preferences. Prior research demonstrates that individuals more often than not deviate from purely rational and self-interested behavior. Positional and status-seeking motives are one explanation for such inconsistencies; another lies in social and other-regarding preferences. These include altruism, envy, inequality aversion, and reciprocity, and have been

found to be heterogeneous across individuals, yet remain relatively stable across experimental games (Cooper and Kagel, 2016). Similarly, we therefore expect:

H3) Positional concerns are associated with social preferences, including altruism, envy, and inequality aversion.

To do so, we first employ different sets of preference measures (including economic, social, and distributional preferences) using established instruments and validated scales (see Appendix A2.2 for a full description of variables). Among these is the *stated satisfaction* method adapted from Diaz *et al.* (2023), in which participants rate their satisfaction with various payoff allocations. We then apply clustering techniques following Fallucchi, Luccasen III and Turocy (2019, 2022) to group individuals into behavioral types based on the structure of their satisfaction responses. While the framework provided by Diaz *et al.* (2023) classifies individuals using fixed cutoff points, the clustering approach is flexible enough to handle even those observations that may fall into multiple categories. The clustering method helps us assign individuals to latent profiles emerging from the data, and thus identify coherent responses that may not align with pre-specified thresholds.

2.2.2 Secondary hypotheses (H4–H5)

Differences across domains. We investigate positional concerns across both economic and health-related contexts. Prior experimental evidence suggests that individuals are generally less inclined to make relative trade-offs in health domains than in economic ones (Celse and Grolleau, 2021; Wouters *et al.*, 2015), however methodological differences across studies make these findings difficult to compare directly. We test the robustness of these findings with the following hypothesis:

H4) Positional concerns vary across domains and are less prevalent in health-related outcomes than in economic ones.

Visibility of goods. Lastly, earlier work has often pointed to the visibility and salience of goods as intuitive explanations for why positional concerns vary across domains (Alpizar,

Carlsson and Johansson-Stenman, 2005; Celse and Grolleau, 2021; Charles, Hurst and Roussanov, 2009; Frank, 2013; Heffetz and Frank, 2011; Leites, Rivero and Salas, 2024), though this view has not gone unchallenged in the literature (e.g., Hillesheim and Mechtel, 2013). We hypothesize that in the context of health and healthcare, goods that are more visible (e.g., physical fitness or the quality and availability of healthcare) may serve as more salient markers of relative standing than less publicly observable attributes like life expectancy or child health. Thus:

H5) Positional concerns are more pronounced for visible goods (e.g., income, physical activity, or healthcare access) than for less visible goods (e.g., life expectancy and child health).

2.2.3 Measuring positional concerns

In the literature on positional preferences, relatively little attention has been paid to the rationale for selecting specific values in survey instruments. We are only aware of one paper that explains their choice of value selection for their study on relative comparisons (Mageli, Mannberg and Heen, 2022). A common issue in previous studies on positionality is that survey questions often present participants with options and values that are either unrealistically high or low, which might give rise to validity concerns, and thus make scenarios difficult to evaluate plausibly. Another issue arises from the lack of consistency in survey question formats and value ranges across studies, undermining the direct comparability of results not only between studies investigating the same domain, but also within studies examining multiple domains. To address these challenges, our study adopts a systematic approach in determining the values for survey items, guided by two primary criteria. First, the values are calibrated to fall within plausible and empirically grounded ranges. For instance, in economic domains, income values are based on the average post-tax monthly earnings (OECD, 2020). In the health domain, indicators of health and healthcare outcomes were informed, among others, by official statistics, the OECD, and the WHO[3]. Second, since individuals with positional concerns care about how they compare to others, we assume that this preference structure is reflected in individuals' utility functions, which incorporate both the absolute and relative

[3]Refer to Appendix A2.2 for an overview of the empirical data used to design the hypothetical choice scenarios.

value of the respective goods (Alpizar, Carlsson and Johansson-Stenman, 2005; Duesenberry, 1949).

Positional concerns can enter the utility function through various means. While many studies employ an additive comparison utility formulation, another approach involves expressing these concerns using a ratio comparison utility function, which has been shown to perform slightly better than additive alternatives in explaining economic behavior (Johansson-Stenman, Carlsson and Daruvala, 2002). Accordingly, we specify individual utility as $U_i = v_i(x_i^b, \frac{x_i^b}{\bar{x}^b})$, where x_i^b is the individual i's consumption level of good b and $\bar{x}^b$ is the average consumption level of b in the given reference group (Alpizar, Carlsson and Johansson-Stenman, 2005; Duesenberry, 1949). Finally, to measure positional concerns, we follow Alpizar, Carlsson and Johansson-Stenman (2005) and Leites, Rivero and Salas (2024): $v_{i,s} = (x_{i,s}^b)^{(1-\gamma_i^b)} \cdot (\frac{x_{i,s}^b}{\bar{x}_s^b})^{\gamma_i^b}$, where $x_{i,s}^b$ denotes individual i's consumption level of b in society s, $\bar{x}_s^b$ is the reference group's average consumption of b in society s and γ_i^b corresponds to individual i's degree of positionality concerning good b. It follows that, if $\gamma_i^b = 0$, individual i only cares about their own consumption of b. If $\gamma_i^b > 0$, i's utility depends on both their own consumption levels as well as on how their consumption compares to that of other people.

To illustrate this concept, consider two identical hypothetical societies A and B that only differ in terms of absolute and relative consumption levels of good b. An individual is indifferent between these two societies if $v_{i,A} = v_{i,B}$. Based on the above utility specification, the degree of positionality γ can be determined from Equation 2.1:

$$\hat{\gamma}_i^b = \frac{\log\left(\frac{x_{i,B}^b}{x_{i,A}^b}\right)}{\log\left(\frac{\bar{x}_B^b}{\bar{x}_A^b}\right)} \tag{2.1}$$

While the degree of positionality can, in theory, be computed by adjusting the values of $x_{i,s}^b$ in alternative societies A and B, its practical application is often limited by the difficulty of eliciting consistent and realistic consumption scenarios. To date, previous studies have addressed this issue following two distinct approaches: some studies, for example Alpizar, Carlsson and Johansson-Stenman (2005) and Carlsson, Johansson-Stenman and Martinsson (2007), have utilized varying parameter values for the degree of positionality,

where $0 \leq \gamma_i^b \leq 1$. Although this framework offers the advantage of allowing comparisons of respondents' positional concerns across various goods, the majority of studies adopt different lower bounds for γ within their respective domains (Celse, 2012; Celse and Grolleau, 2021; Solnick and Hemenway, 2005; Wouters *et al.*, 2015), thereby challenging comparability both within and across studies. One exception is Mageli, Mannberg and Heen (2022), which consistently applies a constant parameter value for γ across all domains, facilitating more meaningful and consistent cross-domain analyses of positionality. In the present study, we follow this practice and set $\gamma = 0.25$ uniformly across all domains. Importantly, $\gamma = 0.25$ serves as a design parameter to construct trade-offs, not as a direct empirical estimate of participants' positionality.

We hold γ constant and thus standardize the benchmark for positional trade-offs. This assumption makes responses across domains directly comparable and avoids the inconsistencies observed in previous studies, where γ often varies by domain[4]. In the present study, the value $\gamma = 0.25$ is informed by empirical estimates suggesting that the average positionality for variables such as income, vacation time, and life expectancy falls between 0.20 and 0.37 (Carlsson, Johansson-Stenman and Martinsson, 2007; Celse and Grolleau, 2021; Solnick and Hemenway, 2005; Wouters *et al.*, 2015). Although values up to 0.8 have been reported in certain domains (Aronsson and Johansson-Stenman, 2013), the lower value is more representative of those estimates reported in prior work. Additionally, it makes it easier to observe differences in positional behavior across domains, especially in areas like health, where status concerns are assumed to be weaker. Accordingly, we used $\gamma = 0.25$ to calibrate all choice scenarios in our survey experiment, constructing the trade-offs so that an individual with moderate positional concerns (specifically, $\gamma = 0.25$) would be indifferent between the two hypothetical options. This further implies that the share of participants choosing the relative option in each domain reflects the strength of positional concerns relative to this benchmark. In this sense, the proportion of participants who choose the relatively better option is interpreted as an indicator of the strength of positional concerns in the given domain. For example, if 8% of

[4]While a constant γ improves the interpretability of our results and allows for relative comparisons across domains (for instance, "participants exhibit stronger positional concerns for healthcare access than for waiting time"), these conclusions are mainly based on differences with regard to positional choice proportions. They do not represent direct estimates of the underlying degree of positionality in each domain. Thus, our results should be interpreted as deviations from the common benchmark, not as estimates of participants' *true* concern with relative standing.

study participants opt for worse air quality locally but better relative standing compared to other areas, it suggests that most participants derive less than one-quarter of their utility v_i from consumption relative to others, and, in other words, that the vast majority prioritizes fewer polluted days where they live over their position relative to others.

2.2.4 Sample and preregistration details

In February 2024, we preregistered our hypotheses, primary outcomes, and analysis plan in the Open Science Framework (OSF) registry (`https://doi.org/10.17605/OSF.IO/MJZKV`). Ethical approval for the study was granted by the committee for research integrity of the University of Bergamo [2024_06_05].

We recruited a total of 1,003 participants from the United States ($N = 502$) and the United Kingdom ($N = 501$) via the online survey platform Prolific. Eligible participants were aged between 18 and 55 years and balanced by gender. Prior to data collection, we conducted a power analysis using G*Power (Faul *et al.*, 2009). Based on a two-tailed test at the 5% significance level, we determined that a total sample size of 694 would ensure 80% power to detect small differences in proportions between the two countries (US vs. UK). Within each country, participants were then randomly assigned to one of two experimental reference group conditions (general population vs. friends and acquaintances). Additional power calculations confirmed that with about 250 participants per condition per country, our design was also well suited to detect comparable treatment effects as those observed in previous studies (e.g., Mageli, Mannberg and Heen, 2022; Wouters *et al.*, 2015).

As preregistered, two attention checks were embedded in the survey, following standard practice to ensure data quality. Participants who failed either of these checks were excluded from the main analysis, resulting in a final sample of $N = 981$ respondents after data cleaning. All participants received a flat participation fee of £1.50, with the opportunity to earn a bonus of £0.50 through an incentivized second-order belief elicitation task (Krupka and Weber, 2013). The median completion time for the survey was approximately 8 minutes.

Descriptive statistics for the final sample are presented in Appendix A2.1. Table A2.1 summarizes the distribution of demographic characteristics, mean survey responses, and

the proportion of participants selecting the positional (relative) option across experimental conditions. If participants indicated "I don't know" or "Prefer not to answer", responses were coded as not available and excluded from regressions.

2.2.5 Experimental design

We conducted an experimental online survey to assess positional concerns across economic and health-related domains in the United States and the United Kingdom – two countries that differ significantly in income and health inequalities, especially with regard to healthcare infrastructure and access (Schneider *et al.*, 2021). The experiment was divided into five blocks, which all participants completed in sequential order.

Block 1: Positional choice tasks. For the main study design, we build on Solnick and Hemenway's empirical approach. Following their methodology, we presented study participants with a randomized set of questions covering a broad range of goods and categories that vary in properties such as visibility, tangibility, and necessity, to compare positional concerns. For each question, participants were asked to indicate which of the two presented hypothetical choice scenarios they preferred, each illustrating a distinct society and outcome. The two societies differed in only one domain-specific feature intended to elicit positional concerns, while we emphasized that they were identical in all other respects, including prices and the variety of goods. Participants were instructed to select the society in which they would feel most content, even if neither option was strongly preferred, and were explicitly asked not to choose based on what they believed was best for society as a whole.

An example is provided below:
Which society would you prefer?

- *Society A:* Your life expectancy at birth is 84 years.
  Others' life expectancy at birth is 87 years.
- *Society B:* Your life expectancy at birth is 79 years.
  Others' life expectancy at birth is 68 years.

The societies are the same in all other matters.

These, as well as the values used in the other scenarios, were informed by official statistics from

the OECD and WHO, and in part by other studies. In Appendix Table A2.2 (Section A2.2) we provide an overview of the full set of reference values used to design the parameters.

Additionally, questions pertaining to the measurement of economic preferences (risk and time preference, altruism) (Falk *et al.*, 2023), social comparison (Gibbons and Buunk, 1999), and social image concerns (Alba *et al.*, 2014) were randomly[5] interspersed among the positional choice items.

Treatment: Reference group framing. To examine the role of reference groups in positional comparisons, that is, the people individuals compare themselves to, participants were randomly assigned to one of two conditions: One half of the sample received questions framed relative to the general population, while the remaining half were presented with questions framed relative to friends and acquaintances. This manipulation, as preregistered, allows us to examine whether positional preferences are more pronounced when comparisons involve socially closer peers, as hypothesized in Hypothesis **H2** in Section 2.1.

Block 2: Indifference thresholds via slider tasks. Block 2 was shown only to a subset of participants. After completing the main choice tasks, participants who had previously indicated a preference for the positional choice in either the income or life expectancy domain proceeded to a follow-up slider task: respondents who had chosen the positional option for income in Block 1 were presented with a continuous slider to indicate the minimum acceptable income at which they would be willing to accept Society A (where they would be better off in absolute terms but relatively worse off compared to the given reference group) over Society B (where they would be worse off in absolute terms but relatively better off). Similarly, those who indicated positional preferences for life expectancy in Block 1 were provided with a range of life expectancies (in years) to indicate the threshold at which they would accept the trade-off and switch to Society A.

Note that, conceptually, these responses also enable a more precise estimation of individuals' degree of positionality $\hat{\gamma}_i^b$, by applying Equation 2.1 from the preceding Section 2.2.3, given that the observed indifference thresholds provide the required inputs for its computation.

[5]We chose to intersperse and randomize these items from different questionnaire modules to avoid order effects and reduce participant fatigue, which we deemed likely if a long battery of positional questions were presented consecutively.

Methodologically, the sliders provide a useful tool for directly estimating $\hat{\gamma}_i^b$, the individual degree of positionality. Although we do not compute the individual estimates in the main analysis due to limited sample sizes (e.g., $N = 36$ for the life expectancy slider), the indifference points elicited through the tasks provide the necessary data for doing so.

Block 3: Incentivized belief elicitation. In the third part of the study, all participants completed an incentivized belief-elicitation task. They were presented with three hypothetical choice scenarios similar to those they had read earlier, and were asked to predict which option the majority of other study participants had preferred. Participants were instructed to choose the option they believed most others would feel most content with, and not the one they considered best for society as a whole. This task, based on Krupka and Weber (2013), was incentivized: participants were informed that one of the three scenarios would be randomly selected, and if their response matched the majority choice for that scenario, they would receive a £0.50 bonus in addition to their base payment.

Block 4: Stated satisfaction and social preferences. The fourth part of the experiment asked participants to evaluate how satisfied they would feel in a series of income allocations, implementing a reduced version of the stated satisfaction approach introduced by Diaz *et al.* (2023). Participants rated their satisfaction with monetary allocations in which their own payment was fixed and another (unspecified) person's payment varied. This task was identical across countries, differing only in currency formatting. Based on their satisfaction responses, individuals were classified into nine stylized social preference profiles, reflecting sensitivity to inequality, fairness, or efficiency. Principally, these types reflect how participants react to advantageous and disadvantageous inequality, and allow us to examine whether individuals with particular distributive preferences are more likely to exhibit positional concerns.

Modeling social preference types through clustering methods. To classify participants into the aforementioned social preference types, we applied clustering techniques to their full pattern of satisfaction ratings. Rather than relying on single responses or strict thresholds, we grouped individuals based on the overall structure of their satisfaction statements. This allowed us to identify mixed or ambiguous cases that do not fit neatly into one category. Our method builds on the clustering framework introduced by Fallucchi, Luccasen III and Turocy

(2019, 2022), which we adapt to a stated satisfaction context. To do so, we used K-means clustering to identify the common motivational patterns across individuals. The resulting types, their characteristic satisfaction profiles, and differences are detailed in Appendix A2.2.2.

Importantly, this clustering approach goes beyond simple linear associations and allows us to test Hypothesis **H3** at the type level: by linking participants' stated social preferences to their revealed positional concerns from earlier blocks, we can assess whether certain types, such as those exhibiting inequality-averse or envious tendencies, are more likely to choose the positional option over absolute outcomes. Thus, we can investigate whether positional preferences are influenced by more than just contextual factors (e.g., domain or reference group) but also by people's underlying social motivations.

Block 5: Demographics. Finally, the last block concluded with a set of demographic questions including gender, age, education, employment status, annual household income and financial comfort, political orientation, and self-rated health status. These variables are used in our analysis as descriptive covariates and controls.

Together, these blocks provide a comprehensive framework for testing our main hypotheses (H1–H3), enabling us to relate behavioral choices, belief elicitation, and stated preferences to both contextual factors and individual-level determinants of positional concerns.

2.3 Results

2.3.1 Heterogeneity across and within domains

2.3.1.1 Positional preferences in economic and health domains

Figure 2.1 presents the proportion of positional responses across all surveyed domains, including the second-order belief items, which refer to participants' perceptions of what most others would choose[6]. We plotted second-order beliefs alongside their own preferences to allow

[6]Note that the second-order belief items capture what participants *believe* most others would choose in these hypothetical scenarios, rather than their own preferences. The survey item stated "Do you think the majority of participants preferred Society A or Society B?", followed by the same hypothetical scenarios they had previously expressed their preferences for. These questions were incentivized, and participants could earn a bonus if their prediction matched the majority choice. We report these responses alongside own preferences

for a direct comparison between individual concerns and perceived norms surrounding the trade-offs in each domain, illustrating whether participants systematically overestimate or underestimate the positional preferences of others. The bars reflect average values pooled across both treatment conditions. The distribution highlights the salience of income as the most frequently cited positional concern (29%), followed by physical activity (21%), vacation time (15%), and hospital beds per capita (15%). Days of illness (10%) and doctor density (10%) are next, then good child health (7%), air quality (6%), cancer screening wait time (5%), and life expectancy (4%). Second-order beliefs fall below own preferences only in the income domain. For example, 25% of respondents believed most others would select the positional income option (compared to 29% selecting the positional outcome for themselves), and 9% believed most would choose the positional option with regard to life expectancy (compared to 4% own). Chi-squared tests confirm that these differences between own choices and beliefs about others are highly statistically significant across income, life expectancy and waiting time (all $p < 0.001$), and that these differences persist when stratified by country and by reference group.

2.3.1.2 Comparison with previous research

While previous literature in economics has investigated the extent of positional concerns across a wide range of goods and domains, empirical work on how these preferences manifest in health and healthcare represents a comparatively newer strand. One of the principal aims of this study is therefore to examine whether health-related goods that are typically presumed to be less positional evoke similar status sensitivity as economic goods.

Table 2.1 presents our findings alongside estimates from selected prior studies, enabling a direct comparison of the prevalence of positional choice across both economic and health domains. While perfect comparisons are not feasible due to differences in conceptualization and implementation of study designs, we restrict this comparison to participants in the *general population* treatment group to ensure consistency and to align as closely as possible with earlier experimental designs. In brief, our results suggest that previous studies may have overstated the extent of positionality, particularly in health-related contexts.

across domains to highlight how perceived social norms align with (or diverge from) individuals' actual choices.

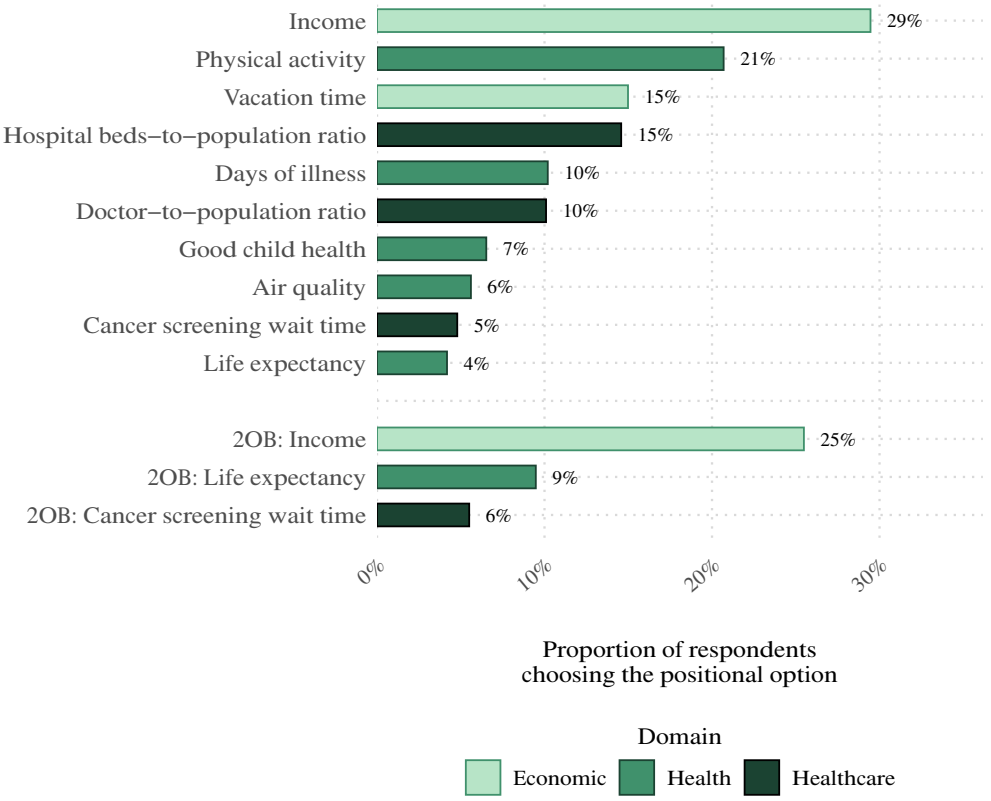


Figure 2.1: Proportion of positional choices across domains, including second-order beliefs (2OB)

Table 2.1: Comparison of positional choices across selected studies

	This Study[b]	Solnick and Hemenway (2005)	Carlsson et al. (2007)	Wouters et al. (2015)	Celse and Grolleau (2021)	Min Δ, Max Δ[c]
Economic domain						
Income	0.41	0.41	0.75	0.30	–	−0.34, 0.00
Vacation time	0.18	0.16	–	0.24	–	−0.06, 0.02
Health domain						
Life expectancy	0.05	0.11	–	0.21	0.04	−0.16, 0.01
Air quality	0.08	0.11	–	–	–	−0.03, –
Days of illness	0.13	0.11	–	–	–	0.02, –
Cancer screening wait time[a]	0.07	–	–	0.16–0.18	–	−0.11, −0.09

Notes:
[a] We focus on positional concerns regarding cancer screening wait times; Wouters *et al.* (2015) examine wait times for other medical procedures, including knee, cataract, and open-heart operations.
[b] Estimates from this study are based on participants assigned to the *general population* treatment group only.
[c] Min Δ, Max Δ denote the range of differences between this study and prior studies (This study – Others), illustrating where our estimates fall relative to earlier findings.

For income, reported positional choice proportions in earlier studies range from approximately 30% to 75% (Carlsson, Johansson-Stenman and Martinsson, 2007; Solnick and Hemenway, 2005; Wouters *et al.*, 2015). Our estimate of 41% lies toward the lower–middle end of this distribution. In contrast, the prevalence of positional responses for health-related variables in our sample consistently falls at the lower bound of existing ranges. For instance, Solnick and Hemenway (2005) find that 11% of respondents preferred more days of illness and 11% also preferred worse air quality – as long as they remained better off than others. Similarly, estimates of positionality with respect to life expectancy vary from 4% (Celse and Grolleau, 2021) to 21% (Wouters *et al.*, 2015), whereas our estimate (5%) closely matches the lower end of that spectrum.

A plausible explanation for these discrepancies concerns the methodological treatment of trade-off strength across domains. In contrast to previous studies that allow the degree of positionality γ, that is the weight on relative standing in the utility function, to vary implicitly by context, we fix γ at 0.25 in all choice scenarios. This calibration is informed by prior empirical work suggesting an average degree of positionality between 0.20 and 0.37 (Carlsson, Johansson-Stenman and Martinsson, 2007; Celse and Grolleau, 2021; Solnick and Hemenway, 2005; Wouters *et al.*, 2015). While this approach facilitates comparability across goods, it may attenuate positional responses in domains where respondents would otherwise require steeper trade-offs.

Consistent with theoretical work by Aronsson, Ghosh and Wendner (2023), our results support the view that positional concerns are more pronounced for goods that are socially visible or salient. As many of the health-related dimensions studied here involve outcomes that are relatively private or less subject to peer comparison, the low prevalence of positional choice appears consistent with the literature on the social visibility of consumption and status.

2.3.1.3 Sensitivity analysis: Trade-off decisions under varying degrees of positionality

In the second block of the experiment, participants who had expressed positional concerns regarding income or life expectancy were presented with a slider task designed to elicit their marginal degree of positionality. In this task, they were asked to indicate the point at which

they would be willing to forgo a relatively better status (e.g., earning $3700 while others earn $3200) in favor of a higher absolute outcome (e.g., earning $4100 while others earn $4900)[7]. Subsequently, in the income domain, participants were asked to identify the minimum income level at which they would switch to the non-positional option, and thereby accept a relatively worse standing compared to their treatment-specific reference group.

We began by comparing participants with and without positional concerns across their preferences in other domains (e.g., vacation time, health-related outcomes) and background characteristics (e.g., age, gender). To account for potential non-normality in the data, we employed Mann-Whitney U tests, which are appropriate for comparing independent groups under such conditions. Bonferroni corrections were applied to control for multiple testing. Perhaps not surprisingly, those expressing positional concerns about income were also more likely to do so in several other domains, including vacation time, air quality, days of illness, good child health, physical activity, and waiting time. No significant differences were observed for preferences related to life expectancy or healthcare provision.

We found no statistically significant differences in age, gender, education, or income between respondents with and without positional concerns. Where the groups did differ, though, was in their social preference types, following the *stated satisfaction* approach (Diaz *et al.*, 2023). At the median, the non-positional group showed no signs of envy, while the positional group did. Conversely, satisfaction responses among non-positional individuals were indicative of self-interest, suggesting that their choices were also driven by own payoff considerations. The positional group displayed no such tendency, consistent with behavioral models on inequality aversion (Fehr and Schmidt, 1999) and the view that individuals are motivated not only by absolute outcomes but also by where they stand relative to others (Bolton and Ockenfels, 2000). Consistent with this, positional respondents reported greater dissatisfaction when others earned more, indicating a broader sensitivity to inequality. We return to these points in more detail in Section 2.3.3. Unlike the fixed $\gamma = 0.25$ benchmark that we used to construct the hypothetical scenarios in the main block, these estimates represent individualized switching points, and provide a revealed measure of respondents' positional sensitivity. Figure 2.2 illustrates the distribution of marginal degrees of positionality for income across the UK and

[7]UK respondents received the equivalent framing in £. In Society A, they earned £2400 while others earned £3000; in Society B, they earned £2100 while others earned £1800.

US samples. Participants' slider responses were translated into implied γ values, ranging from 0.09 to 0.63 (see Appendix A2.3).

These values are plotted on the x-axis as individual switching points, while the y-axis indicates the proportion of respondents falling into each bracket. The dashed vertical line denotes the baseline value of $\gamma = 0.25$ employed in the main choice tasks[8].

UK participants exhibited slightly higher marginal degrees of positionality ($\overline{\gamma} = 0.51$) than their US counterparts ($\overline{\gamma} = 0.45$), suggesting comparatively greater sensitivity to relative income. When pooling all positional respondents together, the average positionality was $\overline{\gamma} = 0.48$, which is substantially above the baseline threshold of 0.25. In monetary terms, this can be interpreted as the average willingness to forgo their relatively better position compared to their reference group, meaning UK participants would require approximately £333.80 more per month, and US participants $405.60 more, to accept being relatively worse off. In addition to these mean differences, a Kolmogorov–Smirnov (K-S) test showed a highly significant ($p < 0.001$) difference between the UK and US distributions, indicating that not only are average levels of positionality different, but also how positional concerns are spread across individuals in each country.

We administered a similar slider task to the subset of participants who reported positional concerns with regard to life expectancy. In this version, individuals faced trade-offs between a higher relative life expectancy (e.g., living to 79 while others live to 68) versus a better absolute one (e.g., living to 84 while others live to 87). Only 36 respondents progressed from the main block to this task, substantially limiting the scope for meaningful statistical analysis. However, although underpowered, the presence of some positional responses tentatively suggests that certain individuals do apply relative considerations even in the context of life expectancy, albeit less pronounced compared to other domains. We return to their discussion in Section 2.4.

Finally, we assessed whether the distribution of positionality differed by treatment condition. Individuals in our baseline scenario who exhibited positional concerns toward income were

[8]Note that these switching points can be interpreted in multiple, yet complementary ways. In this chapter, we mainly follow a model-consistent interpretation and treat them as marginal degrees of positionality to ensure comparability across domains and with previous studies. In the discussion (Section 2.4), we revisit and elaborate on these interpretations.

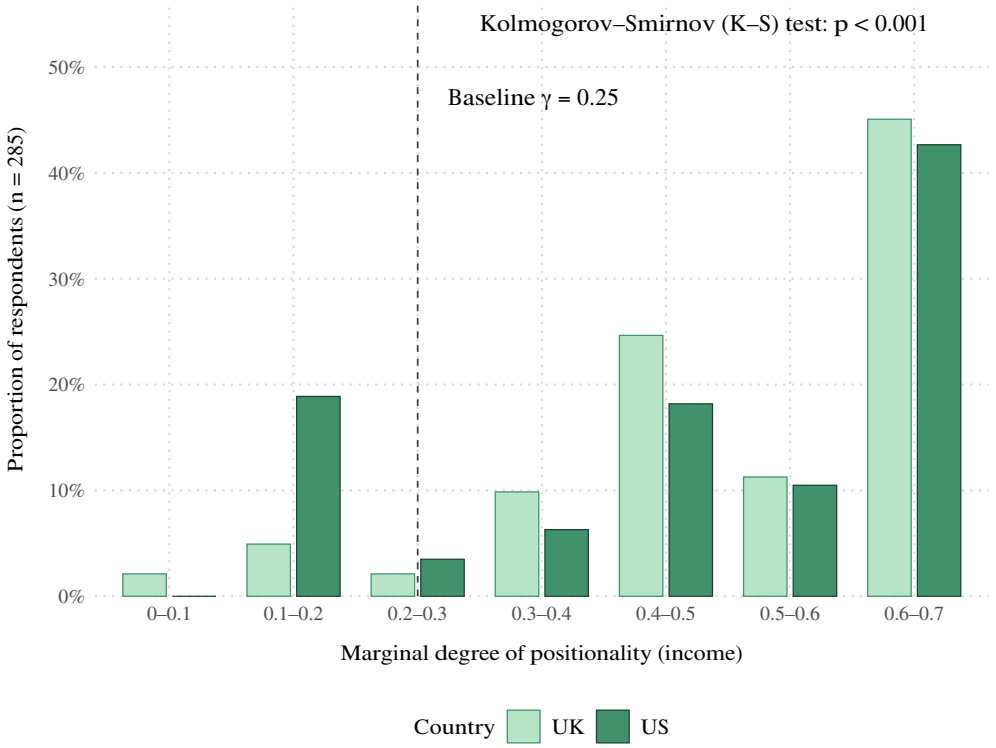


Figure 2.2: Distribution of marginal degree of positionality γ for income ($n = 285$), shown as bin proportions.

more likely to originate from the *general population* treatment than from the *friends and acquaintances* treatment. To account for differences in group sizes, we conducted Mann-Whitney U tests to determine if there were any differences between the treatments in terms of trade-off decisions and responses to the slider task. The test outcomes revealed no statistically significant differences in the positionality distributions of γ between the two groups, suggesting that the median values of positionality are not significantly different at the 5% significance level.

Overall, these findings point to the idea that participants' willingness to accept trade-offs varies across domains, with income eliciting comparatively stronger positional responses. While this interpretation is based on a limited comparison between only two domains (income and life expectancy), the observed differences are consistent with the notion that economic outcomes provoke greater status sensitivity. The observations also lend some first, tentative support to Hypothesis **H3**, as we find significant differences in social preference types between positional and non-positional respondents, with the former group exhibiting higher levels of envy and

lower self-interest. Although we cannot claim causality, these results indicate an underlying association between positional concerns and certain preference types or motivations.

2.3.2 Contextual influences on positional concerns

2.3.2.1 Do friends and acquaintances heighten positional concerns?

We conducted chi-squared tests to test for reference-group effects on positional preferences, and compare the proportion of positional choices across the two framing conditions: Do individuals display greater positional concerns across domains when they make their choices with respect to friends and acquaintances than when they choose referring to the general population? Our results are shown in Table 2.2, based on the full pooled sample[9].

Table 2.2: Differences in proportions of positional choices by reference group (Full pooled sample)

	Friends and acquaintances	General population	Diff.	p-value
Economic domains				
Income	0.183	0.408	-0.225	0.000***
Vacation time	0.120	0.180	-0.061	0.010*
Health domains				
Life expectancy	0.039	0.045	-0.007	0.725
Air quality	0.032	0.080	-0.047	0.002**
Days of illness	0.071	0.133	-0.062	0.002**
Good child health	0.057	0.074	-0.017	0.344
Physical activity	0.197	0.217	-0.020	0.476
Cancer screening wait time	0.026	0.070	-0.043	0.002**
Doctor-to-population ratio	0.073	0.129	-0.056	0.005**
Hospital beds-to-population ratio	0.132	0.160	-0.028	0.249
Second-order beliefs				
Income	0.203	0.307	-0.105	0.000***
Life expectancy	0.079	0.111	-0.032	0.115
Cancer screening wait time	0.037	0.074	-0.037	0.016*

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Diff. = Difference in proportions, i.e., share of respondents selecting the positional option in each framing condition. All p-values are based on two-sided chi-squared tests.

As shown in Table 2.2, participants expressed greater positional concerns when comparing

[9]Country-specific results are presented in Table A2.8 in the Appendix.

themselves to the general population than to closer social ties. Across the full sample, the proportion tests reveal statistically significant differences ($p < 0.05$) between treatments for income, vacation time, air quality, days of illness, waiting time, and the doctor-to-population ratio, as well as for second-order beliefs concerning income and waiting time. In the income domain, participants assigned to the *general population* condition were not only more likely to choose the positional option themselves, but also more likely to believe that most others would do the same, although the proportion of participants who believed that most others would choose the positional option was around 10 percentage points lower than the share who actually did so.

Interestingly, this tendency to overestimate positionality extended to life expectancy, though the pattern was more pronounced in the US: there, participants estimated that others would prioritize relative standing nearly twice as often as they actually did. In the UK, however, estimates were nearly identical across both treatment groups (Table A2.8). In the *friends and acquaintances* condition, the estimated majority was nearly twice the actual rate, and in the *general population* condition, the estimate was approximately 2.5 times higher. Finally, about 20% of participants expressed positional concerns in the domain of physical activity, with no significant difference between the two reference group conditions.

These findings contradict our initial prediction that individuals would display stronger positional concerns when comparing themselves to friends and acquaintances: rather than increasing positional concerns, comparisons with closer social ties were linked to lower levels of positional choice in most domains. One explanation could be that individuals are less motivated by competition and relative standing in close social contexts, where relationships may be characterized by more familiarity or cooperation than by competition and status-seeking behaviors (Vandegrift and Duke, 2015). Alternatively, altruistic motives may play a role: respondents might derive additional utility from knowing that their friends or acquaintances are also doing well. We explore the potential mechanisms driving positional choices in the remainder of the results section, specifically in Section 2.3.3.

2.3.2.2 Does inequality heighten positional concerns?

To test whether individuals living in societies characterized by higher social, economic, and health inequalities display greater positional concerns, we administered our survey experiment to two samples: 502 US and 501 UK residents. As a first step, we use non-parametric proportion tests to compare the distribution of positional responses across the two countries. For all but one variable (doctor-to-population ratio) the chi-squared tests revealed no statistically significant differences in the proportion of positional choices between US and UK respondents ($p = 0.05$). Although US participants still reported positional concerns more frequently across most domains (excluding income), the absolute differences were small (ranging from 0.005 to 0.045) and not statistically significant.

A more detailed breakdown is provided in Table A2.8 in the Appendix, which reports the proportion of respondents choosing the positional option across both countries and treatments for all ten domains. To move beyond merely descriptive differences and better isolate potential contextual effects of country of residence, we estimated logistic regression models including covariates for individual preferences and sociodemographic characteristics (age, gender, education, income and political orientation). These controls account for underlying country-level variation that may also influence positional choices. The dependent variable is a binary indicator, defined as $D_i^b = 1$ if individual i selected the positional option for good b, and zero otherwise; thereby estimating the probability of choosing the positional option in a given domain: two economic (income, vacation time) and four health-related (days of illness, physical activity, doctor density, hospital bed density) variables. Results are shown for the six domains in which positional concerns were most pronounced, that is, scenarios with the highest observed rates of positional decision-making (ranging from 10.09% to 29.46%). Table 2.3 summarizes the results for all six models; corresponding unadjusted models are reported in Table A2.10.

The results, reported in Table 2.3, do not provide evidence that being located in the United States is associated with a statistically significant increase in the likelihood of expressing positional concerns. The country effect is positive but small and not significant across all outcomes, except for physical activity (AME = 0.055, SE = 0.026, $p = 0.03$), and to a marginal

Table 2.3: Average marginal effects from logistic regression models predicting positional choice across domains

	DV: Dummy variable (positional choice)					
	Income	Vacation time	Days of illness	Physical activity	Doctor density†	Hospital bed density†
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment						
Group: Friends	−0.228***	−0.068**	−0.072***	−0.028	−0.064**	−0.027
	(0.029)	(0.023)	(0.019)	(0.026)	(0.019)	(0.023)
Country: US	−0.011	0.010	0.015	0.055*	0.036	0.030
	(0.029)	(0.023)	(0.020)	(0.026)	(0.020)	(0.023)
Cluster (ref: 3: Self-interest)						
1: Inequality aversion	0.081*	0.090**	0.071**	−0.001	0.040	0.083**
	(0.038)	(0.031)	(0.027)	(0.031)	(0.027)	(0.031)
2: Envy	0.149***	0.087**	0.076**	0.149***	0.028	0.091**
	(0.035)	(0.027)	(0.023)	(0.033)	(0.023)	(0.028)
Preferences						
Patience	0.003	−0.009	−0.011*	−0.021**	−0.014**	−0.010
	(0.008)	(0.006)	(0.005)	(0.007)	(0.005)	(0.006)
Risk tolerance	0.005	0.002	0.005	0.008	0.007	0.016**
	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.005)
Altruism	−0.011	0.003	0.000	−0.000	−0.002	0.003
	(0.006)	(0.005)	(0.004)	(0.006)	(0.004)	(0.005)
Social comparison	0.001	0.041*	0.040*	0.025	0.014	−0.018
	(0.024)	(0.019)	(0.017)	(0.021)	(0.016)	(0.018)
Controls	Y	Y	Y	Y	Y	Y
Observations	944	944	944	944	944	944

Robust standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Results are average marginal effects (AME) from logistic regression models. † represents healthcare provision ratios: doctor-to-population and hospital beds-to-population ratio. *Friends* refer to the *friends and acquaintances* reference group. All models include controls for age, gender, education, income, and political orientation (coefficients not shown).

extent, the doctor-to-population ratio (AME = 0.036, SE = 0.020, $p = 0.07$). This finding stands in contrast with our initial hypothesis and suggests that country-level differences are negligible in this sample. Overall, the results suggest that national context, as measured by residence in a more unequal society (the US compared to the UK), is not strongly associated with positional concerns.

2.3.3 Underlying drivers of positional concerns

Following the cross-country comparison, we turn to the individual-level mechanisms behind positional decision-making. Our analysis follows the framework of Diaz *et al.* (2023), which draws on the Fehr and Schmidt (1999) model to classify preference types. We cluster these types according to their response profiles and distinguish between three salient profiles (self-interest, inequality aversion, and envy), to explain positional choice and preferences for relative standing. We further investigate how time and risk preferences, altruism, and social comparison relate to the probability of selecting positional options. All estimates are based on the logistic regressions reported in Table 2.3 and focus on the six domains where positional choices were most frequent.

Inequality aversion and envy. Across all domains, both inequality-averse and envious individuals were more likely than the self-interested group to select positional options. For the envy cluster, stronger positional concerns appeared in five of six domains, with the largest effects in income and physical activity (AME = 0.149, $p < 0.001$ for both), with smaller but significant effects for vacation time (0.087, $p < 0.01$), days of illness (0.076, $p < 0.01$), and healthcare provision (hospital beds-to-population ratio) (0.091, $p < 0.01$). Similarly, we found that inequality-averse individuals were more likely than the self-interested to make positional choices in several domains, including income (0.081, $p < 0.05$), vacation time (0.090, $p < 0.01$), days of illness (0.071, $p < 0.01$), and hospital beds-to-population ratio (0.083, $p < 0.01$). Interestingly, neither cluster showed significant effects in the doctor-to-population ratio domain.

Time and risk preferences. More patient individuals were significantly less likely to select the positional options, especially in health-related contexts: Patience was negatively associated with positional choice for physical activity (AME = -0.021, $p < 0.01$), followed by doctor-to-population ratio (-0.014, $p < 0.01$) and days of illness (-0.011, $p < 0.05$). Risk tolerance was not significantly associated with positional choice in most domains, and showed only a positive effect in hospital beds-to-population ratio (0.016, $p < 0.01$).

Altruism and social comparison. Altruism was not a statistically significant predictor of positional choice in any domain. The social comparison index, by contrast, was positively associated with positional choices in vacation time (AME = 0.041, $p < 0.05$) and days of illness (0.040, $p < 0.05$). No statistically significant associations were observed in other domains, and the direction of effects varied.

2.4 Conclusion

This chapter examined positional concerns across economic and health domains using a survey experiment administered to gender-balanced samples in the United States and the United Kingdom. Our descriptive analysis shows that positional choices vary by domain (from nearly 30% for income to 4% for life expectancy). Fewer than 5% expressed status concerns about waiting time and life expectancy, which appears substantially lower than the proportions reported in previous studies (Solnick and Hemenway, 2005; Wouters *et al.*, 2015). These findings lend support to the view that positional concerns are generally less pronounced in health-related domains than in economic ones. One contributing factor could be our fixed γ calibration, whereas other studies implicitly allow the degree of positionality to vary. While our standardization approach enhances comparability across domains, it may also understate how many people truly care about being healthier relative to others. In health domains, the share choosing the positional option could be lower simply because these choices may require a stronger status incentive to trigger the same switch. Although we find the strongest positional signal for income, our participants also expressed concerns about physical activity, which indicates that status sensitivity also extends to health behaviors, particularly those that are visible or socially salient. This tendency becomes more pronounced when we stratify by reference group treatment, as discussed in Section 2.3.2.1 (see Table A2.8 in the Appendix). Our prediction that stronger positional concerns would emerge for visible or publicly observable health outcomes (e.g., physical activity) than for less tangible or diffuse ones (e.g., life expectancy or air quality) is also supported. Respondents were clearly more likely to display status sensitivity in domains tied to everyday behaviors than in those involving long-term outcomes or collective health resources. We also elicited individuals'

second-order beliefs and report that people's own preferences significantly deviate from what they believe others would prefer.

Our analysis further shows that positional choices vary significantly depending on the nature of the reference group, whether respondents compare themselves to the general population or to more proximate social ties, such as friends and acquaintances. We also find that individual social preferences, particularly envy and inequality aversion, are closely associated with these choices. For example, envy emerges as the strongest predictor of positionality in domains such as physical activity.

We find little evidence that positional concerns systematically vary by socioeconomic or demographic factors such as age, gender, or income. This echoes earlier findings noting that these characteristics exert at most a limited influence on positional preferences (Solnick and Hemenway, 2005). In terms of cross-country comparisons, our results are consistent with recent experimental evidence by Daniel, van Exel and Chorus (2023), who observed similarly modest and statistically non-significant differences in positional choices between the US and UK. However, Daniel, van Exel and Chorus (2023) focused more narrowly on healthcare access, i.e., waiting time allocation, neglecting other domains such as health outcomes or service availability. While institutional differences in healthcare provision between the US and the UK might still contribute to some of the variation observed in health-related domains, these cross-country differences tend to be small and mostly not statistically significant. It is also important to acknowledge that beyond economic, social, or health disparities, there may be other contextual factors and socioeconomic characteristics affecting our findings, the precise contribution of which we cannot assess here. Thus, while we selected these two samples to proxy underlying differences in healthcare systems, further research is required to confirm whether, for example, the type of healthcare system, i.e., whether it is predominantly privately or publicly funded, may be driving these results.

Our results further highlight the role of individual-level preference profiles in impacting positional behavior. Respondents classified as envious or inequality-averse were more likely to prioritize relative standing than those categorized as self-interested, across domains. Our composite social comparison variable, comprising social image concerns as well as tendencies toward social comparison, showed limited and domain-specific associations with positionality.

While prior studies often assumed that altruistic concerns might be an opposing force to the display of positional concerns (e.g., Mageli, Mannberg and Heen, 2022), in our study, altruism was not significantly linked to positional choice in any of the six domains examined. Positionality appears neither uniformly driven by status-seeking nor consistently offset by altruistic concerns, and our results point to a more nuanced and context-dependent role of social preference profiles than is reflected in the literature.

Limitations. Several limitations should be acknowledged. First, the study relies on hypothetical trade-offs, which, while standard in this type of experimental research, may not accurately reflect behavior in real-world decision-making contexts. This limits external validity and raises the question of how well stated preferences translate into actual choices. We used a slider task to elicit indifference thresholds and obtain more precise estimates of individual positionality, but this approach was constrained by small sample sizes, particularly for life expectancy preferences (N = 36). While this group did show positional responses, the small sample size does not allow us to draw meaningful conclusions. Second, although our health-related domains clearly expand the empirical scope of the literature, the scenarios remain limited. Areas such as mental health, preventive care, or chronic illness management may involve different mechanisms and warrant further investigation. Third, while our models include envy and inequality aversion, they do not capture the full spectrum of social preferences or motivations that might drive positional behavior. Competitiveness, reciprocity, or fairness norms could also matter but are not directly addressed. Fourth, although we account for basic demographic characteristics, unobserved heterogeneity, such as personality traits or experiences over the life-course, could still influence outcomes. Finally, generalizability remains an open question: our findings come from two WEIRD samples – Western, educated, industrialized, rich, and democratic – whose institutional contexts differ from each other but are not representative of the world as a whole.

Interpretation and implications. Our findings underscore the role of social preferences and comparative processes as significant drivers of positional choices. While we do not directly model competitive motives, we indirectly account for them through a composite indicator of social comparison, an indicator measuring the degree to which individuals who care more about how they compare to others prefer being better off relative to others. Social

preferences have long been recognized as major determinants of individual behaviors and help explain why individuals often make choices that deviate from self-interest (Fehr and Schmidt, 1999; Hill and Buss, 2006). Although some studies suggest that social preferences such as envy may be important determinants of positional concerns (Mujcic and Frijters, 2015; Solnick and Hemenway, 1998), few studies had previously explored the underlying factors of positional preferences (Celse, Galia and Max, 2017). Our results support the idea that envy and inequality aversion are important drivers of positional concerns across and within most domains, while we do not find support for the hypothesis that altruistic preferences are negative correlates of positionality in any of our specifications.

Overall, these findings contribute new evidence to the literature on positional concerns by showing how they manifest not only in traditional economic domains, but also in areas often viewed as less status-oriented, such as health. We show that preferences for relative standing extend beyond income and consumption. We highlight the significant role of envy and inequality aversion in driving positional choices and confirm that relative concerns are not strongly tied to socioeconomic and demographic factors, suggesting a broader applicability across different population groups. Our findings have important implications for policy-making. Policies to improve health or economic outcomes will likely be more successful if they work at the community level rather than exclusively through national campaigns. Local health programs or community-based financial literacy initiatives, for example, can harness the influence of close social ties. By encouraging comparisons within their immediate social circle, with friends, neighbors, or others in daily life, policymakers may be able to design interventions that could connect and resonate better with people and deliver greater impact.

References

Alba, Beatrice, Doris McIlwain, Ladd Wheeler, and Michael P Jones. 2014. "Status consciousness: A preliminary construction of a scale measuring individual differences in status-relevant attitudes, beliefs, and desires." *Journal of Individual Differences*, 35(3): 166.

Alpizar, Francisco, Fredrik Carlsson, and Olof Johansson-Stenman. 2005. "How much do we care about absolute versus relative income and consumption?" *Journal of Economic Behavior & Organization*, 56(3): 405–421.

Alvarez-Cuadrado, Francisco, and Ngo Van Long. 2012. "Envy and inequality." *The Scandinavian Journal of Economics*, 114(3): 949–973.

Aronsson, Thomas, and Olof Johansson-Stenman. 2008. "When the Joneses' consumption hurts: Optimal public good provision and nonlinear income taxation." *Journal of Public Economics*, 92(5-6): 986–997.

Aronsson, Thomas, and Olof Johansson-Stenman. 2013. "Veblen's theory of the leisure class revisited: implications for optimal income taxation." *Social Choice and Welfare*, 41: 551–578.

Aronsson, Thomas, Sugata Ghosh, and Ronald Wendner. 2023. "Positional preferences and efficiency in a dynamic economy." *Social Choice and Welfare*, 61(2): 311–337.

Blanchflower, David G, Bert Van Landeghem, and Andrew J Oswald. 2009. "Imitative obesity and relative utility." *Journal of the European Economic Association*, 7(2-3): 528–538.

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

Bowles, Samuel, and Yongjin Park. 2005. "Emulation, inequality, and work hours: Was Thorsten Veblen right?" *The Economic Journal*, 115(507): F397–F412.

Bull, Fiona C, Salih S Al-Ansari, Stuart Biddle, Katja Borodulin, Matthew P Buman, Greet Cardon, Catherine Carty, Jean-Philippe Chaput, Sebastien Chastin, Roger Chou, *et al.* 2020. "World Health Organization 2020 guidelines on physical activity and sedentary behaviour." *British journal of sports medicine*, 54(24): 1451–1462.

Carlsson, Fredrik, Olof Johansson-Stenman, and Peter Martinsson. 2007. "Do you enjoy having more than others? Survey evidence of positional goods." *Economica*, 74(296): 586–598.

Carrieri, Vincenzo. 2012. "Social Comparison and Subjective Well-Being: Does the Health of Others Matter?" *Bulletin of Economic Research*, 64(1): 31–55.

Celse, Jérémy. 2012. "Is the positional bias an artefact? Distinguishing positional concerns from egalitarian concerns." *The Journal of Socio-Economics*, 41(3): 277–283.

Celse, Jérémy, and Gilles Grolleau. 2021. "Keeping up with the Joneses: Examining relative concerns in health-related domains." *Journal de gestion et d'économie de la santé*, 39(1): 21–44.

Celse, Jérémy, Fabrice Galia, and Sylvain Max. 2017. "Are (negative) emotions to blame for being positional? An experimental investigation of the impact of emotional states on status preferences." *Journal of Behavioral and Experimental Economics*, 67: 122–130.

Charles, Kerwin Kofi, Erik Hurst, and Nikolai Roussanov. 2009. "Conspicuous consumption and race." *The Quarterly Journal of Economics*, 124(2): 425–467.

Clark, Andrew E, and Claudia Senik. 2010. "Who compares to whom? The anatomy of income comparisons in Europe." *The Economic Journal*, 120(544): 573–594.

Clark, Andrew E, and Conchita d'Ambrosio. 2015. "Attitudes to income inequality: Experimental and survey evidence." In *Handbook of income distribution*. Vol. 2, 1147–1208. Elsevier.

Cooper, David J, and John H Kagel. 2016. "Other-regarding preferences." *The handbook of experimental economics*, 2: 217.

Daniel, Aemiro Melkamu, Job van Exel, and Caspar G Chorus. 2023. "Self-interest, positional concerns and distributional considerations in healthcare preferences." *The European Journal of Health Economics*, 1–24.

Diaz, Lina, Daniel Houser, John Ifcher, and Homa Zarghamee. 2023. "Estimating social preferences using stated satisfaction: Novel support for inequity aversion." *European Economic Review*, 155: 104436.

Duesenberry, James Stemble. 1949. "Income, saving and the theory of consumer behavior." *Harvard University Press, Cambridge.*

Esguerra, Emilio, Leonhard Vollmer, and Johannes Wimmer. 2023. "Influence Motives in Social Signaling: Evidence from COVID-19 Vaccinations in Germany." *American Economic Review: Insights*, 5(2): 275–291.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2023. "The preference survey module: A validated instrument for measuring risk, time, and social preferences." *Management Science*, 69(4): 1935–1950.

Fallucchi, Francesco, R Andrew Luccasen III, and Theodore L Turocy. 2019. "Identifying discrete behavioural types: a re-analysis of public goods game contributions by hierarchical clustering." *Journal of the Economic Science Association*, 5(2): 238–254.

Fallucchi, Francesco, R Andrew Luccasen III, and Theodore L Turocy. 2022. "The sophistication of conditional cooperators: Evidence from public goods games." *Games and Economic Behavior*, 136: 31–62.

Faul, Franz, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. "Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses." *Behavior research methods*, 41(4): 1149–1160.

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

Frank, Robert H. 1985. "The demand for unobservable and other nonpositional goods." *The American Economic Review*, 75(1): 101–116.

Frank, Robert H. 2008. "Should public policy respond to positional externalities?" *Journal of Public Economics*, 92(8-9): 1777–1786.

Frank, Robert H. 2013. *Falling behind: How rising inequality harms the middle class.* Vol. 4, Univ of California Press.

GBD 2019 Demographics Collaborators. 2020. "Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019." *The Lancet*, 396(10258): 1160–1203.

Gibbons, Frederick X, and Bram P Buunk. 1999. "Individual differences in social comparison: development of a scale of social comparison orientation." *Journal of personality and social psychology*, 76(1): 129.

Grolleau, Gilles, and Sandra Saïd. 2008. "Do you prefer having more or more than others? Survey evidence on positional concerns in France." *Journal of Economic Issues*, 42(4): 1145–1158.

Grolleau, Gilles, Lisette Ibanez, and Naoufel Mzoughi. 2012. "Being the best or doing the right thing? An investigation of positional, prosocial and conformist preferences in provision of public goods." *The journal of socio-economics*, 41(5): 705–711.

Gugushvili, Alexi. 2021. "Which socio-economic comparison groups do individuals choose and why?" *European Societies*, 23(4): 437–463.

Heffetz, Ori, and Robert H Frank. 2011. "Preferences for status: Evidence and economic implications." In *Handbook of social economics*. Vol. 1, 69–91. Elsevier.

Hemenway, David. 2021. "Positional goods and the underfunding of public health." *Journal of Public Health Policy*, 42(3): 522–524.

Hillesheim, Inga, and Mario Mechtel. 2013. "How much do others matter? Explaining positional concerns for different goods and personal characteristics." *Journal of Economic Psychology*, 34: 61–77.

Hill, Sarah E, and David M Buss. 2006. "Envy and positional bias in the evolutionary psychology of management." *Managerial and Decision Economics*, 27(2-3): 131–143.

Huțu, Daniela, Carmen Marinela Cumpăt, Andreea Grădinaru, and Bogdan Rusu. 2024. "The Impact of Moral Hazard on Healthcare Utilization in Public Hospitals from Romania: Evidence from Patient Behaviors and Insurance Systems." Vol. 12, 2519, MDPI.

Johansson-Stenman, Olof, Fredrik Carlsson, and Dinky Daruvala. 2002. "Measuring future grandparents' preferences for equality and relative standing." *The economic journal*, 112(479): 362–383.

Krupka, Erin L, and Roberto A Weber. 2013. "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association*, 11(3): 495–524.

Leites, Martín, Analía Rivero, and Gonzalo Salas. 2024. "The positionality of goods and the positional concern's origin." *Journal of Behavioral and Experimental Economics*, 102184.

Mageli, Ingvild, Andrea Mannberg, and Eirik Eriksen Heen. 2022. "With whom, and about what, do we compete for social status? Effects of social closeness and relevance of reference groups for positional concerns." *Journal of Behavioral and Experimental Economics*, 98: 101867.

Mathieu-Bolh, Nathalie, and Ronald Wendner. 2020. "We are what we eat: obesity, income, and social comparisons." *European Economic Review*, 128: 103495.

Mujcic, Redzo, and Paul Frijters. 2015. "Conspicuous consumption, conspicuous health, and optimal taxation." *Journal of Economic Behavior & Organization*, 111: 59–70.

Mullen, Kathleen J, and Stephanie Rennane. 2017. "Worker absenteeism and employment outcomes: a literature review." *Unpublished manuscript, RAND Corporation*.

Müller, Nathalie, Francesco Fallucchi, Francesco Principe, and Marc Suhrcke. 2024. "Preregistration: Positional preferences for health and well-being – a survey experiment." *https://doi.org/10.17605/OSF.IO/MJZKV*, Registered on the Open Science Framework, February 14, 2024.

Mumtaz, Zubia, Adrienne Levay, Afshan Bhatti, and Sarah Salway. 2013. "Signalling, status and inequities in maternal healthcare use in Punjab, Pakistan." *Social Science & Medicine*, 94: 98–105.

Neal, RD, P Tharmanathan, B France, NU Din, S Cotton, J Fallon-Ferguson, W Hamilton, A Hendry, M Hendry, Ruth Lewis, *et al.* 2015. "Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review." *British journal of cancer*, 112(1): S92–S107.

OECD. 2020. "How's Life? 2020: Measuring Well-being." *OECD Publishing*.

OECD. 2023. "OECD Data Explorer – Health Statistics." *https://data-explorer.oecd.org/*, Accessed: 13 December 2023. Figures used in analysis reflect values available at that time.

ONS. 2023. "Sickness absence in the UK labour market: 2022." *Office for National Statistics (ONS)*.

Postlewaite, Andrew. 1998. "The social basis of interdependent preferences." *European Economic Review*, 42(3-5): 779–800.

Ray, Rebecca, Milla Sanes, and John Schmitt. 2013. "No-vacation nation revisited." *Center for Economic and Policy Research*, 1–22.

Schneider, Eric C, Arnav Shah, Michelle M Doty, Roosa Tikkanen, Katharine Fields, and Reginald D Williams II. 2021. "Mirror, mirror 2021 – Reflecting Poorly: Health Care in the U.S. Compared to Other High-Income Countries." *New York: The Commonwealth Fund*, 4.

Solnick, Sara J, and David Hemenway. 1998. "Is more always better? A survey on positional concerns." *Journal of Economic Behavior & Organization*, 37(3): 373–383.

Solnick, Sara J, and David Hemenway. 2005. "Are positional concerns stronger in some domains than in others?" *American Economic Review*, 95(2): 147–151.

Statista. 2025. "Share of people with tertiary education in OECD and affiliated countries in 2022, by country." *https://www.statista.com/statistics/1227287/share-of-people-with-tertiary-education-in-oecd-countries-by-country*, Accessed: 11 June 2025.

Vandegrift, Donald, and Kristen Duke. 2015. "Competitive behavior, impact on others, and the number of competitors." *Journal of Behavioral and Experimental Economics*, 57: 37–44.

Veblen, Thorstein. 1899. "The theory of the leisure class." *Journal of Political Economy*, 7(4): 425–455.

Velandia-Morales, Andrea, Rosa Rodríguez-Bailón, and Rocío Martínez. 2022. "Economic inequality increases the preference for status consumption." *Frontiers in Psychology*, 12: 809101.

WHO. 2021. "WHO global air quality guidelines. Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide." *Geneva: World Health Organization*.

Wouters, Sofie, NJA van Exel, M Van de Donk, Kim Rohde, and WBF Brouwer. 2015. "Do people desire to be healthier than other people? A short note on positional concerns for health." *The European Journal of Health Economics*, 16: 47–54.

Appendix

A2.1 Descriptive statistics

Table A2.1: Summary statistics for full sample (UK and US, $N = 981$)

Statistic	Mean	SD	Min	Max	N
Experimental assignment and study sample					
Reference group: General population (%)	49.7	–	–	–	981
Country: US (%)	50.3	–	–	–	981
Positional choices – Economic domains					
Income (%)	29.5	45.6	0	100	981
Vacation time (%)	15.0	35.7	0	100	981
Positional choices – Health domains					
Life expectancy (%)	4.2	20.0	0	100	981
Air quality (%)	5.6	23.0	0	100	981
Days of illness (%)	10.2	30.3	0	100	981
Good child health (%)	6.5	24.7	0	100	981
Physical activity (%)	20.7	40.5	0	100	981
Cancer screening wait time (%)	4.8	21.4	0	100	981
Doctor-to-population ratio (%)	10.1	30.1	0	100	981
Hospital beds-to-population ratio (%)	14.6	35.3	0	100	981
Second-order beliefs – Positional choices					
Income (%)	25.5	43.6	0	100	981
Life expectancy (%)	9.5	29.3	0	100	981
Cancer screening wait time (%)	5.5	22.8	0	100	981
Slider task – Indifference thresholds					
Income – UK (GBP)	2733.8	196.0	2200	2900	142
Income – US (USD)	4505.6	332.7	3900	4800	143
Life expectancy (years)	84	2.7	80	87	36
Economic, social and distributional preferences					
Patience	6.89	1.92	0	10	981
Risk tolerance	5.32	2.32	0	10	981
Altruism	6.88	2.47	0	10	981
Social image concerns	3.02	1.02	1	5	981
Social comparison	3.04	0.62	1	5	981
Political attitude	3.2	1.48	1	7	981
Cluster 1: Inequality aversion (%)	20.5	40.4	0	100	981
Cluster 2: Envy (%)	29.6	45.7	0	100	981
Cluster 3: Self-interest (%)	49.9	50.0	0	100	981
Demographic characteristics					
Female (%)	48.5	50.0	0	100	981
Age (years)	36.2	9.7	21	59.5	981
Higher education (%)	58.2	49.3	0	100	981
Employed full-time (%)	59.2	49.2	0	100	981
Financial well-being	4.6	1.2	1	6	961
Self-reported health status	3.3	0.9	1	5	981
Income quintile 1 (%)	20.2	40.2	0	100	981
Income quintile 2 (%)	20.1	40.1	0	100	981
Income quintile 3 (%)	19.8	39.9	0	100	981
Income quintile 4 (%)	19.8	39.9	0	100	981
Income quintile 5 (%)	20.2	40.2	0	100	981

Note: Percentages (%) reflect the share of participants falling into the indicated category. Continuous variables are reported as means and standard deviations (SD). *Social image concerns* and *social comparison* are composite variables based on the average of two and three items, respectively, each rated on a 5-point Likert scale (see Table A2.6). Demographic variables are coded as follows: *Female* is a binary indicator; *Age* is the mean of age-bin midpoints; *Higher education* is a binary indicator (university bachelor's degree or higher); 1 (part-time); *Income quintile* dummies are sample-based income quintiles (1 = lowest, 5 = highest), computed within each country.

A2.2 Design and justification of experimental variables

A2.2.1 Binary choice scenario design

Table A2.2: Reference values and sources for hypothetical scenario design

Variable and description	Hypothetical scenario: Society A vs. B	Empirical data and data source
Economic domains		
Income – UK Monthly income after tax (GBP)	A: You £2,400, Others £3,000; B: You £2,100, Others £1,800	**UK average:** £2,177/month (OECD, 2020)
Income – US Monthly income after tax (USD)	A: You $4,100, Others $4,900; B: You $3,700, Others $3,200	**US average:** $4,262/month (OECD, 2020)
Vacation time Annual leave in weeks	A: You 4, Others 6; B: You 3, Others 2	**UK minimum:** 28 days (statutory) **US average:** 10 days (0 guaranteed) (Ray, Sanes and Schmitt, 2013)
Health domains		
Life expectancy Lifespan at birth (years)	A: You 84, Others 87; B: You 79, Others 68	**UK:** 80.4 years **US:** 76.4 years (OECD, 2023)
Air quality Days of high pollution/year	A: Your area 13, Others' 8; B: Your area 18, Others' 29	**WHO guideline:** max. 3–4 days/year exceeding recommended thresholds (WHO, 2021)
Days of illness Sick days/year	A: You 4, Others 2.5; B: You 5, Others 6	**UK average:** 5.7 days lost/year **US estimates:** 3–4 days missed/year (Mullen and Rennane, 2017; ONS, 2023)
Good child health Child healthy life expectancy (years)	A: Yours 63, Others' 67; B: Yours 60, Others' 55	**UK:** 68.9 years **US:** 65.2 years (GBD 2019 Demographics Collaborators, 2020)
Physical activity Weekly minutes of activity	A: You 125, Others 160; B: You 110, Others 95	**WHO recommendation:** 150–300 min/week moderate-intensity (Bull *et al.*, 2020)
Cancer screening wait time Weeks to cancer screening	A: You 3, Others 1; B: You 5, Others 7	**UK:** 3–5 weeks **US:** 1–2 (insured), 5–7 (limited access) (Neal *et al.*, 2015)
Doctor-to-population ratio Number of people per doctor	A: Your area 200, Others' 140; B: Your area 250, Others' 340	**UK:** 3.2 physicians/1,000 people **US:** 3.5 physicians/1,000 people (OECD, 2023)
Hospital beds-to-population ratio Number of people per hospital bed	A: Your area 160, Others' 125; B: Your area 200, Others' 300	**UK:** 2.4 beds/1,000 people **US:** 2.8 beds/1,000 people (OECD, 2023)

Notes: Each item describes a binary choice between two hypothetical societies, A and B, differing on a single domain-specific attribute. The values are based on empirical data and used as design parameters to create scenarios consistent with a degree of positionality of $\gamma \approx 0.25$. The incentivized bonus scenarios, in which participants were asked to predict the majority preference, were constructed similarly. Reference groups ("Others") refer to either the *general population* or to *friends and acquaintances*, depending on randomized treatment assignment. For health care access indicators (doctor and hospital beds-to-population ratio), scenario values are expressed as people per doctor (or per hospital bed, respectively) to emphasize scarcity. The empirical benchmarks use the inverse (units per 1,000 people), following public health reporting standards.

A2.2.2 Stated satisfaction, social preferences and clustering

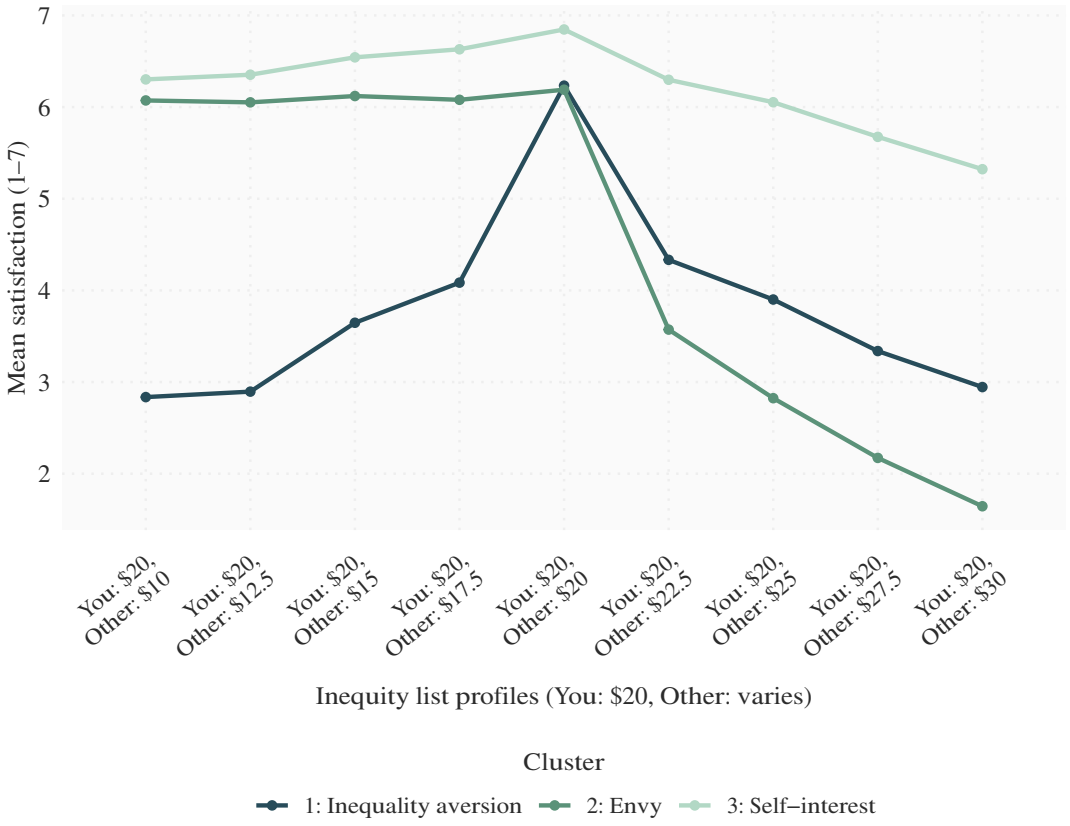


Figure A2.1: Mean satisfaction patterns across distributional choice scenarios by cluster

Table A2.3: Stated satisfaction profiles by cluster (mean scores)

Cluster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
1: Inequality aversion	2.84	2.90	3.65	4.08	6.23	4.33	3.90	3.34	2.95
2: Envy	6.07	6.05	6.12	6.08	6.19	3.57	2.82	2.17	1.64
3: Self-interest	6.30	6.35	6.54	6.63	6.85	6.30	6.05	5.68	5.32

Notes: Values represent mean satisfaction statements for each inequity list profile across clusters, following the stated satisfaction method adopted from Diaz *et al.* (2023). Ratings were based on an inequity list (based on Diaz *et al.*, 2023). "Q1–Q9" correspond to satisfaction with nine allocations in which the participant always received $20, while the partner's amount varied from $10 to $30. Responses were given on a 7-point Likert scale (1 = *Extremely dissatisfied*, 7 = *Extremely satisfied*). Cluster sizes: 1: $n = 201$, 2: $n = 290$, 3: $n = 490$.

Table A2.4: Cluster descriptions based on satisfaction profiles

Cluster	Label	Profile description
1	Inequality aversion	Satisfaction follows a ∩-shaped pattern, peaking at equal distributions and declining as either person becomes better off, consistent with aversion to both advantageous and disadvantageous inequality.
2	Envy	Satisfaction is stable as long as the respondent is better off and declines sharply as the other (hypothetical) person earns more. Average response pattern is consistent with aversion to disadvantageous inequality.
3	Self-interest	Satisfaction remains relatively high and stable when participants are better off compared to a hypothetical other person, and decreases only slightly otherwise, indicating that preferences are primarily driven by own payoffs.

Notes: Clusters were derived using K-means clustering on standardized satisfaction ratings from nine distributional choice scenarios. Following Diaz *et al.* (2023), stated satisfaction was used to infer individual-level distributional preferences. While the original design featured a more extensive set of scenarios, I simplified the task to nine conditions to reduce participant burden. Cluster labels reflect the best-fitting profile types in Diaz *et al.*'s nomenclature. See Table A2.5 for distributions by treatment condition. For further discussion of pattern classification and alternative models, see Diaz *et al.* (2023).

Table A2.5: Participant distribution across clusters, by treatment condition and country

Condition		Cluster		
		1: Self-interest (%)	2: Inequality aversion (%)	3: Envy (%)
Reference group	Friends and acquaintances	20.1	33.4	46.5
	General population	20.9	25.8	53.3
Country	United Kingdom	17.2	33.4	49.4
	United States	23.7	25.8	50.5

Note: Percentage share of participants per cluster, by treatment condition (reference group: friends and acquaintances vs. general population) and country (UK vs. US).

A2.2.3 Orientation toward social evaluation: variable coding and index construction

Table A2.6 summarizes the five additional items that were included in the survey experiment to measure individual differences in status-related motivations, namely *social image concerns* and *social comparison*. For the analysis, we constructed three indices to represent these complementary motivations, in addition to the economic, social, and distributional preference measures employed.

All responses were recoded to a uniform five-point ordinal scale, with higher values indicating stronger agreement. Two indices were constructed for analytical purposes: the social image index reflects the average of two items measuring concerns related to impressing others and being perceived favorably. The social comparison index reflects the average of three items capturing tendencies to compare oneself with others in terms of performance, life circumstances, and health.

A third composite index aggregates all five items to form an indicator of an overall orientation toward social evaluation, both inward-looking (comparison) and outward-focused (image concerns), and is used to predict variance in positional preferences across individuals or groups.

Table A2.6: Description of social comparison and social image variables

Variable	Type	Survey item	Notes
Social image			
Impressing others	Ordinal	"It is important to me to impress others."	Based on Alba *et al.* (2014) and Esguerra, Vollmer and Wimmer (2023)
Being perceived by others	Ordinal	"It is important to me how I am perceived by others."	Based on Alba *et al.* (2014) and Esguerra, Vollmer and Wimmer (2023)
Social comparison			
Performance	Ordinal	"I always pay a lot of attention to how I do things compared with how others do things."	Item 2 from INCOM Gibbons and Buunk (1999)
Life situation	Ordinal	"I never consider my situation in life relative to that of others." (reverse-coded)	Item 11 from INCOM Gibbons and Buunk (1999)
Health	Ordinal	"I often compare myself on how I am doing in terms of health."	Custom item, based on item 4 from INCOM Gibbons and Buunk (1999)

Notes: All items were preceded by the statement "Most people compare themselves from time to time with others. For example, they may compare the way they feel, their opinions, their abilities, and/or their situation with those of other people. There is nothing particularly 'good' or 'bad' about this type of comparison, and some people do it more than others. We will now show you 5 statements that might apply to you." adapted from the Iowa-Netherlands Comparison Orientation Measure (INCOM) framing; see also Gibbons and Buunk (1999). Participants rated each item on a 5-point Likert scale ranging from 1 = *Strongly disagree* to 5 = *Strongly agree*.

A2.3 Conversion tables

Mapping individual choices. The degree of positionality γ can be determined from Equation 2.1, and is calculated as $\gamma = \frac{\log(x_B/x_A)}{\log(\bar{x}_B/\bar{x}_A)}$, where x_A and x_B are the individual's chosen options in societies A and B, and $\bar{x}_A$, $\bar{x}_B$ are the corresponding societal averages. Higher γ reflects greater concern with relative standing.

Table A2.7: Degree of positionality (γ) conversion for income and life expectancy scenarios

(a) Income (UK)

x_A	x_B	$\bar{x}_A$	$\bar{x}_B$	γ
2200	2100	3000	1800	0.091
2300	2100	3000	1800	0.178
2400	2100	3000	1800	0.261
2500	2100	3000	1800	0.341
2600	2100	3000	1800	0.418
2700	2100	3000	1800	0.492
2800	2100	3000	1800	0.563
2900	2100	3000	1800	0.632

(b) Income (US)

x_A	x_B	$\bar{x}_A$	$\bar{x}_B$	γ
3900	3700	4900	3200	0.124
4000	3700	4900	3200	0.183
4100	3700	4900	3200	0.241
4200	3700	4900	3200	0.298
4300	3700	4900	3200	0.353
4400	3700	4900	3200	0.407
4500	3700	4900	3200	0.459
4600	3700	4900	3200	0.511
4700	3700	4900	3200	0.561
4800	3700	4900	3200	0.611

(c) Life expectancy

x_A	x_B	$\bar{x}_A$	$\bar{x}_B$	γ
80	79	88	68	0.049
81	79	88	68	0.097
82	79	88	68	0.145
83	79	88	68	0.192
85	79	88	68	0.284
86	79	88	68	0.329
87	79	88	68	0.374

A2.4 Comparison of proportions and chi-squared tests: positional preferences by reference group and by country

Table A2.8: Chi-squared tests and differences in proportions of positional choices: friends and acquaintances vs. general population (UK and US)

	Friends^{UK}	Others^{UK}	Diff.^{UK}	Friends^{US}	Others^{US}	Diff.^{US}
Economic domain						
Income	0.176	0.420	-0.244^{***}	0.190	0.396	-0.206^{***}
Vacation time	0.110	0.185	-0.075^{*}	0.129	0.176	-0.046
Health domain						
Life expectancy	0.029	0.049	-0.021	0.048	0.041	0.008
Air quality	0.029	0.066	-0.037	0.036	0.094	-0.058^{**}
Days of illness	0.061	0.123	-0.062^{*}	0.081	0.143	-0.062^{*}
Good child health	0.045	0.078	-0.033	0.069	0.069	-0.001
Physical activity	0.171	0.198	-0.026	0.222	0.237	-0.015
Cancer screening wait time	0.029	0.053	-0.025	0.024	0.086	-0.062^{**}
Doctor-to-population ratio	0.057	0.107	-0.050^{*}	0.089	0.151	-0.062^{*}
Hospital beds-to-population ratio	0.110	0.140	-0.030	0.153	0.180	-0.026
Second-order beliefs						
Income	0.188	0.263	-0.076^{*}	0.218	0.351	-0.133^{**}
Life expectancy	0.086	0.086	-0.001	0.073	0.135	-0.062^{*}
Cancer screening wait time	0.029	0.058	-0.029	0.044	0.090	-0.045^{*}

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Friends refer to the *friends and acquaintances* reference group, and *others* to the *general population*, respectively.

Table A2.9: Chi-squared tests and differences in proportions of positional choices: UK vs. US (full sample).

	UK	US	Diff.	p-value
Economic domain				
Income	0.297	0.292	−0.416	0.862
Vacation time	0.148	0.152	−0.696	0.840
Health domain				
Life expectancy	0.039	0.045	−0.911	0.656
Air quality	0.047	0.065	−0.870	0.226
Days of illness	0.092	0.112	−0.777	0.317
Good child health	0.061	0.069	−0.862	0.635
Physical activity	0.184	0.229	−0.542	0.083
Cancer screening wait time	0.041	0.055	−0.890	0.312
Doctor-to-population ratio	0.082	0.120	−0.761	0.050
Hospital beds-to-population ratio	0.125	0.166	−0.667	0.067
Second-order beliefs				
Income	0.225	0.284	−0.432	0.035*
Life expectancy	0.086	0.103	−0.793	0.353
Cancer screening wait time	0.043	0.067	−0.866	0.101

Significance level: * $p < 0.05$.

Robustness checks: logistic and probit models

Table A2.10: Average marginal effects from unadjusted logistic regression models predicting positional choice across domains

	DV: Dummy variable (positional choice)					
	Income	Vacation time	Days of illness	Physical activity	Doctor density†	Hospital bed density†
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment						
Group: Friends	−0.225***	−0.061**	−0.062**	−0.021	−0.056**	−0.028
	(0.028)	(0.023)	(0.019)	(0.026)	(0.019)	(0.022)
Country: US	−0.005	0.005	0.019	0.045	0.038*	0.041
	(0.028)	(0.023)	(0.019)	(0.026)	(0.019)	(0.022)
Controls	N	N	N	N	N	N
Observations	981	981	981	981	981	981

Robust standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Results are average marginal effects (AME) from logistic regression models. † denotes healthcare provision ratios: doctor-to-population and hospital beds-to-population. *Friends* refer to the *friends and acquaintances* reference group.

Table A2.11: Probit model results: preference correlates of positional choice

Variables	DV: Dummy variable (Positional choice)					
	Income	Vacation time	Illness	Physical activity	Doctor density†	Hospital bed density†
	(1)	(2)	(3)	(4)	(5)	(6)
Preferences						
Patience	0.016	−0.039	−0.063*	−0.076**	−0.085**	−0.048
	(0.024)	(0.028)	(0.031)	(0.026)	(0.030)	(0.028)
Risk tolerance	0.019	0.011	0.041	0.040	0.052*	0.078***
	(0.019)	(0.023)	(0.026)	(0.021)	(0.025)	(0.023)
Altruism	−0.031	0.024	0.003	−0.005	−0.009	0.009
	(0.018)	(0.022)	(0.024)	(0.020)	(0.023)	(0.021)
Social comparison	0.123*	0.147*	0.242**	0.110	−0.059	−0.183**
	(0.059)	(0.070)	(0.081)	(0.064)	(0.075)	(0.068)
Cluster (ref: 3: Self-interest)						
1: Inequality aversion	0.246*	0.511***	0.379**	0.191	0.259*	0.349**
	(0.100)	(0.117)	(0.132)	(0.110)	(0.128)	(0.117)
2: Envy	0.490***	0.497***	0.478***	0.546***	0.252	0.413**
	(0.109)	(0.128)	(0.140)	(0.116)	(0.143)	(0.130)
Observations	981	981	981	981	981	981

Standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: † doctor-to-population and hospital beds-to-population. *Patience*, *Risk tolerance*, *Altruism* per Falk *et al.* (2023). *Social Comparison* from Alba *et al.* (2014); Gibbons and Buunk (1999). Clusters via stated-satisfaction as in Diaz *et al.* (2023), see Fallucchi, Luccasen III and Turocy (2019, 2022).

Table A2.12: Probit estimates of the probability of positional choice in economic and health domains

	DV: Dummy variable (Positional choice)												
	Income	Vacation time	Days of illness	Physical activity	Doctor density†	Hospital bed density†	Life expectancy	Air quality	Good child health	Wait time	2OB Income	2OB Life exp.	2OB Wait time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Treatment													
Group: Others	0.659***	0.251*	0.373**	0.064	0.333**	0.100	0.085	0.435**	0.113	0.463**	0.323***	0.186	0.359*
	(0.089)	(0.103)	(0.116)	(0.094)	(0.114)	(0.102)	(0.148)	(0.146)	(0.130)	(0.154)	(0.089)	(0.115)	(0.142)
Country: US	0.020	0.033	0.090	0.200*	0.241*	0.190	0.054	0.174	0.110	0.168	0.188*	0.083	0.220
	(0.089)	(0.103)	(0.115)	(0.094)	(0.114)	(0.103)	(0.149)	(0.141)	(0.131)	(0.148)	(0.089)	(0.116)	(0.140)
Preferences													
Patience	0.008	−0.039	−0.071*	−0.073**	−0.091**	−0.051	−0.071	−0.028	−0.023	−0.022	0.019	−0.044	−0.044
	(0.025)	(0.028)	(0.031)	(0.026)	(0.031)	(0.028)	(0.040)	(0.039)	(0.036)	(0.040)	(0.025)	(0.031)	(0.038)
Risk tolerance	0.015	0.022	0.047	0.047*	0.053*	0.072**	0.089*	0.057	0.062*	0.057	0.015	0.010	0.064*
	(0.020)	(0.024)	(0.027)	(0.022)	(0.026)	(0.023)	(0.035)	(0.033)	(0.030)	(0.034)	(0.020)	(0.026)	(0.032)
Altruism	−0.030	0.010	0.001	−0.013	−0.011	0.011	−0.026	0.025	0.027	−0.037	−0.002	0.043	−0.002
	(0.019)	(0.022)	(0.025)	(0.020)	(0.024)	(0.022)	(0.031)	(0.031)	(0.029)	(0.031)	(0.020)	(0.026)	(0.030)
Social comparison	0.107	0.115	0.223**	0.096	−0.053	−0.183**	0.039	0.318**	−0.034	−0.098	0.136*	−0.063	−0.038
	(0.062)	(0.073)	(0.084)	(0.066)	(0.078)	(0.070)	(0.103)	(0.106)	(0.090)	(0.101)	(0.062)	(0.079)	(0.095)
Cluster (ref: 3: Self-interest)													
1: Inequality aversion	0.237*	0.472***	0.368**	0.179	0.252	0.347**	0.197	0.366*	0.206	0.276	0.149	0.348**	0.518**
	(0.104)	(0.119)	(0.135)	(0.111)	(0.130)	(0.118)	(0.168)	(0.169)	(0.155)	(0.176)	(0.104)	(0.131)	(0.160)
2: Envy	0.433***	0.444***	0.442**	0.556***	0.232	0.416**	0.046	0.525**	0.434**	0.470*	0.124	0.171	0.425*
	(0.113)	(0.131)	(0.144)	(0.118)	(0.147)	(0.133)	(0.200)	(0.174)	(0.162)	(0.183)	(0.116)	(0.153)	(0.182)
Sociodemographics													
Female	0.075	0.339**	0.110	0.123	0.056	0.005	−0.195	0.102	0.296*	0.531***	0.116	0.219	0.131
	(0.093)	(0.107)	(0.120)	(0.098)	(0.118)	(0.107)	(0.158)	(0.147)	(0.137)	(0.161)	(0.093)	(0.120)	(0.144)
Age	−0.010*	−0.002	−0.007	0.001	0.002	−0.002	0.004	−0.001	−0.016*	−0.005	−0.012*	−0.017**	−0.003
	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.008)	(0.008)	(0.007)	(0.008)	(0.005)	(0.006)	(0.007)
Employment (Part-time)	0.042	−0.217	−0.055	−0.274	−0.094	0.223	−0.082	−0.137	−0.077	−0.186	−0.015	−0.204	0.104
	(0.143)	(0.161)	(0.176)	(0.150)	(0.185)	(0.165)	(0.247)	(0.223)	(0.224)	(0.232)	(0.144)	(0.190)	(0.213)
Employment (Full-time)	0.177	−0.109	−0.145	−0.146	0.052	0.218	−0.025	−0.081	0.243	−0.035	0.134	0.080	0.026
	(0.113)	(0.125)	(0.140)	(0.115)	(0.143)	(0.135)	(0.188)	(0.172)	(0.168)	(0.180)	(0.113)	(0.141)	(0.177)

Standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: † denotes healthcare provision ratios: doctor-to-population and hospital beds-to-population. Second-order beliefs (2OB) were measured by asking participants "Do you think the majority of participants preferred Society A or Society B?".

A2.5 Participant instructions and survey items

A2.5.1 Block 1: Positional choice tasks

Positional choice task design (Solnick and Hemenway, 1998). Participants were asked to choose between two hypothetical societies that differ in a single domain-specific attribute, including income, consumption levels, or health- and healthcare-related conditions (see Table A2.13). All other aspects of these societies, including the variety and prices of goods, were held constant. There was no correct answer. Respondents were instructed to indicate the society in which they personally would feel most content, rather than the one they believed to be best for society as a whole.

Table A2.13: Positional choice task items

Domain	Which society would you prefer?
Income	Society A: $4100 vs $4900
Your vs average monthly income after tax	Society B: $3700 vs $3200
Vacation time	Society A: 4 vs 6 weeks
Your vs others' weeks of vacation	Society B: 3 vs 2 weeks
Air quality	Society A: 13 vs 8 days
Your vs others' areas with moderate to high air pollution	Society B: 18 vs 29 days
Days of illness	Society A: 4 vs 2.5 days
Your vs others' days of significant illness per year	Society B: 5 vs 6 days
Life expectancy	Society A: 84 vs 87 years
Your vs others' life expectancy at birth	Society B: 79 vs 68 years
Good child health	Society A: 63 vs 67 years
Your vs others' children's years lived in good health	Society B: 60 vs 55 years
Physical activity	Society A: 125 vs 160 minutes
Your vs others' physical activity per week	Society B: 110 vs 95 minutes
Cancer screening wait time	Society A: 3 vs 1 weeks
Your vs others' waiting time	Society B: 5 vs 7 weeks
Doctor-to-population ratio	Society A: 1 per 200 vs 1 per 140 people
Your vs others' area's doctor availability	Society B: 1 per 250 vs 1 per 340 people
Hospital beds-to-population ratio	Society A: 1 per 160 vs 1 per 125 people
Your vs others' area's hospital bed availability	Society B: 1 per 200 vs 1 per 300 people

Notes: The items listed reflect the version presented to US participants in the *general population* reference group. Values were adapted for UK participants and for the *friends and acquaintances* reference group.

Social comparison and preferences. Participants were asked to indicate how well each of the following statements described them personally. Table A2.14 summarizes the items across two scale types. Items measuring social image concerns and social comparison were rated on a 5-point Likert scale ranging from 1 = *Strongly disagree* to 5 = *Strongly agree*. Items assessing risk tolerance, time preferences (patience), and altruism were rated on an 11-point Likert scale ranging from 0 = *Completely unwilling to do so* to 10 = *Very willing to do so*.

Table A2.14: Social comparison and preference measures

Indicator	Statement
Rated on a 1–5 scale (Strongly disagree to Strongly agree):	
Social image	"It is important to me to impress others."
Social image	"It is important to me how I am perceived by others."
Social comparison	"I always pay a lot of attention to how I do things compared with how others do things."
Social comparison	"I never consider my situation in life relative to that of others."
Social comparison	"I often compare myself on how I am doing in terms of health."
Rated on a 0–10 scale (Completely unwilling to Very willing):	
Risk tolerance	"How willing are you to take risks?"
Patience	"How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?"
Altruism	"How willing are you to give to good causes without expecting anything in return?"

Attention checks. Participants were instructed to select a specific response to confirm they had read the instructions carefully. Failure to follow these instructions resulted in exclusion from payment. Table A2.15 presents the two attention check items, which were displayed at the beginning of Block 1 (Section A2.5.1) and after Block 3 (Section A2.5.3).

Table A2.15: Attention check questions

Question	Options and expected answer
(AC1) What color is the sky on a clear day?	Options: Blue | Grey | Purple | Yellow | Orange Expected answer: *Purple*
(AC2) What color is a stop sign?	Options: Red | Yellow | Orange | White | Green Expected answer: *Green*

A2.5.2 Block 2: Slider tasks

Indifference thresholds via slider tasks. Participants who selected the positional option in the income domain were presented with an ordinal slider to indicate the minimum acceptable income at which they would be willing to accept Society A (where they would be better off in absolute terms but relatively worse off compared to the given reference group) over Society B (where they would be worse off in absolute terms but better off in relative terms). Figure A2.2 illustrates the task administered to US participants assigned to the *general population* condition. Similarly, participants who expressed positional concerns regarding life expectancy were asked to indicate their indifference threshold in years, specifying the point at which they would switch to preferring Society A.

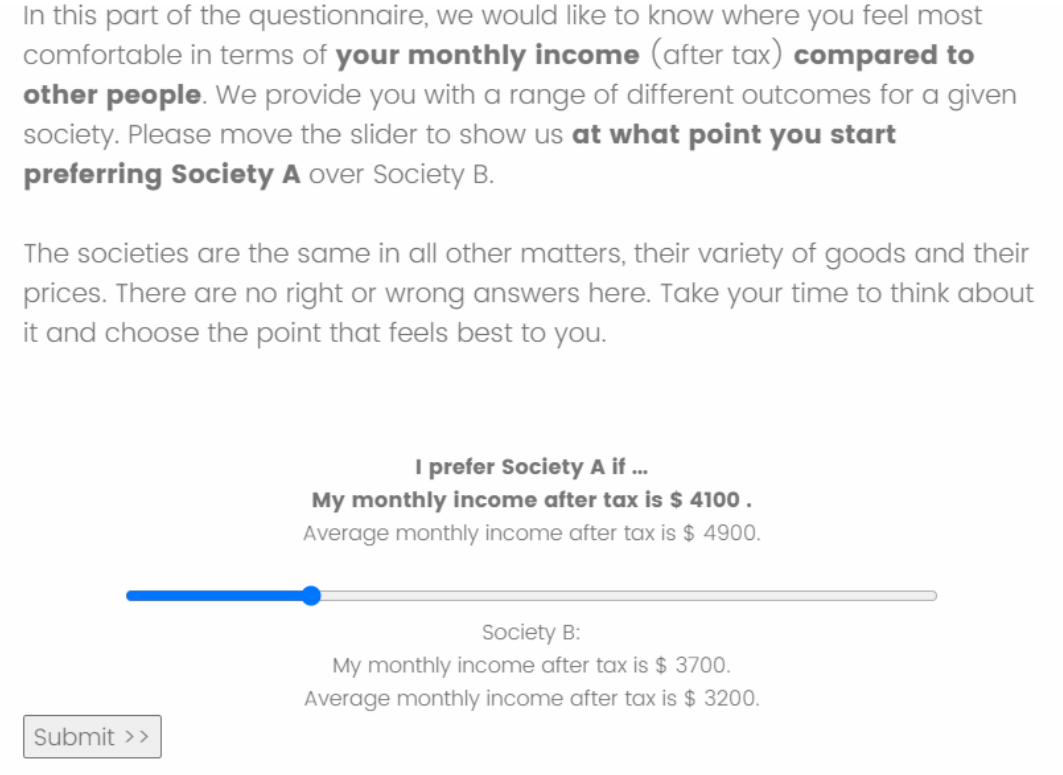


Figure A2.2: Screenshot of slider task measuring US participants' indifference threshold for income (reference: general population)

A2.5.3 Block 3: Incentivized belief elicitation

Incentivized second-order beliefs. Participants were asked to predict which of two hypothetical societies the majority of respondents preferred in three separate scenarios (see Table A2.16). Preference was defined as the society in which the average participant would feel most content. One of the three scenarios was randomly selected for bonus payment purposes (£0.50 bonus if their prediction matched the actual majority choice). Participants were instructed to focus only on perceived majority choice, not their own preference.

Table A2.16: Second-order belief items

Domain	Which society did the majority of participants prefer?
Income	Society A: \$4100 vs \$4900
Participants' vs average monthly income after tax	Society B: \$3700 vs \$3200
Life expectancy	Society A: 85 vs 90 years
Participants' vs others' life expectancy at birth	Society B: 80 vs 75 years
Cancer screening wait time	Society A: 3 vs 1 weeks
Participants' vs others' waiting time for screening	Society B: 5 vs 7 weeks

Notes: The items listed reflect the version presented to US participants in the *general population* reference group. Values were adapted for UK participants and for the *friends and acquaintances* reference group.

A2.5.4 Block 4: Stated satisfaction

Stated satisfaction approach (Diaz *et al.*, 2023). Participants were asked to evaluate how satisfied they would feel with various income allocations. In each scenario, their own payment was fixed at \$20, while the payment to another (unspecified) person varied from \$10 to \$30, in increments of \$2.50.[10] This task was identical across countries, with only currency formatting adjusted. Each scenario was rated on a 7-point Likert scale ranging from 1 = *Extremely dissatisfied* to 7 = *Extremely satisfied*.

[10]Due to a typographical error, one allocation presented \$27.7 instead of \$27.5. Given the small numerical difference, this is not expected to affect stated satisfaction.

A2.5.5 Block 5: Demographic questions

Table A2.17: Demographic questions

Question item	Options and coding	
How old are you?	• Under 18 (1) • 18–24 years old (2) • 25–34 years old (3) • 35–44 years old (4)	• 45–54 years old (5) • 55–64 years old (6) • 65+ years old (7)
How do you describe yourself?	• Male (1) • Female (2) • Non-binary/third gender (3)	• Prefer to self-describe(4) • Prefer not to say (5)
How would you rate your overall health?	• Poor (1) • Fair (2) • Good (3)	• Very good (4) • Excellent (5)
What is the highest level of education you have completed?	• Some Primary (1) • Completed Primary School (2) • Some Secondary (3) • Completed Secondary School (4) • Vocational or similar (5)	• Some university but no degree (6) • University Bachelors Degree (7) • Graduate/professional degree (8) • Prefer not to say (9)
What describes your employment status (last three months)?	• Working full-time (1) • Working part-time (2) • Unemployed/looking for work (3) • Homemaker/stay-at-home parent (4)	• Student (5) • Retired (6) • Other (7)
[US] What was your total household income before taxes (past 12 months)?	• Less than $25,000 (1) • $25,000–49,999 (2) • $50,000–74,999 (3) • $75,000–99,999 (4)	• $100,000–149,999 (5) • More than $150,000 (6) • Prefer not to say (7)
[UK] What was your total household income before taxes (past 12 months)?	• Less than £20,000 (1) • £20,000–39,999 (2) • £40,000–59,999 (3)	• £60,000–99,999 (4) • More than £100,000 (5) • Prefer not to say (6)
How does your household manage its monthly income to cover all living expenses?	• Very poorly (1) • Fairly poorly (2) • Somewhat poorly (3) • Somewhat well (4)	• Fairly well (5) • Very well (6) • I don't know (7) • Prefer not to answer (8)
Political attitude from extremely liberal (left) to extremely conservative (right)	• Extremely liberal (1) • Moderately liberal (2) • Slightly liberal (3) • Neutral (4)	• Slightly conservative (5) • Moderately conservative (6) • Extremely conservative (7)

The hidden struggle: navigating disclosure, social backlash, and incentives – experimental evidence from adults with ADHD

The hidden struggle: Navigating disclosure, social backlash, and incentives – experimental evidence from adults with ADHD

3.1 Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a clinically heterogeneous neurodevelopmental condition that, once seen primarily as a childhood and adolescent disorder, is now widely recognized as a condition that often persists into adulthood. Although reported incidence rates have fluctuated in recent years, especially so during the COVID-19 pandemic, the evidence is mixed; some studies find no clear indication of an actual increase in prevalence (Martin *et al.*, 2025), which is currently estimated to range from 2.5% to 10% globally (Ayano *et al.*, 2023; Song *et al.*, 2021). What is undisputed, however, is that ADHD has moved steadily to the forefront of public and academic attention. Much of this debate turns on three points: the apparent gap between epidemiological prevalence and clinical diagnosis rates (Banaschewski *et al.*, 2024); persistent diagnostic disparities among underrepresented groups, including women and ethnic minorities (Abdelnour, Jansen and Gold, 2022); and the uneasy coexistence of rising public awareness with enduring stigma. These discussions have increasingly permeated both traditional and social media, contributing to a greater public sensitivity toward the experiences of individuals affected by ADHD, and, as a result, a growing recognition of neurodivergence and neurodiversity (Martin *et al.*, 2025). It is against this backdrop that this chapter presents two linked experimental studies. The first focuses on neurodivergent individuals with ADHD, who must decide whether to disclose their condition in a strategic context. The second follows up with non-neurodivergent adults, who evaluate those disclosures and make corresponding interaction choices[1].

Even though awareness of neurodivergence is growing, disclosing conditions like ADHD can still pose significant challenges. Unlike race, gender, or physical disabilities, ADHD is often not immediately visible and constitutes a *concealable stigmatized identity* (Quinn and Earnshaw, 2011), which individuals may choose to reveal or conceal. This choice, however, often has significant implications for the individual affected, across educational, occupational, psychological, and social dimensions. Disclosure can lead to understanding, accommodations, and support, yet it also exposes individuals

[1]Some treatment conditions in the present study refer to the term "neurodivergence" instead of "ADHD" for framing purposes. The variation in language is deliberate and reflects the experimental manipulation of framing rather than a difference in diagnostic status.

to potential judgment, discrimination, or rejection (Brohan *et al.*, 2012; Donnelly, 2017). Although some studies suggest that disclosure has even been shown to reduce negative social perceptions (Jastrowski *et al.*, 2007; Thompson-Hodgetts *et al.*, 2020), many individuals remain "in the closet" due to the aforementioned risk, even if doing so means forgoing access to support, for example, through workplace accommodations such as flexible hours and mentorship opportunities. Thereby, especially in environments where image concerns and perceptions of fairness are high (e.g., in schools, universities, and workplaces), people often choose to keep certain aspects of themselves hidden to avoid any risks (Benndorf, Kübler and Normann, 2015).

Studies on the disclosure of mental health conditions, and ADHD in particular, suggest that such decisions involve a balancing act of anticipated reactions, fear of stigma, perceived social norms, legal rights, and trust (Greene *et al.*, 2012; McGrath *et al.*, 2023; Murphy and Latham, 2022). Thus, individuals often avoid disclosing neurodivergent conditions like ADHD due to fear of discrimination (Morris, Begel and Wiedermann, 2015), concerns about being perceived as incompetent (Gupta and Priyadarshi, 2020; Santuzzi *et al.*, 2014), or as exploiting the system (Godard, Hebl and Nittrouer, 2022). As a result, they often leave accommodations untapped even when they are available. Employee Assistance Programs (EAPs), for example, offer support for individuals with a broad spectrum of personal issues, including mental health and neurodevelopmental disorders. However, although many employers offer initiatives like EAPs to promote the well-being of their employees, overall uptake rates have been shown to remain low (Attridge, 2019; Mio, 2023).

Although many studies have examined this disclosure dilemma using qualitative methods and surveys, experimental research with incentive-compatible designs remains limited. Earlier studies on information disclosure, such as Benndorf, Kübler and Normann (2015) and Shulman *et al.* (2024), often rely on vignettes, hypothetical scenarios, or self-reports. Although widely used in economic and behavioral research, these methods are limited in their ability to represent how people actually act when the decisions they face have real social or economic consequences, and may be susceptible to social desirability or hypothetical bias (Haghani *et al.*, 2021). As a result of these biases, little is known about how anticipated social consequences, contextual framing, and individual characteristics interact to affect disclosure behaviors, and this may hinder the design of effective interventions for settings where protecting one's social image is a major concern.

Disclosure decisions are rarely binary: they depend on when, where, and to whom one discloses potentially stigmatizing information. While previous empirical studies have addressed the barriers to disclosure, and proposed institutional solutions to support neurodiverse populations, there remains a

need to better understand the behavioral drivers of disclosure decisions under conditions of strategic uncertainty and asymmetric information (Benndorf, Kübler and Normann, 2015). This becomes apparent through evidence suggesting that even theoretically incentive-compatible mechanisms often fail to elicit truthful reporting of beliefs, for instance due to risk preferences, loss aversion, or bounded rationality (Danz, Vesterlund and Wilson, 2024). Since image concerns and the risk of rejection can shape individual behavior in such environments, disclosure choices are likely further influenced by the framing of personal information, beliefs about others' perceptions and reactions (Bénabou and Tirole, 2006), internalized norms, and how individuals perceive their traits relative to normative expectations (Bicchieri, 2005; Grossman and Van Der Weele, 2017).

Theoretical framework. Based on these deliberations, this study builds on economic theories of signaling, social image concerns, and strategic disclosure. In settings characterized by asymmetric information, where identities or characteristics are concealable, disclosure can serve as a costly signal of credibility or need, particularly when it is linked to access to resources such as workplace accommodations. In such contexts, disclosure (the signal) can become a strategic act for individuals in order to influence beliefs and the outcomes they receive[2]. If, however, individuals anticipate negative judgments, they may be inclined to withhold certain information, especially when social approval or economic outcomes are at stake (Bénabou and Tirole, 2006). In this regard, information framing itself can influence how costly or credible the disclosure appears, while anticipated social reactions further manipulate the (perceived) cost of disclosure.

Social image theory similarly suggests that people care deeply about how others perceive them[3]. In environments where stigma or reputational loss is expected, behavior may therefore be influenced by the objective to maintain a favorable image in the eyes of their peers or the public, leading individuals to hide potentially reputation-damaging information, and thus, to under-disclose. In such settings, where individuals weigh whether, what, when, and to whom to disclose, framing effects and

[2]See, for example, Spence (1973) and Crawford and Sobel (1982) for their early contributions to strategic communication between two agents under information asymmetry. While Spence (1973) shows how job seekers use observable traits to signal their abilities to employers, where the cost of the signal ensures its credibility, Crawford and Sobel (1982) considers settings where private information is conveyed without any cost and through non-binding messages – which later laid the foundational framework of *cheap talk* (Farrell and Rabin, 1996).

[3]Note that image concerns also include self-image, i.e., how individuals perceive themselves by adopting the role of an internal observer (Grossman, 2015). People may act in certain ways to feel good about themselves. Grossman and Van Der Weele (2017) extend this idea, showing experimentally how individuals use willful ignorance to preserve a positive self-image. In the present study, though, this mechanism appears less likely. The primary study population consists of individuals who had already self-identified as having ADHD on the recruitment platform. While their self-concept may still be influenced by the experimental framing cues, I consider any such effect to be minimal.

second-order beliefs about how others perceive one's attributes may substantially influence behavior (Acquisti, John and Loewenstein, 2013).

To empirically test how these signaling mechanisms unfold in practice, particularly in settings marked by asymmetric information and reputational risk, the present study implemented a two-part experimental design. By manipulating both the framing of neurodivergent traits (diagnostic vs. identity-based) and the potential for social backlash (no rejection vs. rejection with or without cost), the design tests whether and when individuals are willing to send costly signals to access accommodations, and how such signals may be interpreted. The time advantage offered for disclosure simulates a realistic performance-based accommodation, mirroring settings where disclosure grants access to a resource or support that would otherwise be unavailable.

The study design also draws on additional theoretical models that inform how disclosure signals are received, by both the discloser and receiver. Models of social preferences, such as those developed by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), suggest that individuals are sensitive to inequality and fairness, which can influence both the decision to disclose and how disclosures are judged by others. On the one hand, disclosers might forgo accommodations out of guilt if they took advantage of the performance boost. Receivers, on the other hand, may form judgments about disclosers based on the perceived legitimacy of disclosed traits, if they were to learn about the access to an accommodation. To isolate the latter mechanism, receivers in Study 2 were not informed about the availability of accommodations to their matched partner, thereby allowing the study to isolate reactions to the disclosure itself. In addition to examining the decision to disclose, the study also investigates how such disclosures are received and interpreted, capturing both sides of the interaction. Lastly, belief-based and Bayesian signaling models (e.g., Kamenica and Gentzkow, 2011; Konow, 2000; Wu, 2019) imply that receivers update their beliefs based on the structure and framing of disclosure messages. In these models, senders strategically choose how to communicate information in anticipation of how receivers will interpret and respond to it. The specific framing thus shapes posterior beliefs about the sender's type, for instance in terms of ability or competence. Accordingly, my manipulation of message framing allows testing how different framings can affect how the discloser's abilities are perceived. In conclusion, these considerations raise a central research question guiding the design and implementation of the present study: How do neurodivergent individuals make strategic disclosure decisions when potential economic benefits are weighed against risks of stigma and rejection, particularly under different framings of identity and varying levels of perceived backlash?

Study overview. This chapter presents two preregistered online experiments examining disclosure decisions involving private health information, focusing on a sensitive, concealable trait such as ADHD. While the experiments take place in an online and anonymous setting, the design incorporates real social and economic stakes, including the possibility of social backlash, consequential choices, and performance-based financial incentives, to simulate the trade-off between access to performance-enhancing accommodations and the potential social costs, even in the absence of face-to-face interaction.[4] Study 1 examines disclosure behavior under varying backlash risks and framings, whereas Study 2 follows up with a separate, non-neurodivergent sample to examine how those disclosures are received. This two-study design allows for an analysis of both sides of the interaction, without having to rely on hypothetical responses. For clarity, participants in Study 1, the *disclosers*, are referred to as Player 1 (P1) hereafter, while those in Study 2, the *receivers*, are referred to as Player 2 (P2).

In Study 1, 671 participants were recruited, of whom 659 met the inclusion criteria and were included in the analysis. In Study 2, 750 participants took part, with 749 forming the final sample. In both studies, participants were randomly assigned to one of six treatment arms in a 3×2 between-subjects design. The treatments varied (1) how ADHD was framed (clinical diagnosis vs. identity-focused) and (2) the potential for and consequence of social rejection. Given that the primary research questions concern disclosure decisions by neurodivergent individuals, this chapter focuses on the analysis of Study 1. The full designs of Study 1 and Study 2 are reported in Section 3.2 and 3.3, respectively. Participants first completed three rounds each of two cognitive tasks commonly used to assess impulse control and working memory (Go/No-Go and N-back). They were then informed that they would be paired with another participant. Depending on the condition, participants were also told that their partner might have the opportunity to reject the match. Finally, they were told either that they had "symptoms that are commonly associated with a clinical ADHD diagnosis", or "fall within a category reflecting neurodivergent cognitive traits, often described as neurodivergence". They could then choose to disclose this information to their matched partner in exchange for a time advantage, which may help them boost their performance and increase their chances of earning a higher bonus.[5]

Based on the theoretical framework and the experimental setup outlined above, this study formulates

[4]However, the online context still imposes limitations that may affect the external validity of observed behavior. Refer to Section 3.7.1 for a discussion of these constraints and their implications.

[5]Note that in Study 2, the receivers of said information were not informed of the performance-based accommodation linked to disclosure in Study 1. This asymmetry is a deliberate design choice and allowed me to isolate reactions due to stigma, independent of fairness or envy over resource allocation.

the following preregistered main hypotheses: First, I predict that participants exposed to the most consequential backlash condition (i.e., social rejection and forfeiture of bonus) are less likely to disclose their neurodivergent traits, including ADHD, than those in the Control condition, who do not face any risk of rejection. Second, I hypothesize that individuals exposed to a diagnostic framing are expected to disclose less frequently than those receiving an identity-based framing. Third, I expect the lowest disclosure rates among participants who received both the High Backlash condition and diagnostic framing. This chapter focuses on the preregistered main effects of backlash and framing manipulations on disclosure decisions, but also incorporates exploratory analyses of potential moderators such as fairness perceptions, and second-order beliefs.

The chapter is organized as follows: Sections 3.2 and 3.3 describe the two-study design of the present experiment. Section 3.4 presents the main findings. Section 3.7 revisits the findings and limitations of the study.

3.1.1 Experimental hypotheses

Hypotheses **H1** through **H3** were formulated prior to data collection and documented in the preregistration protocol for Study 1.

H1a) Participants in the High Backlash condition will be less likely to disclose their neurodivergence (by opting for extra time) than participants in the Control condition.

H1b) Participants in the High Backlash condition will be less likely to disclose their neurodivergence than participants in the Low Backlash condition.

H1c) Participants in the Low Backlash condition will be less likely to disclose their neurodivergence than participants in the Control condition.

H2) Participants exposed to the diagnostic framing ("symptoms (...) associated with a clinical ADHD diagnosis") will be less likely to disclose than those presented with the identity-based framing ("neurodivergent cognitive traits").

H3) Participants in the High Backlash condition who receive the diagnostic framing will exhibit the lowest disclosure rates of all.

Secondary research objectives. In addition to the preregistered hypotheses, the present study also aimed to examine whether participants' disclosure rates varied with other pre- and post-treatment outcomes. While the experimental design captures the consequences of anticipated rejection and financial loss, I was also interested in exploring how individual characteristics, including preferences and beliefs, impact disclosure decisions.

To complement the primary study, a follow-up experiment that focused on Player 2's reactions to Player 1's disclosure was also preregistered. The following hypotheses were developed for this as part of Study 2's preregistration protocol:

H4a) Participants who view a neurodivergence-framed (either diagnostic or identity-based) profile of their assigned partner will be less likely to accept the match than those who receive no health-related information.

H4b) Participants exposed to a diagnostic framing ("symptoms associated with a clinical ADHD diagnosis") will be less likely to accept the match than those who receive an identity-based message ("neurodivergent cognitive traits").

H4c) Participants who receive a diagnostic message will exhibit the lowest acceptance rates overall.

3.2 Study 1: Disclosure decisions (Player 1)

Study 1 is the first study of two connected experiments, designed to investigate how adults with self-reported ADHD make strategic disclosure decisions in a performance setting where disclosure grants an accommodation but may also invite social backlash. Participants face varying levels of rejection risk and receive different framings of their condition, allowing for an experimental test of signaling behavior under asymmetric information. As outlined in Section 3.1.1, I pose three hypotheses concerning how disclosure rates change with backlash risk, message framing, and the interaction between the two.

3.2.1 Sample and preregistration details

The full study design, hypotheses, treatment conditions, outcome measures, and analysis plan for Study 1 were preregistered on May 13, 2025, in the Open Science Framework registry (`https://doi.org/10.17605/OSF.IO/TWYJN`).

Data collection took place in June 2025. A total of 671 individuals participated in the study, recruited through the online survey platform Prolific. I excluded 8 people whose disclosure decision was missing, as well as any duplicate entries, resulting in a final sample of 659 participants. Eligible participants were U.S.-based adults (ages 18–65) who self-reported ADHD according to Prolific's audience screener.

As preregistered, the target sample size was 500 participants, which is more than the minimum suggested by an a priori power analysis ($N = 334$) for a significance level of $\alpha = 0.05$, and 95% power. The analysis assumed a small-to-medium effect size ($f^2 = 0.08$), informed by prior research on norm-directed behavior and message framing (Bicchieri and Xiao, 2009; Krupka and Weber, 2013; Nyhan and Reifler, 2019; McQueen *et al.*, 2011).

Compensation. All participants received a base payment of £3 (≈ $4). The maximum bonus they could earn throughout the experiment was £1.50 (≈ $2), determined by task performance and the final matching outcome. Detailed instructions appear in the Appendix.

Ethical approval and consent. The study received ethical approval from the LISER Research Ethics Committee under reference number [LISER REC/2024/127.HealthDisclosure/1]. Before accessing the study on Prolific, participants were informed that it would cover sensitive health topics. The consent form explained that participation was voluntary and anonymous, detailed their rights, and made clear that they could withdraw at any time without penalty.

3.2.2 Experimental design and treatment conditions

Study 1 uses a fully randomized 3×2 between-subjects design to examine how adults with ADHD make disclosure decisions when facing varying levels of anticipated social backlash. Participants were randomly assigned to one of three backlash conditions (Control, Low Backlash, High Backlash), which differed in the likelihood and consequences of rejection, and to one of two framings of their condition (diagnostic vs. identity-based). The experiment was implemented online via Qualtrics. Table 3.1 briefly summarizes the six resulting experimental arms.

Table 3.1: Treatment conditions for Player 1: backlash context and rejection consequences

Backlash	Framing	Rejection consequence (if $\text{Reject}_i = 1$)
Control	Diagnostic or identity-based	No rejection possible; remains matched.
Low Backlash	Diagnostic or identity-based	20%: both reassigned; no bonus loss.
High Backlash	Diagnostic or identity-based	20%: P2 reassigned; P1 forfeits bonus.

Note: $\text{Reject}_i = 1$ if the matched partner (Player 2), to whom information could have been revealed, chooses to reject Player 1, requesting reassignment to another study participant.

Backlash factor (3 levels). The backlash manipulation varied the potential consequences of disclosure. In the Low and High Backlash conditions, participants were informed that they would be paired with another participant (Player 2), who might have the opportunity to reject the match. The three levels of backlash were as follows:

- **Control:** No rejection risk. Participants remained matched throughout the experiment regardless of their disclosure decision.
- **Low Backlash:** Participants were informed that their matched partner could reject the match, with a 20% probability of implementation. If rejection were implemented, both participants were reassigned to new partners without any loss in earned bonuses.
- **High Backlash:** Participants were informed of the same rejection mechanism, but with higher stakes: if rejection was implemented (20% probability), only Player 2 was reassigned, while Player 1 forfeited their earned bonus.

Message framing factor (2 levels). The message framing manipulation varied how participants' neurodivergent traits (ADHD) were described in the disclosure signal:

- **Diagnostic framing:** Participants' characteristics were described as "symptoms associated with a clinical ADHD diagnosis".
- **Identity-based framing:** Participants' characteristics were described as "neurodivergent cognitive traits".

The inclusion and design of these framing conditions are informed by empirical evidence that labels can shape social responses independently of observable behavior. A recent systematic review by O'Connor *et al.* (2022) found that, in the case of ADHD, diagnostic labels can exacerbate negative perceptions, with the consequence being stronger stereotyping and reduced willingness to engage. Conversely, reduced stigma and greater acceptance could arise if ADHD is not just thought of as a medical condition but as part of a person's identity, as has been suggested in previous research (Grummt, 2024; Porras Pyland *et al.*, 2025). Based on this, I hypothesize that the label not only influences how receivers interpret the disclosure, but also whether people with ADHD choose to disclose in the first place, given that they may anticipate certain social responses.

Thus, before making their disclosure decision, each participant was shown a preview of the information that would be shared with a matched partner if they chose to disclose. The assigned message was shown to the matched participant only in cases where disclosure occurred, in exchange for a performance-boosting time advantage on the final bonus task. Although participants were not informed of their partner's identity, they were told that, if they opted to disclose, the information (including their neurodivergent status) would be visible to their match. If they chose not to disclose, no personal information was revealed.

Implementation probability. In the Low and High Backlash treatments, Player 2 could choose to reject the match with their assigned partner. However, this rejection was implemented in only 20% of cases; in the remaining 80%, the pair remained matched regardless of Player 2's choice. This probabilistic element was added for two reasons: first, it ensured that the risk of rejection felt real and consequential without becoming too predictable. Second, to maintain realistic beliefs and prevent participants from assuming that disclosure would always lead to rejection, which could have disproportionately influenced risk-averse Player 1s or social-desirability-driven Player 2s. Similar designs have been used in economic experiments on discrimination (Ridley, 2022), social signaling (Grossman, 2015), and moral decision-making (Dana, Weber and Kuang, 2007) to introduce uncertainty or ambiguity, while still maintaining the credibility of consequences and mitigating overly strategic behavior.

Matching procedure. After all participants in Study 1 had completed the experiment, including the main tasks and their disclosure decisions, a separate sample of non-neurodivergent adults (Player 2) was recruited via Prolific (see Section 3.3). Participants from Study 1 were then post hoc matched with individuals from this new sample using pre-generated, randomly distributed Qualtrics links

that embedded both Player 1's original treatment condition and their disclosure choice.

While the behavior of the neurodivergent sample (Player 1) was of primary interest, including actual participants as Player 2s ensured that the potential of rejection was credible rather than posing only a hypothetical risk to those acting as Player 1. Player 2's role was to view the disclosed information (if any) and decide whether to remain matched or request reassignment. Their decision had implications not only for their own outcome but also for their matched partner's outcome.

3.2.3 Procedure

After providing informed consent, participants completed demographic and health questionnaires, including the ASRS-v1.1 (ADHD), PHQ-2 (depression), and GAD-2 (anxiety) screeners (Kessler *et al.*, 2005; Kroenke, Spitzer and Williams, 2003; Kroenke *et al.*, 2007).

They then completed three rounds each of a Go/No-Go task and an N-back task (1-, 2-, and 3-back). The order in which participants completed the two task blocks was randomized to control for order effects. Performance on these tasks was used to compute the Balanced Integration Score (BIS), which combines standardized accuracy and reaction time (Liesefeld and Janczyk, 2019). The top 5% of participants received a $0.50 performance bonus.

Participants received information about their assigned backlash condition immediately prior to the disclosure decision stage. This ensured that the potential social consequences of disclosure were salient at the time of decision-making.

Next, participants were shown their assigned message framing and asked whether they would accept extra time on a final bonus task in exchange for disclosing their neurodivergent status to another participant. To ensure that they understood the implications of the matching and bonus structure, participants had to complete a brief comprehension check after making their disclosure decision. Participants were randomly assigned to one of two comprehension checks. If they failed to answer correctly, they were provided with corrective feedback and could only continue to the next screen once they answered correctly.

The bonus task that followed was either a Go/No-Go or 2-back task, structurally identical to the ones they completed earlier. Participants who chose to disclose received an extended, potentially performance-boosting version of the task; those who did not disclose completed the standard version.

Performance in the final bonus task was again evaluated using the BIS. Participants could earn

an additional bonus ranging from \$0.40 to \$1.50, depending on their task performance and final matching outcome. In the High Backlash condition, if participants were rejected by their matched partner and the rejection was implemented, the bonus was forfeited entirely (\$0).

The study concluded with post-task measures assessing distributional preferences (via stated satisfaction), fairness perceptions, and second-order beliefs about stigma and discrimination.

Data quality and comprehension checks. All participants included in the final sample ($N = 659$) completed the experiment and submitted a valid Prolific ID. Participants were not required to pass any formal attention check, but they were flagged for one or more of the following criteria: (1) scoring below four on the ASRS-v1.1 screener[6], (2) extreme performance patterns on cognitive tasks suggestive of inattention or reduced engagement, or (3) answering their assigned comprehension question incorrectly on the first attempt[7].

Although all participants were pre-screened for ADHD, approximately two-thirds ($N = 436$) scored below the screening threshold on the ASRS. In addition, 333 participants (50.5%) failed their assigned comprehension question on the first attempt. Each participant was randomly assigned to one of two comprehension questions, administered immediately after their disclosure decision. Participants who failed received corrective feedback and could only proceed after correcting their response. While the comprehension check assignment did not differ across conditions (backlash: $p = 0.891$; message framing: $p = 0.414$), comprehension pass rates were found to vary significantly across backlash conditions ($p = 0.004$): 58.4% of participants in the Low Backlash condition passed their comprehension check, compared to 46.5% in the Control and 43.5% in the High Backlash conditions[8].

As preregistered, all participants were included in the main analyses irrespective of task performance or passing the comprehension checks on the first attempt. To account for potential selection bias stemming from confounding differences in participant quality, I constructed inverse probability weights (IPW) using pre-treatment covariates and treatment dummies to predict inclusion in a higher-quality subsample (defined by performance and comprehension flags[9]). The resulting stabilized

[6]The ASRS-v1.1 is a self-report screening tool for adult ADHD, developed in collaboration with the WHO (Kessler *et al.*, 2005). In this study, I use its 6-item subset as an additional control to assess whether pre-screened participants would still meet the screening threshold. A score of four or above is generally considered indicative of symptoms consistent with ADHD.

[7]See Table A3.1 in the Appendix for details on flagged observations.

[8]Additional chi-squared (χ^2) and t-tests revealed no significant differences in comprehension performance by message framing ($p = 0.414$), ADHD status ($p = 0.544$), or Go/No-Go BIS ($p = 0.633$).

[9]Detailed in Appendix Table A3.1 and Section A3.1.2.

and truncated weights were then applied to the full sample, allowing me to reduce the impact of responses of lower quality while preserving statistical power.

3.2.4 Outcome measures

The outcome measures presented here provide the empirical basis for the main and secondary analyses presented in Section 3.2.6.

The primary outcome is a binary indicator of disclosure, defined by participants' decision to reveal sensitive health information in exchange for a time advantage in the final bonus task. Participants could either accept the accommodation (disclose = 1) or continue without additional time (disclose = 0). This choice determined whether their matched partner received a disclosure message and which version of the bonus task (extended vs. standard) was presented.

Secondary outcomes include the following:

- **Task performance:** Participants' performance in the Go/No-Go and N-back tasks was assessed using accuracy, average reaction time (RT), number of missed trials, false alarms (in the Go/No-Go tasks only), and incorrect responses (in the N-back tasks only). To create a single index for performance, I computed the *Balanced Integration Score* (BIS) (Liesefeld and Janczyk, 2019) by standardizing accuracy and RT (across participants and within each round), and then subtracting standardized RT from standardized accuracy. For each participant, BIS values were first computed separately for the three rounds of each task type and then averaged to yield an overall Go/No-Go BIS and N-back BIS. Higher BIS values indicate better performance, and reflect both greater accuracy and faster reaction time. Participants in the top 5% based on BIS across initial tasks were eligible for a $0.50 bonus. Performance in the final bonus task (completed with or without time advantage) determined eligibility for an additional performance-based bonus ranging from $0.40 to $1.50. In the High Backlash condition, this bonus could be forfeited if the participant was rejected by their matched partner and rejection was implemented.

- **Social preferences:** Participants' aversion to inequality was elicited via the *stated satisfaction* approach adapted from Diaz *et al.* (2023). All participants were prompted to rate how satisfied they would feel under hypothetical endowment distributions in which their own payoff was fixed at $20, and the partner's payoff varied from $10 to $30. Ratings were recorded on a 7-point

Likert scale (1 = "Extremely dissatisfied", 7 = "Extremely satisfied"). This approach allows me to examine whether and how disclosure behavior may be shaped by individual differences in inequity concerns (Fehr and Schmidt, 1999). Additionally, individual ratings were later used to derive certain social preference profiles via K-means clustering (see Appendix A3.4 for details).

- **Fairness perceptions:** Participants rated how fair they found it for a matched partner to receive a performance advantage due to (a) a health condition or (b) based on luck. Responses were recorded on a 4-point Likert scale (1 = "Completely fair", 4 = "Very unfair"). Fairness beliefs are a central component of decision-making and decision-makers' utility in allocation models (Konow, 2000).

- **Second-order beliefs:** To assess perceived stigma, participants were randomly assigned to one of two versions of a vignette task asking them to rate how they believed (a) most other participants or (b) the general population would perceive a person with each of the following traits: physical disability, low-income background, a history of mental health challenges, or neurodivergence (e.g., ADHD, autism). Perception ratings were recorded on a 5-point Likert scale (1 = "Very negative", 5 = "Very positive").

The final set of controls includes age, gender, employment status, symptoms of anxiety and depression, ADHD symptom score (ASRS-v1.1), and Go/No-Go BIS as a measure of inhibition and impulsivity. Adjusting for these potential confounds follows best practices in experimental economics to account for selection and context effects (Levitt and List, 2007). While I also collected participant data on ethnicity, education, general health, and stimulant use on the day of the online experiment, including these had little effect on the results and model fit. Thus, they were omitted to keep the specification parsimonious (Appendix Table A3.3).

3.2.5 Descriptive statistics

Table 3.2 summarizes the baseline characteristics across the three experimental conditions: Control, Low Backlash, and High Backlash. These include demographic measures (age, gender, education, employment, ethnicity), self-reported health indicators (general health, PHQ-2 and GAD-2 scores, ADHD symptoms, stimulant use), as well as baseline task performance from the initial Go/No-Go and N-back rounds. A binary indicator for same-day stimulant intake (e.g., caffeine, prescription

Table 3.2: Summary of baseline characteristics by condition

Variable	Control	Low Backlash	High Backlash	p-value
n	215	221	223	
Demographic characteristics				
Age	35.74 (10.42)	36.84 (10.91)	36.26 (10.38)	0.759
Female (%)	107 (49.8)	111 (50.2)	129 (57.8)	0.483
Ethnicity (%)				0.759
Black	56 (26.0)	50 (22.6)	51 (22.9)	
Hispanic/Latino	10 (4.7)	9 (4.1)	8 (3.6)	
White	140 (65.1)	154 (69.7)	160 (71.7)	
Other	9 (4.2)	8 (3.6)	4 (1.8)	
Education level (%)				0.759
High school or less	18 (8.4)	19 (8.6)	17 (7.6)	
Some college	33 (15.3)	39 (17.6)	36 (16.1)	
Bachelor's degree	76 (35.3)	59 (26.7)	76 (34.1)	
Postgraduate	88 (40.9)	104 (47.1)	94 (42.2)	
Employment status (%)				0.483
Full-time	133 (61.9)	133 (60.2)	128 (57.4)	
Part-time	56 (26.0)	59 (26.7)	77 (34.5)	
Unemployed/Other	26 (12.1)	29 (13.1)	18 (8.1)	
(Mental) Health and ADHD				
General health	2.86 (0.92)	2.76 (0.89)	2.77 (0.92)	0.454
Depression (PHQ-2 score)	2.14 (1.56)	2.14 (1.63)	1.85 (1.42)	0.420
Anxiety (GAD-2 score)	2.32 (1.71)	2.37 (1.77)	2.12 (1.61)	0.594
ADHD (ASRS-v1.1 score)	2.57 (1.91)	2.57 (1.92)	2.43 (1.83)	0.759
ADHD (Score ≥ 4) (%)	71 (33.0)	72 (32.3)	81 (36.7)	0.759
Stimulant use (%)				0.594
Prescription medication	17 (7.9)	13 (5.9)	10 (4.5)	
Caffeine/stimulants	33 (15.3)	35 (15.8)	47 (21.1)	
None	165 (76.7)	173 (78.3)	166 (74.4)	
Performance				
Go/No-Go BIS	0.06 (0.82)	-0.01 (0.87)	-0.04 (0.75)	0.427
N-back BIS	-0.04 (1.47)	0.00 (1.37)	0.03 (1.48)	0.891
Total BIS	0.03 (1.00)	0.00 (0.88)	-0.02 (0.92)	0.842

Notes: Continuous variables are reported as mean (SD). Categorical variables are reported as n (%). All p-values were corrected for multiple comparisons using the Benjamini-Hochberg method. The Balanced Integration Score (BIS) combines standardized accuracy and reaction time across Go/No-Go and N-back tasks following Liesefeld and Janczyk (2019).

medication) is also included. In order to enhance interpretability, categorical variables such as gender, ethnicity, education and employment status have been collapsed into the most frequent categories. What becomes quickly apparent is that while all participants were initially pre-screened for ADHD, only one-third of our sample met the ASRS-v1.1 threshold for symptoms indicative of a clinical ADHD diagnosis. Given this unexpected variation, I used both a continuous symptom measure (centered score) and a binary indicator (score ≥ 4) in the analyses. The data suggest effective pre-treatment randomization: there are no statistically significant differences by condition after correcting for multiple comparisons.

3.2.6 Empirical specification

Treatment effects on the binary disclosure outcome are estimated using logistic regression models. Both unadjusted and adjusted specifications are reported. The unadjusted model includes only the randomized treatment indicators and their interactions. The full, adjusted model incorporates a set of preregistered covariates to account for observable heterogeneity, as specified in Equation 3.1:

$$\begin{aligned}\text{Disclosure}_i = \beta_0 &+ \beta_1 \text{LowBacklash}_i + \beta_2 \text{HighBacklash}_i + \beta_3 \text{DiagnosticFrame}_i \\ &+ \beta_4(\text{LowBacklash}_i \times \text{DiagnosticFrame}_i) + \beta_5(\text{HighBacklash}_i \times \text{DiagnosticFrame}_i) \\ &+ X_i'\gamma + \epsilon_i\end{aligned} \tag{3.1}$$

where Disclosure_i is a binary variable equal to 1 if participant i chose to disclose their neurodivergent status to their matched partner, and 0 otherwise. LowBacklash_i and HighBacklash_i are binary indicators for assignment to the respective treatment conditions, with the Control group serving as the reference category. DiagnosticFrame_i is a binary indicator equal to 1 if participant i received the diagnostic message framing, and 0 if the identity-based framing was shown.

β_1 and β_2 estimate the effects of potential social rejection under the Low and High Backlash conditions, respectively. β_3 captures the effect of receiving a more clinical (*diagnostic*) framing of neurodivergence, with specific reference to ADHD. β_4 and β_5 represent interaction effects, allowing the impact of message framing to vary across backlash conditions.

X_i' denotes the vector of preregistered control variables, including age, gender, education, ethnicity, employment status, general health, ADHD symptom score, and baseline task performance; ϵ_i captures

unobserved factors influencing disclosure decisions.

Robustness checks. Robustness checks examine the sensitivity of treatment effects to the inclusion of control variables and explore potential heterogeneity by ADHD symptom severity. Additional secondary analyses investigate whether disclosure is further associated with participants' fairness perceptions, second-order beliefs, or social preferences profiles. As noted in Section 3.2.3, all participants, including those flagged for an ASRS-v1.1 score below four, disengagement or reduced attentiveness, as well as initial failure to correctly answer the comprehension question, were included in the main analyses. However, these participants were analyzed in separate robustness and sensitivity analyses. To facilitate the interpretation and comparability of results, average marginal effects (AMEs) are reported alongside logit coefficient estimates where appropriate.

3.3 Study 2: Social reactions to disclosure (Player 2)

Study 2 is a follow-up study to investigate how a sample of non-neurodivergent adults respond to disclosure behavior from a matched sample with neurodivergent traits, specifically ADHD. Each participant is matched to a participant from the preceding study (Player 1) after Study 1 concludes, so that they view the actual disclosure decision and message framing assigned to their matched partner. While Study 2 was designed to complement Study 1, the primary variable of interest is whether Player 2 rejects their partner based on the framing of the disclosed information (diagnostic or identity-based) and the anticipated consequences of rejection.

3.3.1 Sample and pre-registration details

The full study design, including hypotheses, treatment conditions, outcome measures, and analysis plan for Study 2 (a follow-up to Study 1; see Section 3.2) were pre-registered on May 24, 2025 (OSF: `https://doi.org/10.17605/OSF.IO/A5P2E`).

The target sample size was 500 participants, chosen to mirror Study 1 and to allow for a one-to-one matching between participants. To further support this choice, an a priori power analysis was conducted. Drawing on prior literature, I assumed a 15% rejection rate under identity-based framing and a 30% rejection rate under diagnostic framing. This 15 percentage point difference is consistent with effect sizes commonly reported in lab studies of discriminatory behaviors in incentivized tasks (for a meta-analysis of experimental studies, see e.g., Lane, 2016). Research on mental health likewise

reports avoidance behaviors toward stigmatized conditions, such as depression or anxiety (Corrigan and Watson, 2002; Phelan, Link and Dovidio, 2008; Ridley, 2022). Based on these assumptions, a small-to-moderate effect size (Cohen's $d = 0.25$–0.35) was anticipated, suggesting a sample of 131 participants per group to provide at least 80% power to detect this difference (two-tailed test, $\alpha = 0.05$).

Participant recruitment and eligibility. Ultimately, a total of 750 individuals participated in Study 2, recruited via the online survey platform Prolific. One observation had to be excluded because its rejection choice was missing, resulting in a final sample of 749 participants. Data collection took place in June 2025. Eligibility was restricted to U.S.-based adults aged 18 to 65 who did not self-identify as neurodivergent[10] and who had not previously participated in any pilot for Study 1 or Study 2, nor in Study 1 itself. Each eligible participant received a unique, pre-generated Qualtrics link that embedded their matched Player 1's match ID, treatment assignment, and disclosure decision. If a participant failed to complete the survey, for example, because they dropped out before a decision could be recorded, or if a participant returned the survey, they were replaced until the target sample size was reached[11].

Compensation. All participants received a base payment of £2.70 (≈ $3.60). In addition, they could earn up to £1.50 (≈ $2.00), depending on their matched Player 1's performance bonus and the final matching outcome. Detailed instructions are in the Appendix.

Ethical approval and consent. Ethical approval was obtained from the LISER Research Ethics Committee under reference number [LISER REC/2024/127.HealthDisclosure/1]. Before accessing the study on Prolific, participants were informed that it would cover sensitive health topics. The consent form explained that participation was voluntary and anonymous, detailed their rights, and made clear that they could withdraw at any time without penalty. Participants were also explicitly informed that, in the High Backlash condition, rejecting their partner would potentially cause the partner to forfeit their earned bonus.

[10] That is, they responded "No" to the question, "Do you consider yourself to be neurodivergent?"

[11] Although the preregistered sample size was 500, I oversampled to ensure complete matching to Study 1 participants and to account for attrition or incomplete responses.

3.3.2 Experimental design and treatment conditions

Study 2 mirrored the design structure of Study 1 but did not implement a fully randomized 3×2 factorial design for Player 2. Instead, each Player 2 was matched *post hoc* to a Study 1 participant, such that Player 2's treatment was quasi-random, determined by the treatment assignment and disclosure decision of their matched Player 1. Player 1's treatment assignment and disclosure decision were embedded in unique, pre-generated survey links used to randomize Player 2's assignment at the distribution stage in Prolific, yet conditional on Player 1's actual choice. Note that, because Player 1's decision to disclose varied across the six Study 1 arms, the resulting Player 2 sample could not be perfectly balanced by (Backlash $\times$ Framing $\times$ Disclosure) cell. Thus, individual cell sizes reflected the actual frequency of disclosures in Study 1 rather than being equal by design (see Appendix A3.1.3).

Table 3.3 provides a summary of Player 2's treatment assignment, including (1) whether they were assigned to the Control or one of two Backlash conditions, inherited from Player 1, (2) whether Player 1 disclosed, (3) the information shared (if any), and (4) whether rejection was possible.

This matching structure aimed to preserve the structural variation from Study 1 and allowed me to analyze Player 2 behavior by condition. The three-level backlash manipulation for Player 2 therefore varied whether and how participants could reject their assigned partner:

In (1) the *Control* condition, participants were told that neither they nor their partner could reject the match, and that they would remain paired for the remainder of the study.

In (2) the *Low Backlash* condition, participants could choose to reject their match. If they did, their decision was implemented in 20% of cases, leading to both players receiving new partners. Otherwise, the match remained unaffected and continued as if no rejection had occurred.

Similarly, in (3) the *High Backlash* condition, participants were informed that they could reject their match, with a 20% chance of implementation. In this condition, however, if rejection was implemented, only Player 2 was reassigned, while Player 1 remained unmatched, forfeiting the final bonus they had already earned.

Participants were also exposed to a two-level message framing manipulation, shown only if their matched Player 1 had opted to disclose their status, namely (1) a *diagnostic framing*, stating that they reported symptoms commonly associated with a clinical ADHD diagnosis; or (2) an *identity-based*

Table 3.3: Experimental conditions of Player 2

Treatment	Disclosure & Framing	Option to reject?	Details
Control	No Disclosure	No	Sees no health information; cannot reject, remains matched.
	Disclosed: Diagnostic	No	Sees diagnostic framing; cannot reject.
	Disclosed: Identity-based	No	Sees identity-based framing; cannot reject.
Low Backlash	No Disclosure	Yes	Sees no health information; may reject and request reassignment.
	Disclosed: Diagnostic	Yes	Sees diagnostic framing; if reject and implementation occurs ($p = 0.2$), both are reassigned; otherwise, remain matched.
	Disclosed: Identity-based	Yes	Sees identity-based framing; if reject and implementation occurs ($p = 0.2$), both are reassigned.
High Backlash	No Disclosure	Yes	Sees no health information; may reject and request reassignment.
	Disclosed: Diagnostic	Yes	Sees diagnostic framing; if reject and implementation occurs ($p = 0.2$), only P2 is rematched; P1 loses bonus.
	Disclosed: Identity-based	Yes	Sees identity-based framing; if reject and implementation occurs ($p = 0.2$), only P2 is rematched; P1 loses bonus.

Note: Treatment reflects the combination of backlash condition and message framing, conditional on whether Player 1 disclosed. In Study 1, each combination (Backlash × Framing × Disclosure) was embedded in a unique Prolific survey link (with a match ID), ensuring Player 2's treatment cell matched Player 1's. Because Player 1's disclosure varied, Player 2's cell sizes were not perfectly balanced (see Appendix Table A3.2 for details).

framing, emphasizing that they fell into a category reflecting neurodivergent traits.[12]

3.3.3 Procedure

After providing informed consent, participants completed demographic and health questionnaires, including the ASRS-v1.1 (ADHD), PHQ-2 (depression), and GAD-2 (anxiety) screener. As in Study 1, participants completed three rounds each of a Go/No-Go task and an N-back task (1-, 2-, and 3-back), with randomized task order. Performance on these tasks was used to calculate the Balanced Integration Score (BIS), and the top 5% of participants received a $0.50 performance-based bonus.

[12]Note that, to isolate discriminatory responses to neurodivergent traits from inequality concerns of fairness or envy, participants in Study 2 were not informed about the time advantage related to disclosure granted to their assigned partners.

Next, participants read the profile information of their assigned partner from Study 1. They were informed that their final bonus payment would depend on the performance of this matched partner. In the Control condition, participants remained matched by default and were not given a choice. In the Low and High Backlash conditions, participants could choose to reject their assigned partner, where rejection was implemented with a probability of 20%. The consequences of rejection were fully disclosed at the decision stage:

- **Low Backlash:** If the rejection was implemented (20% probability), both players were reassigned to new partners, with no negative financial consequences. Player 2 would then receive the same final bonus as their new partner, and Player 1's bonus remained unaffected.
- **High Backlash:** If the rejection was implemented (20% probability), only Player 2 was reassigned; Player 1 continued unmatched and would not receive the bonus they had already earned.

Following this decision stage, the study concluded with the same post-task measures used in Study 1 (stated preferences, fairness perceptions, and second-order beliefs regarding stigma and discrimination). While the overall structure and procedure of Study 2 closely mirrored Study 1, Player 2 participants did not complete a final bonus task. Instead, their decision directly affected whether their matched Player 1 partner received the bonus they had earned in the prior study.

To ensure that study subjects understood the implications of the experimental matching and bonus structure, they had to complete a brief manipulation and comprehension check after they made their decision about their assigned match. The manipulation check assessed participants' understanding of the content of the message they had just read. Before proceeding to the final part of the study, participants were further randomly assigned to one of two comprehension checks. Similar to Study 1, if participants failed to answer correctly, they were provided with corrective feedback and asked to provide the correct answer before they could continue. Although participants did not have to pass any formal attention check, they were flagged if (1) they exhibited extreme performance patterns on cognitive tasks suggestive of inattention or reduced engagement, or (2) they answered their assigned comprehension question incorrectly on the first attempt[13].

[13]See Table A3.1 in the Appendix for details on flagged observations.

3.3.4 Outcome measures

The primary outcome is a binary indicator of rejection, capturing whether participants in the backlash conditions chose to reject their assigned partner. Participants in the Control condition were not eligible to make a choice. The rejection decisions were coded as 1 (reject) or 0 (stay matched), with a probability of 20% of implementation, and in some cases determined the forfeiture of the final bonus for Player 1.

Secondary outcomes are the same as reported for Study 1 (see Section 3.2.4): Task performance (accuracy, RT, missed trials, false alarms (Go/No-Go), and incorrect responses (N-back), summarized via the BIS), social preferences (via the *stated satisfaction* approach), fairness perceptions and second-order beliefs pertaining to individuals with specific stigmatized characteristics, including neurodivergence. As preregistered, the control variables include age, gender, ethnicity, education, employment status, general health, mental health symptoms, ADHD symptom score, and stimulant use on the day of the experiment.

3.3.5 Empirical specification

Treatment effects on the binary rejection decision are estimated using logistic regression. Since participants in the Control condition always remained matched with their assigned partners (i.e., rejection is always zero), the main empirical analysis of rejection is conducted exclusively among participants assigned to the one of the two backlash conditions. The main specification thus includes indicators for message framing (diagnostic vs. identity-based), backlash intensity (High vs. Low), and their interaction, alongside a set of control variables:

$$\begin{aligned} \text{Reject}_i = \alpha_0 &+ \alpha_1 \text{DiagnosticFrame}_i + \alpha_2 \text{HighBacklash}_i \\ &+ \alpha_3 (\text{DiagnosticFrame}_i \times \text{HighBacklash}_i) \\ &+ X_i' \delta + \nu_i \end{aligned} \tag{3.2}$$

where Reject_i is a binary variable indicating whether the participant rejected their assigned partner (1 = reject, 0 = accept), DiagnosticFrame_i denotes the diagnostic framing (vs. identity-based framing), and HighBacklash_i distinguishes between high (1) and low (0) backlash conditions. X_i' represents the transpose of the vector of control variables, including age, gender, education, ethnicity, employment status, general health, ADHD symptom score, and baseline task performance; ν_i captures unobserved

factors influencing rejection decisions.

Pooling backlash conditions. To increase statistical power in testing message framing effects, Low and High Backlash conditions are pooled:

$$\text{Reject}_i = \beta_0 + \beta_1 \text{DiagnosticFrame}_i + X_i'\gamma + \varepsilon_i \tag{3.3}$$

where β_1 directly estimates the difference in rejection rates between diagnostic and identity-based framings, controlling for other covariates.

Robustness checks. Robustness analyses re-estimate the logistic regression separately for Low and High Backlash conditions, to test for non-linear differences in rejection rates across the two backlash conditions. Given that the severity of consequences following rejection may influence decisions in a non-linear way, i.e. participants may respond very differently to Low vs. High Backlash conditions. As rejection in the High Backlash condition may result in forfeiture of another participant's earned bonus, comparing rejection by condition helps provide additional insight into the question whether participants are more sensitive to the potential financial consequences of their rejection or the signaling power of rejection alone.

3.4 Disclosure decisions under the risk of social backlash

In what follows, I first present the results pertaining to the likelihood of disclosure among the $N = 659$ participants with self-reported ADHD, subject to three experimental conditions: Control, Low Backlash, and High Backlash. Disclosure was defined as the choice to reveal one's neurodivergent status, framed either as "symptoms commonly associated with a clinical ADHD diagnosis" (diagnostic framing) or as "neurodivergent cognitive traits" (identity-based framing).

I formally test differences across treatments by estimating both unadjusted and adjusted logistic regression models, under the hypothesis that participants were less likely to disclose their neurodivergent condition when facing a risk of rejection and potential economic loss, compared to participants who would only face a psychological cost in terms of potential judgment by their matched partner. Across all specifications, neither backlash condition nor message framing had a statistically significant impact on the predicted probability of disclosure ($p > 0.59$). Similarly, non-parametric tests indicated no significant differences in disclosure rates across treatment cells

($\chi^2(2) = 0.30, p = 0.863$). For robustness, I re-estimated the models using inverse probability weights to adjust for potential bias from variation in performance and comprehension. The results were similar, and none of the treatment effects reached significance.

Even though both backlash conditions imposed real (albeit probabilistic) consequences such as revealing sensitive information to a partner and, in the High Backlash arm, risking the loss of earned bonus payments, participants disclosed their neurodivergent status at similar rates, even when facing a substantial risk of losing their bonus. As shown in Figure 3.1, average disclosure rates all hover around 67–71%, standing in sharp contrast to my predictions that higher risk of rejection would reduce willingness to disclose. Overall, neither the risk of social backlash nor the specific message framing led to observable changes in disclosure rates.

3.4.1 Are higher ADHD scores predictive of disclosure?

In addition to the preregistered tests of treatment and framing effects, I examine whether variation in ADHD symptom severity predicts disclosure behavior under different social and economic costs. Participants were grouped by symptom severity using the ASRS-v1.1 cutoff (score ≥ 4), revealing that within this niche sample of self-reported ADHD adults, roughly one-third (34%) met the suggested threshold for symptoms indicative of ADHD (see also Table 3.2 in Section 3.2.5).

Figure 3.1 shows disclosure rates by treatment arm and message framing, stratified by ASRS score. Disclosure appears largely stable across symptom groups in the Control and High Backlash conditions. However, in the Low Backlash condition, a gap becomes visible: 78% of participants with higher symptoms chose to disclose, compared to 61% of low-symptom participants. Table 3.4 shows that this difference is marginally significant ($p = 0.056$), suggesting that when rejection risk is present but not too punitive (in terms of potential forfeiture of a bonus), individuals with more pronounced symptoms may perceive a greater benefit in disclosure.

Corresponding logistic regressions (Table A3.4 in the Appendix) show no significant main effects of ADHD severity on disclosure, whether modeled continuously ($OR = 1.05$, $p = 0.25$) or as a binary indicator ($OR = 1.10$, $p = 0.59$). However, once interaction terms are introduced, the story becomes more textured: In adjusted models (Panel B), individuals above the ADHD threshold were significantly more likely to disclose in the Low Backlash condition relative to Control ($OR = 2.47$, 95% CI $[1.02, 6.05]$, $p = 0.047$)[14]. In the more robust, inverse probability weighted model (Panel

[14]This finding is supported by a post hoc power analysis, which indicates 94% power to detect an effect

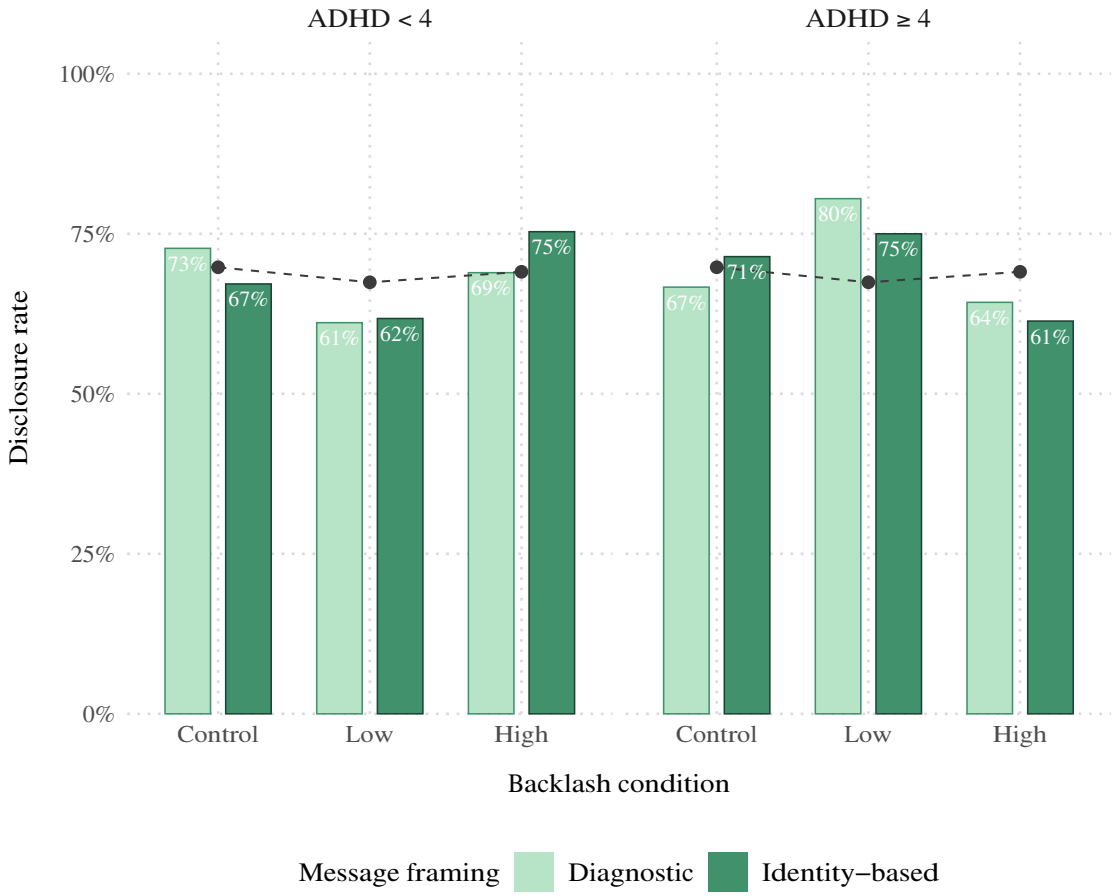


Figure 3.1: Disclosure rates by backlash condition and message framing, separated by ADHD symptom severity

Notes: Bars represent the proportion of participants who chose to disclose in each condition, by message framing (diagnostic vs. identity-based). Facets separate ASRS-v1.1 results for participants with lower (ADHD < 4) and higher (ADHD ≥ 4) symptom scores. The dashed line marks the average disclosure rate for each backlash condition across the full sample. Pairwise comparisons are Holm-adjusted for multiple testing. No differences are statistically significant after adjustment (all $p > 0.05$).

C), the interaction effect increases in magnitude, leading to an estimate for the ADHD × Low Backlash interaction of $OR = 2.84$ (95% CI $[1.11, 7.27]$, $p = 0.03$). Modeling ADHD as a continuous variable, centered at the mean, results in a similar but slightly less pronounced effect ($OR = 1.22$, 95% CI $[0.97, 1.54]$, $p = 0.094$). Overall, these estimates imply that the impact of the Low Backlash treatment depends on participants' symptom severity: while the main effect of Low Backlash is significantly negative compared to the control group ($OR = 0.49$, $p = 0.010$), deterring disclosure

of this magnitude ($OR = 2.47$) given the observed group sizes within the Low Backlash condition ($n = 81$ for ADHD score ≥ 4 and $n = 140$ for score < 4), offering tentative empirical support for the presence of an interaction in the full sample, despite the marginal p-value.

Table 3.4: Disclosure rates by ADHD cutoff score and backlash condition

	Disclosure rate		*p*-value	
Backlash condition	Score < 4	Score ≥ 4	Unadjusted	Holm-adj.
Control	0.70	0.69	0.991	0.991
Low Backlash	0.61	0.78	0.019*	0.056
High Backlash	0.72	0.63	0.191	0.382
Observations	435	224		

Significance level: * $p < 0.05$.
Notes: Disclosure rates are shown separately for participant groups scoring below and above the thresholds on the 6-item Adult ASRS-v1.1 screener (Kessler *et al.*, 2005), for each backlash condition. *p*-values are from Pearson's χ^2 tests, reported with and without Holm adjustments for multiple comparisons.

among participants below the ASRS threshold, the interaction term more than offsets this for those above the threshold and appears to encourage disclosure among those with more pronounced ADHD symptoms.

3.4.2 Do fairness perceptions predict disclosure of neurodivergent traits?

Table 3.5 presents the findings from both ordinal and binary logistic regressions testing whether participants' perceptions of fairness predict their likelihood of disclosing their neurodivergent status. Panel A reports odds ratios from ordinal logistic models estimating how fairness perceptions differ between those who disclosed and those who did not. Panel B presents logistic regression results assessing whether fairness perceptions are predictive of disclosure, and Panel C replicates these models using inverse probability weighting.

Panel A reports that across both fairness scenarios, participants who disclosed their neurodivergence were significantly less likely to perceive the receipt of accommodations as unfair. The odds of rating the scenario as more unfair (i.e., moving from "acceptable" to "unfair") were about 35% lower among disclosers, regardless of whether the advantage was framed as due to health ($OR = 0.65$, $p = 0.010$) or luck ($OR = 0.65$, $p = 0.009$). Panel B similarly shows that perceived unfairness may discourage disclosure: moving one unit up the unfairness scale (e.g., from *acceptable* to *unfair*) was associated with a 24% drop in the odds of disclosure (health: $OR = 0.76$, $p = 0.017$; luck: $OR = 0.76$, $p = 0.008$). The corresponding marginal effects imply a 5.8 percentage point decline in

Table 3.5: Fairness perceptions and willingness to disclose neurodivergence

	OR	95% CI	p-value	AME	SE
Panel A: Fairness perceptions					
Advantage based on health	0.65^{**}	0.47, 0.90	0.010	–	–
Advantage based on luck	0.65^{**}	0.47, 0.90	0.009	–	–
Panel B: Disclosure					
Advantage based on health	0.76^{*}	0.60, 0.95	0.017	-0.058^{*}	0.024
Advantage based on luck	0.76^{**}	0.62, 0.93	0.008	-0.058^{**}	0.021
Panel C: Disclosure (IPW)					
Advantage based on health	0.87	0.68, 1.12	0.290	−0.029	0.027
Advantage based on luck	0.80	0.64, 1.01	0.063	−0.046	0.024

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Panel A reports odds ratios (OR) from ordinal logistic regressions of fairness perceptions on disclosure. Given the ordinal nature of the dependent variable (DV), I do not report marginal effects, which vary across outcome categories and are not straightforwardly interpretable as a single change in probability. Panel B presents logistic regressions with disclosure as DV and fairness ratings as predictors; average marginal effects (AMEs) reflect the predicted change in disclosure probability from 1 (*Completely fair*) to 4 (*Very unfair*). Panel C replicates Panel B using inverse probability weighted (IPW) estimators. All models adjust for experimental condition and framing, age, gender, employment, mental health, ADHD, and Go/No-Go performance (coefficients not shown).

predicted disclosure probability per unit increase, indicating that regarding any granted advantage for someone else as unfair corresponds to being less willing to disclose one's own condition. These associations attenuate once model adjustments via inverse probability weights are introduced (Panel C): although the odds ratios and marginal effects still point in the same direction, they are no longer statistically significant (health: $OR = 0.87$, $p = 0.290$; luck: $OR = 0.80$, $p = 0.063$), suggesting that the earlier findings may partly reflect selection related to task engagement or comprehension.

Social preference types. The previous results indicated a possible association between the perceived fairness or legitimacy of accommodations and participants' disclosure decisions. This prompted me to explore the underlying social and distributional preferences more thoroughly, and to test whether individuals with different preference profiles might respond differently in the face of potential discrimination. Thus, instead of assuming predefined categories, I used K-means clustering on participants' satisfaction ratings from distributional choice scenarios to classify them into four distinct preference types. These clusters, described in detail in Appendix A3.4, are then used to examine whether participants with certain preference attributes are more or less inclined to disclose. Table 3.6 reports the results.

Compared to the reference cluster, participants classified into the *guilt* cluster were significantly

Table 3.6: Odds of disclosure across social preference profiles

	(1)			(2)		
	OR	95% CI	p-value	OR	95% CI	p-value
Cluster (ref: 3: Self-interest)						
1: Guilt	0.62*	0.38, 0.98	0.044	0.58*	0.36, 0.93	0.024
2: Envy	0.90	0.55, 1.45	0.660	0.86	0.53, 1.41	0.560
4: Positional concerns	0.63	0.38, 1.04	0.073	0.62	0.37, 1.04	0.068
Controls		N			Y	
Observations		659			659	

Significance level: * $p < 0.05$.
Notes: Logistic regression of disclosure on social preference clusters and treatment conditions. Model (1) includes only social preference type (cluster); Model (2) adjusts for experimental condition and framing, age, gender, employment, mental health, ADHD, and Go/No-Go performance (coefficients not shown).

less likely to disclose their neurodivergent status to their matched partner. While the reference cluster is composed of predominantly inequality-tolerant individuals whose decisions appear largely driven by self-interest, participants in the guilt cluster could be defined as a group of individuals with a tendency to dislike being better off than others. The odds ratio for this type in the most parsimonious model (column 1) is 0.62 ($p = 0.044$). After adjusting for covariates (column 2), this effect remains statistically significant ($OR = 0.58$, $p = 0.024$) as well as robust to selection bias due to comprehension and performance tasks[15]. The remaining clusters, which resemble two different kinds of disadvantageous inequality aversion, also tend to have lower odds of disclosure, though neither of the effects was statistically significant.

3.4.3 Does anticipated discrimination toward neurodivergence predict disclosure?

Although message framing did not alter overall disclosure rates, the data suggest that it did influence second-order beliefs – primarily among those participants who opted for disclosure (Table 3.7). Specifically, disclosers who received the diagnostic frame expected that anticipated acceptance of neurodivergence would be lower compared to those receiving the identity-focused description ($OR = 0.69$, 95% CI $[0.50, 0.96]$, $p = 0.03$); whereas no such differences were observed among non-disclosers ($p = 0.94$). This finding suggests that framing effects on beliefs may arise primarily if

[15]For robustness, I re-estimate the logit models weighted by inverse probability weights, which yields similar results (see Appendix Table A3.5).

the participant actually decides to disclose their condition. Given this link, a natural next step is to examine whether these post-treatment beliefs about others' acceptance of neurodivergence are themselves associated with disclosure decisions, and if such associations depend on the perceived psychological or economic costs of disclosure.

Table 3.7: Effect of framing on second-order beliefs

	(1) Disclosers	(2) Non-disclosers
Diagnostic framing (ref: Identity-based)	0.69*	0.98
	[0.50, 0.96]	[0.58, 1.66]
Controls	Y	Y
Observations	453	206

Odds ratios (OR) with 95% confidence intervals in brackets. Significance level: * $p < 0.05$.
Notes: Ordered logit models adjust for backlash condition, age, gender, employment status, ADHD symptom score, PHQ-2, GAD-2, and BIS score (coefficients not shown).

While the belief measures were collected post-treatment and cannot credibly be used to claim causation, they still provide an opportunity to examine whether participants' second-order beliefs about others' acceptance of neurodivergence are themselves associated with disclosure decisions, and whether this relationship varies with the perceived psychological or economic costs of doing so. Disclosure risk is manipulated through a backlash condition in which non-disclosure shields participants (Player 1) from reputational or financial consequences. To examine this relationship, I estimate logistic regressions predicting disclosure decisions as a function of second-order beliefs and treatment assignment. The main results are presented in Table 3.8, with complementary models reported in the Appendix (Table A3.6).

Table 3.8 shows that positive expectations about others' acceptance of neurodivergence consistently predicted higher odds of disclosure across all specifications. Column (1) presents the results for the pooled responses, while column (2) and (3) present the results by subgroup[16]. Column (4) applies inverse probability weights. The results clearly show that participants anticipating more favorable attitudes toward neurodivergence are significantly more likely to disclose their own condition ($OR = 1.33$, $p < 0.001$, pooled responses). Similar effects are observed when looking into the effects by subgroup: when statements refer explicitly to the general population ($OR = 1.39$, $p = 0.01$) or to other study participants ($OR = 1.29$, $p = 0.025$).

[16]The survey items prompted participants to rate how either "most participants in this study" (study sample) or "most people in the general population" (population sample) would perceive someone if they knew the person had one of the following attributes: physical disability, low-income background, history of mental health challenges, or was neurodivergent (e.g., ADHD, autism). Note that not everyone received the same version. Instead, the two versions were randomized across participants.

Table 3.8: Perceived social attitudes toward minoritized groups

	Pooled belief responses	General population	Study participants	Pooled belief responses (IPW)
	(1)	(2)	(3)	(4)
Second-order beliefs				
Neurodivergence	1.33^{***}	1.39^{*}	1.29^{*}	1.28^{**}
Mental health history	1.18^{*}	1.20	1.16	1.15
Low-income background	1.07	1.09	1.05	1.03
Physical disability	1.07	1.14	0.97	1.04
Controls	Y	Y	Y	Y
Observations	659	328	331	659

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Results from logistic regression models. Second-order beliefs refer to participants' perceptions of how others would evaluate individuals with specific attributes (e.g., having a physical disability, neurodivergence, etc.). Column (1) – (3) present the unweighted regression results; column (4) applies inverse probability weighting (IPW) for robustness. All models include experimental conditions (backlash, framing) and covariates. Controls: age, gender, employment, mental health, ADHD, and Go/No-Go performance (coefficients not reported).

For a comparative overview, the table also includes the results for all four minoritized identities used in the survey to measure variation in second-order beliefs about others' acceptance. More positive second-order beliefs pertaining to mental health history are also positively associated with disclosure, although the impact appears weaker ($OR = 1.18$, $p < 0.05$, pooled responses). By contrast, beliefs about low-income background and physical disability are not significantly related to disclosure. These cross-regression comparisons suggest that participants are more sensitive to how others may respond to neurodivergence than to other stigmatized traits, when deciding whether to disclose. One possible explanation is that neurodivergence, unlike physical disability or socioeconomic background, is often concealable but still associated with social risk. This makes perceptions of others' attitudes particularly consequential, and participants appear more inclined to disclose when they expect a more accepting response.

Building on this, a natural question arises: do individuals' own beliefs about how society views their condition affect how they behave when the risk of backlash from disclosure changes? Figure A3.2 illustrates the predicted probabilities of disclosure across backlash conditions, conditional on participants' second-order beliefs about how neurodivergence is perceived. Both the baseline and the inverse probability weighted models suggest that participants who believe others hold favorable views of neurodivergence (e.g., ADHD or autism) are more likely to disclose their own condition, regardless of backlash condition. The interaction is most pronounced in the High Backlash group: those with

more positive expectations toward other people's beliefs appear more likely to disclose their own condition, even when the signal is costly. This finding suggests that the impact of anticipated stigma and rejection on information disclosure may be context-dependent, varying with the perceived cost of signaling one's condition.

Appendix Table A3.6 extends this analysis. Panel A first shows the main effects without interactions: a one-unit increase in perceived acceptance is associated with a 5.9 percentage point increase in the probability of disclosure ($AME = 0.059$, $SE = 0.017$). Panel B presents the unweighted interaction model, where second-order beliefs remain positively associated with disclosure ($AME = 0.060$, $p < 0.001$), although the odds ratio does not reach conventional significance levels ($OR = 1.26$, $p = 0.10$). Interaction terms between treatment conditions and second-order beliefs are not statistically significant. These results remain robust to the application of inverse probability weights (Panel C), suggesting that perceptions of others' views matter, even though treatment condition does not significantly moderate this effect.

Heterogeneous second-order beliefs across samples. Table 3.9 shows that participants in the first study (ADHD sample) consistently expected others to view individuals with various stigmatized attributes (such as a physical disability or a history of mental health challenges) more positively than did participants in the second study (non-neurodivergent sample). Each respondent was randomly assigned to one of two versions: either asked about "most participants in this study" or about "most people in the general population". Across both versions, ADHD participants consistently anticipated more positive attitudes (i.e., less stigma) than their non-neurodivergent counterparts. These differences were most pronounced when respondents considered how other study participants would react. When asked about the general population, the results were similar, though significant differences were observed for mental health history ($p = 0.018$) and neurodivergence ($p < 0.001$) only. On average, perceptions of the general population were rated slightly more negatively than those of the study sample, across both study populations.

3.4.4 Summary

Participants' disclosure decisions were not significantly affected by the experimentally induced risk of social backlash or economic loss, contrary to the initial hypothesis. While both treatment and message framing manipulations were designed to simulate rejection and reputational cost, neither condition alone altered overall disclosure rates. Instead, disclosure behavior appears to be more

Table 3.9: Comparison of second-order beliefs: mean stigma ratings by reference group

	Mean (P1)	Mean (P2)	p-value	Mean Diff.
Reference: Other study participants (n)	331	373		
Physical disability	3.11	2.94	0.037^{*}	0.18
Low-income background	3.02	2.86	0.038^{*}	0.16
Mental health history	2.75	2.51	0.005^{**}	0.24
Neurodivergence	2.89	2.65	0.005^{**}	0.24
Reference: General population (n)	328	376		
Physical disability	2.72	2.61	0.148	0.11
Low-income background	2.65	2.55	0.165	0.1
Mental health history	2.21	2.07	0.047^{*}	0.14
Neurodivergence	2.54	2.29	$< 0.001^{***}$	0.26

*, ** and *** Significant mean differences at $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively, between P1 (ADHD sample) and P2 (non-neurodivergent sample), based on independent samples t-tests. Ratings were collected via two randomly assigned vignette conditions. One half of participants rated how they believed *other participants in this study* would perceive individuals with specific characteristics, while the other half rated perceptions on the same characteristics, but in terms of discriminatory attitudes held by the *general population*. All items were preceded by the question: "How do you think most [participants in this study / people in the general population] would perceive someone when deciding whether to remain matched with them — if they knew the person had one of the following attributes?". Responses were recorded on a 5-point Likert scale ranging from 1 = *Very negative* to 5 = *Very positive*.

strongly influenced by perceived fairness and second-order beliefs about social stigma. My results show that participants who considered accommodations as fair were significantly more likely to disclose, suggesting that perceived legitimacy of resource allocation plays an important role in disclosure under uncertainty.

Participants who believed others were more accepting of neurodivergence were more likely to disclose, consistent with the idea that favorable second-order beliefs reduce both the perceived cost and the anticipated harm of rejection. The framing manipulation shifted second-order beliefs, but only among those who disclosed; non-disclosers' beliefs appeared unaffected. This suggests that individuals already inclined to disclose were more responsive to differences in framing, pointing to possible self-selection. Across all specifications, second-order beliefs emerged as the most reliable predictor of disclosure, with their effect especially pronounced in the high backlash condition, where favorable expectations about others' attitudes seemed to offset the deterrent effect of risk.

3.5 Social backlash and rejection decisions

To examine how Player 2s respond to disclosure decisions by their matched Player 1s, and whether these responses vary based on how neurodivergence is framed, I analyze rejection behavior among non-neurodivergent participants using a post hoc matched sample. Participants were exposed to one of two randomly assigned framings of their partner's self-reported ADHD: a diagnostic description or identity-based, neurodiversity-informed language, conditional on Player 1's decision to disclose.

I first show descriptive evidence that the rejection rate among Player 2 participants did vary significantly, conditional on message framing. Among those who were assigned to someone who revealed their ADHD framed in diagnostic terms, 9.5% rejected their match. In contrast, only 4.6% rejected their match when their partner revealed *neurodivergent cognitive traits*. A Fisher's exact test confirms that this difference is statistically significant ($OR = 2.17$, 95% CI $[1.01, 4.92]$, $p = 0.035$), suggesting that the diagnostic framing clearly increases the odds of rejection, conditional on disclosure. These differences are illustrated in Figure 3.2, with rejection rates displayed by message framing and backlash condition.

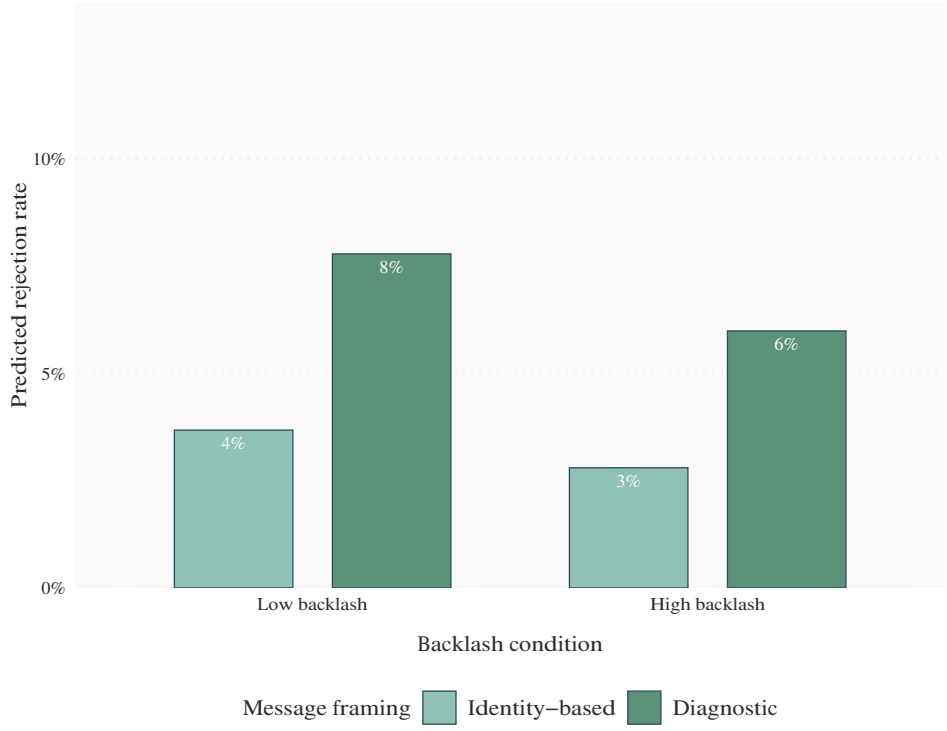


Figure 3.2: Predicted rejection rates by treatment cell

Notes: Bars represent the proportion of Player 2s who rejected their partner across message framings (diagnostic vs. identity-based) and backlash conditions (Low vs. High).

Next, to more precisely estimate the effects of message framing and its potential moderation by backlash condition on rejection decisions, I fit logistic regressions predicting rejection as a function of message framing, backlash condition (Low vs. High), and their interaction, controlling for demographic, health, and performance-related covariates. The results of both unweighted (Panel A) and inverse probability weighted regressions (Panel B) are reported in Table 3.10.

Table 3.10: Effect of message framing on rejection by Player 2

	(1)	(2)	(3)
Panel A: Unweighted logistic regression			
Diagnostic framing (ref: Identity-based)	2.17*	2.38*	1.40
	[1.07, 4.60]	[1.16, 5.11]	[0.48, 4.17]
High Backlash (ref: Low Backlash)			0.74
			[0.21, 2.43]
Diagnostic × High Backlash			2.70
			[0.61, 13.00]
Panel B: IPW-weighted logistic regression			
Diagnostic framing (ref: Identity-based)	2.63*	2.80*	1.33
	[1.20, 5.76]	[1.25, 6.24]	[0.44, 4.01]
High Backlash (ref: Low Backlash)			0.96
			[0.27, 3.41]
Diagnostic × High Backlash			3.45
			[0.77, 15.51]
Controls	N	Y	Y
Observations	504	504	504

Odds ratios (OR) with 95% confidence intervals in brackets. Significance level: * $p < 0.05$.
Notes: Coefficients are from logistic regression models estimated on a non-neurodivergent sample (Player 2), testing whether message framing affects rejection of a matched (neurodivergent) partner. Panel A shows unweighted estimates; Panel B shows inverse probability weighted (IPW) estimates. Model (1) includes only the framing variable. Model (2) adds controls. Model (3) additionally includes an interaction term. Controls: age, gender, ethnicity, employment, education, general and mental health, ADHD, performance BIS (Go/No-Go, N-back), substance use (coefficients not reported).

Regression analyses strongly support the descriptive findings reported above. Across all model specifications, diagnostic framing significantly increases the likelihood of rejection by Player 2s. In the unadjusted (bivariate) model (1), the odds of rejection were more than twice as high in the diagnostic framing condition compared to the identity-based frame ($OR = 2.17$, $p = 0.034$), indicating that players viewing the diagnostic framing were more likely to reject their matched partner compared to those receiving the identity-focused message. This effect remained robust after adjusting for controls (Model 2: $OR = 2.38$, $p = 0.027$) and still appeared present, albeit attenuated, in the interaction model (Model 3). Once rejection outcomes for Player 1s also involve

economic costs, the effect of the diagnostic framing was smaller and no longer statistically significant ($OR = 1.40$, 95% CI $[0.48, 4.17]$). The interaction term suggests a possible moderation effect, with the effect of diagnostic framing possibly concentrated in the Low Backlash condition. Although the wide confidence interval precludes strong conclusions, the findings overall indicate that diagnostic framing increases rejection risk.

To better understand this heterogeneity, I also test whether participants respond differently to the backlash conditions, conditional on whether their partner's condition is framed in diagnostic terms. The results are presented in Table A3.7 in the Appendix. In the Low Backlash group, diagnostic framing significantly increased the odds of rejection ($OR = 4.11$, 95% CI $[1.45, 13.63]$, $p = 0.012$). This effect remains positive and statistically significant when applying inverse probability weights ($OR = 4.55$, $p = 0.015$), indicating a robust influence of diagnostic language on rejection in this group. In contrast, I find no evidence of a framing effect in the High Backlash condition ($OR = 1.37$, $p = 0.564$), and the framing effect appears to be concentrated entirely among participants in the Low Backlash subsample.[17]

3.6 Does disclosure pay off?

Although not explicitly part of the initial preregistration, I also briefly examine how my results link to task performance, and whether disclosure has any measurable effect on participants' later bonus task performance. The analysis is exploratory and motivated by a simple but potentially important question: do individuals who disclose also perform better when it matters?

My results reveal that baseline task performance and disclosure decisions are closely linked in the data. Individuals with higher Go/No-Go performance scores were significantly less likely to disclose their condition: a one standard deviation increase in baseline performance was associated with a 17–18% reduction in the odds of disclosure ($OR = 0.82$–0.83, $p < 0.05$; Table A3.8 in the Appendix). A χ^2 test confirmed that participants with lower Go/No-Go scores (i.e., scores below the median) were significantly more likely to disclose than their high-performing counterparts ($\chi^2(1) = 12.05$, $p < 0.001$). Not surprisingly, as the accommodation was especially designed to provide

[17] Interestingly, scoring higher on the ASRS-v1.1 scale was also associated with fewer rejections in the Low Backlash group, even after reweighting. While noteworthy, I do not examine this or other exploratory findings from Study 2 further, as this chapter's main focus is on disclosure behavior. Study 2 data were collected primarily to provide Player 1 with credible, consequential choice scenarios and to avoid reliance on hypothetical vignettes.

a performance-enhancing advantage, disclosure was strongly associated with better performance in the final incentivized task: disclosers achieved higher BIS scores ($\beta = 1.90$, $p < 0.001$), greater accuracy ($\beta = 0.048$, $p = 0.014$), faster reaction times ($\beta = -0.93$, $p < 0.001$), and substantially lower miss rates ($\beta = -12.46$, $p < 0.001$), even after adjusting for baseline ability, bonus task type, and experimental condition (Table A3.9 in the Appendix).[18]

3.7 Discussion

The results reported in this chapter reveal that disclosure behavior is not as straightforward as initially anticipated. Contrary to my predictions, I find that neither the manipulation of social backlash and its consequences nor the framing of ADHD significantly affected the overall willingness to disclose. The hypotheses assumed a linear deterrence effect, where higher expected costs would lead to progressively lower disclosure rates. Instead, the findings point to a possible non-linear relationship, with some conditions (e.g., Low Backlash for high-symptom participants) linked to unexpectedly higher disclosure rates. Surprisingly, baseline disclosure rates remained high across all experimental arms: nearly 70% of participants chose to reveal their condition, regardless of framing, and even under conditions with a credible risk of social rejection or the forfeiture of already earned bonus payments.

This finding holds across unadjusted and fully adjusted models, as well as robustness checks, and even holds when psychological and economic costs are introduced, standing in sharp contrast to the preregistered hypotheses and economic theories of signaling and social image concerns (see, e.g., Acquisti, John and Loewenstein, 2013; Bénabou and Tirole, 2006; Grossman and Van Der Weele, 2017). One possible explanation for the absence of a significant effect following the framing manipulation lies in a potentially combined effect of limited public familiarity and the incentive structure in the present study. While the diagnostic frame drew on a widely recognized clinical category ("ADHD"), the identity-based frame relied on the term "neurodivergent", which may have been less clearly linked to ADHD for many participants. At the same time, disclosure was linked to a tangible performance advantage on the final bonus task, with the prospect of an immediate benefit, which may have outweighed subtle differences in perceived stigma between the two framings.

[18]Note that the inverse probability weights (IPW) used elsewhere already adjust for performance, which these models also control for; weighted estimates were nearly identical. For robustness, I also tested whether failed comprehension checks influenced the results. Passing the check was unrelated to any bonus task outcome, and adding it as a control left the disclosure estimates unchanged (all $p > 0.25$).

However, framing was not entirely inert: among participants who chose to disclose, diagnostic framing reduced second-order beliefs about others' acceptance of neurodivergence compared to the identity-based framing. This suggests that even if framing does not shift overall disclosure rates, it might have actually changed how participants thought others would react. Thus, if someone disclosed and their condition was described as "symptoms associated with a clinical ADHD diagnosis", they tended to believe that others would be less accepting of neurodivergence than if it had been described pertaining to "neurodivergent cognitive traits". The wording appeared to affect beliefs, even though it did not affect the decision to disclose in the first place.

An alternative explanation for the high baseline disclosure rate could be the occurrence of a possible "ceiling effect", which may have limited the experiment's ability to detect treatment effects. Although participants were not aware of the specific context of the experiment prior to recruitment, self-selection could still have contributed to the high willingness to reveal (potentially stigmatizing) health information in a strategic decision context. Prolific's infrastructure allows researchers to recruit niche samples, based on pre-screened characteristics, such as ADHD or autism, prior to recruitment. However, this self-identification as having ADHD and general eligibility for research may itself be explained by a more liberal attitude toward their condition, and thus with a greater comfort around disclosure, driving their decisions. Moreover, while all participants were included in the study based on their condition, only about one third (34%) reached the suggested cutoff on the ASRS-v1.1 screener, a self-report scale used to assess ADHD symptoms in adults (Kessler *et al.*, 2005). The screener is not diagnostic, yet the low share of participants meeting the screener threshold could have weakened the treatment effects, for example, if those with less severe symptoms were less sensitive to rejection.

While I do not find that participants' disclosure decisions are significantly affected by the main experimental manipulations themselves, my results point to an interaction effect of the treatments with symptom severity. Although one might expect individuals with more pronounced ADHD symptoms to be more reluctant to disclose, possibly due to greater concerns about stigma, the data suggest a more complex relationship. Disclosure rates among higher-symptom participants are not only slightly elevated but also, under specific conditions, markedly so: in the weighted model, participants above the ASRS threshold were substantially more likely to disclose under the Low Backlash than under the Control condition, reaching an odds ratio of 6.50. This interaction with the Low Backlash condition is both statistically significant and stable across model specifications ($OR = 2.47$ in the full sample). This suggests that perceived social costs, such as the prospect of

rejection (without financial loss), may differentially affect disclosure decisions depending on symptom severity.

Additionally, I find evidence that both fairness perceptions regarding accommodations and second-order beliefs about others' attitudes toward neurodivergence are significantly associated with disclosure outcomes. Participants who perceived the receipt of performance-enhancing accommodations as fair (especially when this accommodation was granted based on someone's health condition) were more likely to disclose, whereas those who viewed such benefits as unfair were more likely to continue the experiment withholding their information, even if this meant missing out on the chance of gaining an advantage that could improve their performance. Aside from fairness perceptions, participants' beliefs about how others view neurodivergence also mattered: disclosure was more likely when individuals believed that others in their environment held accepting attitudes toward neurodivergence, particularly under higher potential costs, where favorable beliefs appeared to mitigate the perceived risks of rejection.

Secondary analyses suggest that disclosure may also reflect a strategic choice: individuals who revealed their condition performed better in the bonus task phase, even after adjusting for baseline performance and task type. While the disclosure decision offered an incentive through a performance-enhancing accommodation, one possible explanation could be that those confident in their ability to benefit may have been more likely to disclose.[19] Although the results are consistent with forward-looking behavior under uncertainty, causal claims cannot be established.

My findings align with prior research demonstrating that disclosure decisions often deviate from standard economic predictions under asymmetric information (Benndorf, Kübler and Normann, 2015): participants do not always strategically withhold their information, even when they are presented with a clear reputational or economic risk. Rather than considering only the material incentives, participants in my study appear to be influenced by what they consider fair, and by the attitudes they think others hold toward neurodivergence. These results also matter from a policy perspective.

Because the threat of social backlash or the way ADHD and neurodivergence are framed did not actually deter participants from disclosing, simply lowering the cost or risk of disclosure alone (for example, through introducing new workplace policies) might not be sufficient to encourage individuals to claim accommodations they are legally entitled to. Instead, it may be more effective to

[19] An alternative interpretation is that particularly confident individuals may have opted *against* disclosure, believing they did not need the accommodation.

address underlying fairness norms and stigma at the level of group beliefs. For example, institutions might need to focus on making clear that neurodivergence is accepted and valued, which may prove more effective than reducing formal or economic barriers alone. That said, the sample here was self-selected and willing to reveal sensitive information online, so it remains unclear how far these findings extend to broader populations who may face stronger barriers. Understanding how social norms, image concerns, and fairness perceptions affect real-world health disclosure will be an important next step.

3.7.1 Limitations

As mentioned above, online platforms offer some unique advantages compared to laboratory and field settings, such as access to niche populations (including the one recruited for the present study), lower costs, and reduced social desirability bias, and thus, enhancing internal validity. However, they also introduce well-documented risks to data quality. Self-selection bias, limited control over participants' environments, as well as compromised external validity and/or inattention (Palan and Schitter, 2018) are particularly salient when studying context-dependent behaviors: outcomes measured in an anonymous online setting may not generalize to real-world social contexts like universities or workplaces. Although prior work suggests that online samples represent broader populations in basic demographics and, in the context of this particular study, the prevalence of individuals with ADHD, it remains uncertain whether they meet clinical diagnostic criteria in the same way (Wymbs and Dawson, 2019). Yet there have been prior studies showing that online experiments are capable of reproducing classic behavioral outcomes from the laboratory and field: online participants behave comparably to lab subjects in labor market games (Horton, Rand and Zeckhauser, 2011), public goods settings (Suri and Watts, 2011), and cognitive tasks (Crump, McDonnell and Gureckis, 2013). Moreover, findings pertaining to disclosure rates of sensitive health information in this study may be inflated and, at best, interpreted as setting a clear upper bound on participants' willingness to reveal their condition.

A further limitation concerns the classification of participants in Study 2 as non-neurodivergent. While Study 1 recruited individuals who self-identified as having ADHD, its follow-up study relied on participants replying "No" to the question "Do you consider yourself to be neurodivergent?". However, the term "neurodivergent" may be unfamiliar to some, even among those with relevant diagnoses. For example, a recent YouGov survey found that only 32% of Americans could define the term without help, including just 61% of those who identify as having ADHD themselves (YouGov, 2024).

This raises the possibility that individuals with ADHD or other neurodivergent conditions may have unintentionally been excluded from Study 1 (or incorrectly included in Study 2). To mitigate this concern, I use the ASRS-v1.1 (ADHD) score as a control, and identify potential undiagnosed traits in secondary analyses. Still, some misclassification is likely given the self-reported nature of the data. In addition, without a manipulation check, it is unclear whether participants actually perceived the diagnostic and identity-based framings as intended. The identity-based version may simply have been less salient or less clearly linked to ADHD, which would have narrowed the effective contrast between conditions and, in turn, may help explain the absence of framing effects.

A further concern is the role of the incentive structure embedded in the disclosure decision. Since participants were offered a time advantage to improve their performance at the cost of disclosing their neurodivergent status, their choice may have partially reflected instrumental motivations rather than sensitivity to stigma or social backlash. This performance-based accommodation, while realistic in institutional contexts, introduces a strategic incentive that could obscure the examined mechanisms. However, the study is designed to mitigate this concern in several ways: First, the three-level backlash condition allows for pairwise comparisons that isolate the marginal effects of social rejection risk and economic consequence, allowing for a distinction between pure image concerns (Control vs. Low Backlash) and material self-interest (Low vs. High Backlash). Second, the inclusion of post-task measures on perceived fairness and second-order beliefs provides a means to explore the underlying motivations for disclosure decisions. Third, a set of secondary analyses test whether baseline task performance, ADHD symptom severity, or perceptions of fairness regarding accommodation-based advantages affect disclosure choices. Together, these design features help to reduce potential bias and credibly disentangle fear of stigma from rational, incentive-driven behavior, although some ambiguity may remain.

Similarly, another limitation concerns the selective omission of contextual information from receivers in Study 2. In real-world settings, disclosing a neurodivergent condition such as ADHD is often accompanied by accommodation requests, including extended deadlines or flexible work arrangements. However, in this study, receivers were only exposed to the health-related disclosure itself, without being informed of the time advantage awarded to disclosers. This design choice, however, was intentional as it allowed me to isolate participants' reactions to the disclosure itself, minimizing confounding influences such as fairness-based concerns or envy. Prior research suggests that perceptions of unequal outcomes, whether disadvantageous or advantageous, can incur negative behavioral and emotional responses (such as envy or malice) and thus can hinder the acceptance of

Pareto-efficient outcomes (Beckman *et al.*, 2002; Dal Forno and Merlone, 2021; Fehr and Schmidt, 1999). By omitting information about the accommodation, the study limits these confounds and obtains a cleaner measure of the impact of disclosure and stigma-based discrimination. While this approach enhances the internal validity of the study, generalizability to real-world settings is compromised.

References

Abdelnour, Elie, Madeline O Jansen, and Jessica A Gold. 2022. "ADHD diagnostic trends: increased recognition or overdiagnosis?" *Missouri medicine*, 119(5): 467.

Acquisti, Alessandro, Leslie K John, and George Loewenstein. 2013. "What is privacy worth?" *The Journal of Legal Studies*, 42(2): 249–274.

Attridge, Mark. 2019. "A global perspective on promoting workplace mental health and the role of employee assistance programs." *American Journal of Health Promotion*, 33(4): 622–629.

Ayano, Getinet, Sileshi Demelash, Yitbarek Gizachew, Light Tsegay, and Rosa Alati. 2023. "The global prevalence of attention deficit hyperactivity disorder in children and adolescents: An umbrella review of meta-analyses." *Journal of affective disorders*, 339: 860–866.

Banaschewski, Tobias, Alexander Häge, Sarah Hohmann, and Konstantin Mechler. 2024. "Perspectives on ADHD in children and adolescents as a social construct amidst rising prevalence of diagnosis and medication use." *Frontiers in Psychiatry*, 14: 1289157.

Beckman, Steven R, John P Formby, W James Smith, and Buhong Zheng. 2002. "Envy, malice and Pareto efficiency: An experimental examination." *Social Choice and Welfare*, 19: 349–367.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and prosocial behavior." *American Economic Review*, 96(5): 1652–1678.

Benndorf, Volker, Dorothea Kübler, and Hans-Theo Normann. 2015. "Privacy concerns, voluntary disclosure of information, and unraveling: An experiment." *European Economic Review*, 75: 43–59.

Bicchieri, Cristina. 2005. *The grammar of society: The nature and dynamics of social norms.* Cambridge University Press.

Bicchieri, Cristina, and Erte Xiao. 2009. "Do the right thing: but only if others do so." *Journal of Behavioral Decision Making*, 22(2): 191–208.

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

Brohan, Elaine, Claire Henderson, Kay Wheat, Estelle Malcolm, Sarah Clement, Elizabeth A Barley, Mike Slade, and Graham Thornicroft. 2012. "Systematic review of beliefs, behaviours and influencing factors associated with disclosure of a mental health problem in the workplace." *BMC psychiatry*, 12: 1–14.

Corrigan, Patrick W, and Amy C Watson. 2002. "Understanding the impact of stigma on people with mental illness." *World psychiatry*, 1(1): 16.

Crawford, Vincent P, and Joel Sobel. 1982. "Strategic information transmission." *Econometrica: Journal of the Econometric Society*, 1431–1451.

Criaud, Marion, and Philippe Boulinguez. 2013. "Have we been asking the right questions when assessing response inhibition in go/no-go tasks with fMRI? A meta-analysis and critical review." *Neuroscience & biobehavioral reviews*, 37(1): 11–23.

Crump, Matthew JC, John V McDonnell, and Todd M Gureckis. 2013. "Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research." *PloS one*, 8(3): e57410.

Dal Forno, Arianna, and Ugo Merlone. 2021. "Envy effects on conflict dynamics in supervised work groups." *Decisions in Economics and Finance*, 44(2): 755–779.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness." *Economic Theory*, 33(1): 67–80.

Danz, David, Lise Vesterlund, and Alistair J Wilson. 2024. "Evaluating Behavioral Incentive Compatibility: Insights from Experiments." *Journal of Economic Perspectives*, 38(4): 131–154.

Diaz, Lina, Daniel Houser, John Ifcher, and Homa Zarghamee. 2023. "Estimating social preferences using stated satisfaction: Novel support for inequity aversion." *European Economic Review*, 155: 104436.

Donnelly, Colleen. 2017. "Public attitudes toward disclosing mental health conditions." *Social Work in Mental Health*, 15(5): 588–599.

Drewe, EA. 1975. "Go-no go learning after frontal lobe lesions in humans." *Cortex*, 11(1): 8–16.

Farrell, Joseph, and Matthew Rabin. 1996. "Cheap talk." *Journal of Economic perspectives*, 10(3): 103–118.

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

Godard, Rebecca, Mikki Hebl, and Christine Nittrouer. 2022. "Identity management in the workplace: Coworker perceptions of individuals with contested disabilities." *Journal of Vocational Rehabilitation*, 57(2): 177–186.

Greene, Kathryn, Kate Magsamen-Conrad, Maria K Venetis, Maria G Checton, Zhanna Bagdasarov, and Smita C Banerjee. 2012. "Assessing health diagnosis disclosure decisions in relationships: Testing the disclosure decision-making model." *Health Communication*, 27(4): 356–368.

Grossman, Zachary. 2015. "Self-signaling and social-signaling in giving." *Journal of Economic Behavior & Organization*, 117: 26–39.

Grossman, Zachary, and Joel J Van Der Weele. 2017. "Self-image and willful ignorance in social decisions." *Journal of the European Economic Association*, 15(1): 173–217.

Grummt, Marek. 2024. "Sociocultural perspectives on neurodiversity – An analysis, interpretation and synthesis of the basic terms, discourses and theoretical positions." *Sociology Compass*, 18(8): e13249.

Gupta, Amit, and Pushpendra Priyadarshi. 2020. "When affirmative action is not enough: challenges in career development of persons with disability." *Equality, Diversity and Inclusion: An International Journal*, 39(6): 617–639.

Haghani, Milad, Michiel CJ Bliemer, John M Rose, Harmen Oppewal, and Emily Lancsar. 2021. "Hypothetical bias in stated choice experiments: Part I. Integrative synthesis of empirical evidence and conceptualisation of external validity." *arXiv preprint arXiv:2102.02940*.

Horton, John J, David G Rand, and Richard J Zeckhauser. 2011. "The online laboratory: Conducting experiments in a real labor market." *Experimental economics*, 14: 399–425.

Jaeggi, Susanne M, Martin Buschkuehl, Walter J Perrig, and Beat Meier. 2010. "The concurrent validity of the N-back task as a working memory measure." *Memory*, 18(4): 394–412.

Jastrowski, Kristen E, Kristoffer S Berlin, Amy F Sato, and W Hobart Davies. 2007. "Disclosure of attention–deficit/hyperactivity disorder may minimize risk of social rejection." *Psychiatry*, 70(3): 274–282.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian persuasion." *American Economic Review*, 101(6): 2590–2615.

Kessler, Ronald C, Lenard Adler, Minnie Ames, Olga Demler, Steve Faraone, EVA Hiripi, Mary J Howes, Robert Jin, Kristina Secnik, Thomas Spencer, *et al.* 2005. "The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population." *Psychological medicine*, 35(2): 245–256.

Konow, James. 2000. "Fair shares: Accountability and cognitive dissonance in allocation decisions." *American Economic Review*, 90(4): 1072–1092.

Kroenke, Kurt, Robert L Spitzer, and Janet BW Williams. 2003. "The Patient Health Questionnaire-2: validity of a two-item depression screener." *Medical care*, 41(11): 1284–1292.

Kroenke, Kurt, Robert L Spitzer, Janet BW Williams, Patrick O Monahan, and Bernd Löwe. 2007. "Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection." *Annals of internal medicine*, 146(5): 317–325.

Krupka, Erin L, and Roberto A Weber. 2013. "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association*, 11(3): 495–524.

Lane, Tom. 2016. "Discrimination in the laboratory: A meta-analysis of economics experiments." *European Economic Review*, 90: 375–402.

Levitt, Steven D, and John A List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic perspectives*, 21(2): 153–174.

Liesefeld, Heinrich René, and Markus Janczyk. 2019. "Combining speed and accuracy to control for speed-accuracy trade-offs (?)." *Behavior Research Methods*, 51: 40–60.

Martin, Alex F, G James Rubin, M Brooke Rogers, Simon Wessely, Neil Greenberg, Charlotte E Hall, Angie Pitt, Poppy Ellis Logan, Rebecca Lucas, and Samantha K Brooks. 2025. "The changing prevalence of ADHD? A systematic review." *Journal of Affective Disorders*, 119427.

McGrath, Martina O, Karolina Krysinska, Nicola J Reavley, Karl Andriessen, and Jane Pirkis. 2023. "Disclosure of mental health problems or suicidality at work: A systematic review." *International journal of environmental research and public health*, 20(8): 5548.

McQueen, Amy, Matthew W Kreuter, Bindu Kalesan, and Kassandra I Alcaraz. 2011. "Understanding narrative effects: the impact of breast cancer survivor stories on message processing, attitudes, and beliefs among African American women." *Health psychology*, 30(6): 674.

Mio, Jeffery Scott. 2023. "Employee Assistance Programs." In *Encyclopedia of Mental Health*. Vol. 13rd ed., Eds. Howard S. Friedman and Crystal H. Markey, 761–764. Academic Press, Elsevier.

Morris, Meredith Ringel, Andrew Begel, and Ben Wiedermann. 2015. "Understanding the challenges faced by neurodiverse software engineering employees: Towards a more inclusive and productive technical workforce." 173–184.

Murphy, Kevin, and Patricia Latham. 2022. "Disclosure of ADHD in the workplace: Practical and legal issues to consider." *The ADHD Report*, 30(4): 12–14.

Nyhan, Brendan, and Jason Reifler. 2019. "The roles of information deficits and identity threat in the prevalence of misperceptions." *Journal of Elections, Public Opinion and Parties*, 29(2): 222–244.

O'Connor, Cliodhna, Maryanne Brassil, Sadhbh O'Sullivan, Christina Seery, and Finiki Nearchou. 2022. "How does diagnostic labelling affect social responses to people with mental illness? A systematic review of experimental studies using vignette-based designs." *Journal of Mental Health*, 31(1): 115–130.

Palan, Stefan, and Christian Schitter. 2018. "Prolific. ac—A subject pool for online experiments." *Journal of behavioral and experimental finance*, 17: 22–27.

Phelan, Jo C, Bruce G Link, and John F Dovidio. 2008. "Stigma and prejudice: one animal or two?" *Social science & medicine*, 67(3): 358–367.

Porras Pyland, Claudia, Zachary Zoet, Morrigan Holmes, Hanna Davis, Alannah Shelby Rivers, and Debra Mollen. 2025. "ADHD Identity: How Age, Race, and Socioeconomic Status Shape Rejection or Acceptance." *Identity*, 1–15.

Quinn, Diane M, and Valerie A Earnshaw. 2011. "Understanding concealable stigmatized identities: The role of identity in psychological, physical, and behavioral outcomes." *Social Issues and Policy Review*, 5(1): 160–190.

Ridley, Matthew. 2022. "Essays on the Economics of Mental Illness and Belief Formation." PhD diss. Massachusetts Institute of Technology.

Santuzzi, Alecia M, Pamela R Waltz, Lisa M Finkelstein, and Deborah E Rupp. 2014. "Invisible disabilities: Unique challenges for employees and organizations." *Industrial and organizational Psychology*, 7(2): 204–219.

Shulman, Yefim, Agnieszka Kitkowska, Mark Warner, and Joachim Meyer. 2024. "Conceal or reveal:(non) disclosure choices in online information sharing." *Behaviour & Information Technology*, 43(16): 4125–4149.

Song, Peige, Mingming Zha, Qingwen Yang, Yan Zhang, Xue Li, and Igor Rudan. 2021. "The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis." *Journal of global health*, 11: 04009.

Spence, Michael. 1973. "Job Market Signaling." *The Quarterly Journal of Economics*, 87(3): 355–374.

Suri, Siddharth, and Duncan J Watts. 2011. "Cooperation and contagion in web-based, networked public goods experiments." *ACM SIGecom Exchanges*, 10(2): 3–8.

Thompson-Hodgetts, Sandra, Chantal Labonte, Rinita Mazumder, and Shanon Phelan. 2020. "Helpful or harmful? A scoping review of perceptions and outcomes of autism diagnostic disclosure to others." *Research in Autism Spectrum Disorders*, 77: 101598.

Wu, Wenhao. 2019. "Persuasive Disclosure." Working paper.

Wymbs, Brian T, and Anne E Dawson. 2019. "Screening Amazon's Mechanical Turk for adults with ADHD." *Journal of Attention Disorders*, 23(10): 1178–1187.

YouGov. 2024. "Neurodiversity in the U.S.: 19% of Americans identify as neurodivergent." Accessed: 2025-05-28.

Appendix

A3.1 Data quality and weighting

A3.1.1 Flagged observations

Inclusion of participants with subthreshold ADHD scores. Participants in Study 1 were pre-screened on Prolific based on self-reported ADHD status ("Do you consider yourself to have attention deficit disorder (ADD)/attention deficit hyperactivity disorder (ADHD)?"). As an additional control, I administered the short Adult ADHD Self-Report Scale (ASRS-v1.1). Of the 659 participants, only 223 scored above the common threshold (score ≥ 4) used to flag potential ADHD and warrant further clinical assessment Kessler *et al.* (2005). While this gap may raise concerns about internal validity, the ASRS is a screening tool, not a diagnostic instrument. ADHD cannot be meaningfully reduced to six symptom items, and self-identification may still reflect relevant cognitive experiences and social realities. To account for variation in symptom severity, I include ASRS scores as covariates in all regressions (see Appendix Table A3.4).

Inclusion of participants failing comprehension checks. In both studies, all participants were required to pass at least one comprehension check before proceeding with the main experiment[20]. To preserve power and avoid post-treatment bias, participants who failed to answer correctly on their first attempt were not excluded. However, due to high failure rates (Table A3.1), I constructed a composite indicator of reduced attentiveness to flag those who failed the comprehension check *and* showed signs of disengagement during the performance tasks, including poor Go/No-Go task accuracy, high miss rates in N-back tasks, large accuracy swings across N-back rounds, and extreme reaction times. Based on these criteria, I estimated inverse probability weights (IPW), modeling the propensity of being included under stricter criteria.

Table A3.1: Sample overview and flagged observations

	Study 1		Study 2	
Criterion	N	% of sample	N	% of sample
Total	659	100.0	749	100.0
ASRS-v1.1 score below threshold (score < 4)	435	66.0	679	90.7
Failed comprehension check (first attempt)	333	50.5	404	53.9
High N-back miss rate	491	74.5	524	70.0
Poor Go/No-Go accuracy	132	20.0	143	19.1
Unstable N-back accuracy	19	2.9	30	4.0
RT outlier (any)	7	1.1	13	1.7
2+ flags	320	48.6	356	47.5

Notes: Summary of sample characteristics and flagged observations. The table reports the number and percentage of participants flagged by each criterion.

[20]Procedures are detailed in Sections 3.2.3 and 3.3.3. Appendix Table A3.19 shows question wording and response options.

A3.1.2 Inverse probability weighting

To adjust for potential bias from dropping participants flagged for poor performance or comprehension, I used inverse probability weighting (IPW) to reweight the analysis sample so it better reflected the full recruited sample. I estimated a logistic regression predicting inclusion based on pre-treatment variables (demographics, ADHD symptoms, mental health, and baseline task performance) as well as treatment assignment to ensure balance. Weights were stabilized using the marginal probability of inclusion and truncated at the 1st and 99th percentiles to reduce the influence of outliers. Figure A3.1 shows that the weight distribution was right-skewed, ranging from 0.53 to 3.59.

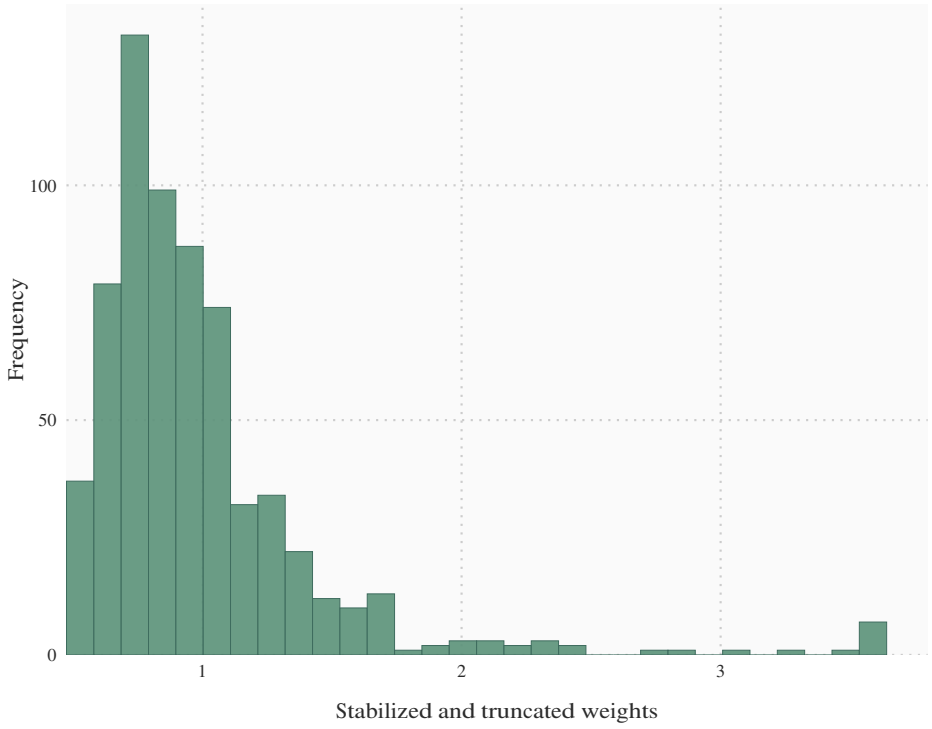


Figure A3.1: Histogram of inverse probability weights

A3.1.3 Treatment assignment and participant responses

Table A3.2: Player 2 assignment to treatment arms and corresponding responses

Treatment	Disclosure & framing	n	Reject$_i = 1$	%
Control	No disclosure	73	0	0.0
	Disclosure: Diagnostic	90	0	0.0
	Disclosure: Identity	82	0	0.0
Low Backlash	No disclosure	82	2	2.4
	Disclosure: Diagnostic	82	14	17.1
	Disclosure: Identity	82	4	4.9
High Backlash	No disclosure	78	2	2.5
	Disclosure: Diagnostic	82	7	8.5
	Disclosure: Identity	98	6	6.1

Notes: Treatment reflects the combination of backlash condition and message framing, conditional on whether Player 1 disclosed. In Study 1, each treatment arm (Backlash × Framing × Disclosure) was embedded in a unique Prolific survey link with a matched ID, ensuring that Player 2's assigned condition matched that of Player 1. Because disclosure was endogenous, Player 2 cell sizes are not perfectly balanced. n reports the number of Player 2 participants in each condition. Reject$_i = 1$ if Player 2 rejected Player 1 and requested reassignment to another participant. The last column reports the percentage of Player 2s who rejected their partner within each treatment condition.

A3.2 Robustness and sensitivity analyses

A3.2.1 Model fit assessment for covariates

Table A3.3: Logit models: examining covariate inclusion and model fit

Panel A. Unadjusted model

	OR	95% CI	*p*-value	AME	SE
Treatment conditions (ref: Control)					
Low Backlash	0.91	0.51, 1.63	0.762	-0.023	0.044
High Backlash	1.08	0.61, 1.91	0.793	-0.008	0.044
Diagnostic framing (ref: Identity)	1.11	0.62, 1.99	0.730	0.003	0.036
Interactions					
Low Backlash × Diagnostic	0.97	0.43, 2.17	0.931		
High Backlash × Diagnostic	0.80	0.35, 1.81	0.589		

Notes: Logistic regression of disclosure on experimental treatment conditions and their interactions. No covariates included. Odds ratios (ORs), 95% confidence intervals (CIs), average marginal effects (AMEs), and standard errors (SEs) reported. Model fit: AIC = 830.05. $N = 659$.

Panel B. Adjusted model with selected covariates

	OR	95% CI	*p*-value	AME	SE
Treatment conditions (ref: Control)					
Low Backlash	0.92	0.51, 1.65	0.775	-0.028	0.044
High Backlash	1.08	0.60, 1.93	0.797	-0.004	0.044
Diagnostic framing (ref: Identity)	1.15	0.63, 2.09	0.642	0.009	0.036
Interactions					
Low Backlash × Diagnostic	0.91	0.40, 2.08	0.825		
High Backlash × Diagnostic	0.82	0.36, 1.88	0.641		

Notes: Logistic regression of disclosure on experimental treatment conditions and their interactions. Covariates include age, gender, employment, mental health, ADHD, and Go/No-Go task performance (coefficients not shown). Model fit: AIC = 835.33. $N = 659$.

Panel C. Adjusted model with full set of covariates

	OR	95% CI	*p*-value	AME	SE
Treatment conditions (ref: Control)					
Low Backlash	0.97	0.53, 1.75	0.907	-0.020	0.044
High Backlash	1.12	0.62, 2.03	0.695	-0.002	0.044
Diagnostic framing (ref: Identity)	1.21	0.66, 2.22	0.533	0.013	0.037
Interactions					
Low Backlash × Diagnostic	0.88	0.38, 2.03	0.767		
High Backlash × Diagnostic	0.77	0.33, 1.78	0.542		

Notes: Logistic regression of disclosure on experimental treatment conditions and their interactions. Full set of covariates included: age, gender, education, employment, ethnicity, general health, mental health, ADHD, stimulant use, and BIS (Go/No-Go, N-back) (coefficients not shown). Model fit: AIC = 847.74. $N = 659$.

A3.2.2 Interaction: ADHD × treatment condition

Table A3.4: Logit models of disclosure by ADHD and condition

Panel A. Baseline model

	(1)			(2)		
	OR	95% CI	p-value	OR	95% CI	p-value
ADHD score	1.05	0.97, 1.15	0.24	1.10	0.78, 1.57	0.59
Low Backlash (ref: Control)	0.90	0.50, 1.61	0.72	0.91	0.51, 1.63	0.75
High Backlash (ref: Control)	1.07	0.60, 1.91	0.81	1.08	0.61, 1.91	0.80
Diagnostic framing (ref: Identity)	1.10	0.61, 1.97	0.75	1.11	0.62, 1.99	0.72
Low Backlash × Diagnostic	0.99	0.44, 2.24	0.98	0.96	0.43, 2.17	0.93
High Backlash × Diagnostic	0.82	0.36, 1.85	0.63	0.80	0.36, 1.82	0.60

Notes: Logistic regressions of disclosure on ADHD score and treatment conditions. Model (1) uses mean-centered ADHD scores; Model (2) uses a binary ADHD indicator (score ≥ 4). No covariates included. $N = 659$.

Panel B. Interaction model

	(1)			(2)		
	OR	95% CI	p-value	OR	95% CI	p-value
ADHD score	0.97	0.82, 1.15	0.70	0.76	0.39, 1.49	0.42
Low Backlash (ref: Control)	0.88	0.58, 1.33	0.55	0.65	0.39, 1.07	0.09
High Backlash (ref: Control)	0.97	0.64, 1.47	0.90	1.10	0.66, 1.83	0.72
Diagnostic framing (ref: Identity)	1.05	0.75, 1.48	0.76	1.02	0.73, 1.44	0.90
Low Backlash × ADHD score	1.18	0.95, 1.47	0.14	2.47*	1.02, 6.05	0.05
High Backlash × ADHD score	0.99	0.79, 1.23	0.92	0.73	0.30, 1.74	0.48

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Logistic regressions including covariates and interaction terms. Model (1) uses the mean-centered ADHD score; Model (2) uses the binary indicator. Controls: age, gender, employment, mental health, and Go/No-Go performance (coefficients not shown). $N = 659$.

Panel C. Interaction model (IPW)

	(1)			(2)		
	OR	95% CI	p-value	OR	95% CI	p-value
ADHD score	0.98	0.81, 1.18	0.82	0.73	0.36, 1.49	0.38
Low Backlash (ref: Control)	0.71	0.45, 1.10	0.13	0.49**	0.29, 0.85	0.01
High Backlash (ref: Control)	0.88	0.56, 1.40	0.60	0.92	0.52, 1.63	0.79
Diagnostic framing (ref: Identity)	0.93	0.64, 1.36	0.71	0.90	0.61, 1.31	0.57
Low Backlash × ADHD score	1.22	0.97, 1.54	0.09	2.84**	1.11, 7.27	0.03
High Backlash × ADHD score	1.00	0.78, 1.26	0.97	0.90	0.35, 2.30	0.82

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Inverse probability weighted (IPW) logistic regressions with the same covariates and interactions as Panel B. Model (1) uses the mean-centered ADHD score; Model (2) uses the binary ADHD variable (score ≥ 4). $N = 659$.

A3.2.3 Social preference clusters

Table A3.5: Disclosure across social preference profiles (IPW)

	(1)			(2)		
	OR	95% CI	p-value	OR	95% CI	p-value
Cluster (ref: 3: Self-interest)						
1: Guilt	0.61	0.36, 1.03	0.07	0.57*	0.34, 0.95	0.03
2: Envy	0.90	0.53, 1.53	0.69	0.88	0.52, 1.49	0.64
4: Positional concerns	0.65	0.37, 1.13	0.13	0.65	0.37, 1.13	0.13
Constant	2.74***	1.83, 4.09	< 0.001	1.97	0.79, 4.88	0.14
Controls		N			Y	
Observations		659			659	

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Inverse probability weighted (IPW) logistic regressions. The dependent variable is disclosure. OR = Odds Ratio; CI = Confidence Interval. Model (1) includes only social preference type (cluster); Model (2) adjusts for experimental condition and framing, age, gender, employment, mental health, ADHD, and Go/No-Go performance (coefficients not reported).

A3.2.4 Second-order beliefs and disclosure: predicted probabilities

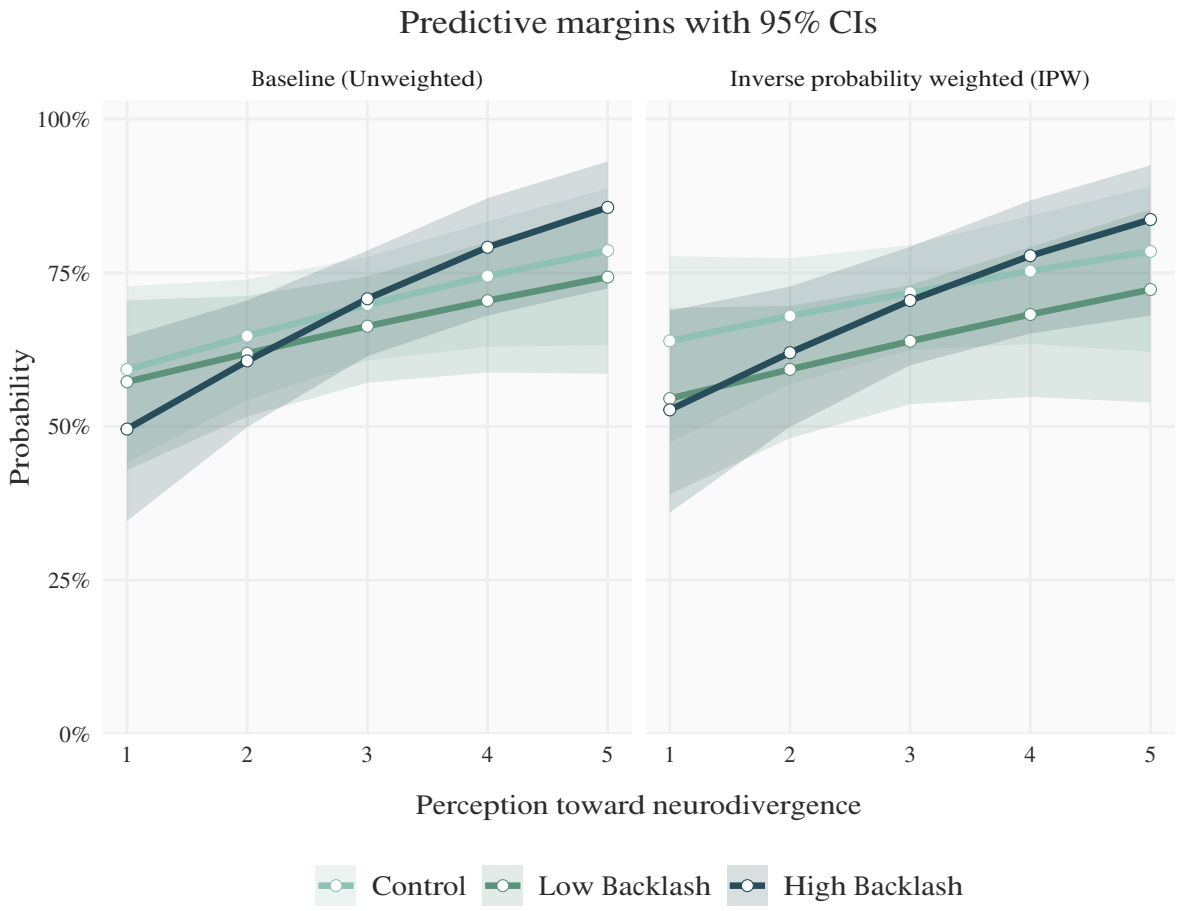


Figure A3.2: Predicted probabilities of disclosure as a function of backlash condition and second-order beliefs toward neurodivergent traits

Figure A3.2 illustrates the marginal effects and 95% confidence intervals for each backlash condition (Control, Low Backlash, High Backlash) across the observed range of second-order beliefs pertaining to others' perceptions of neurodivergence (e.g., ADHD, autism). Responses were given on a 5-point Likert scale (1 = *Very negative*, 5 = *Very positive*). The left panel reports estimates from the unweighted baseline model; the right panel shows predictions from the IPW model that adjusts for potential selection on task performance and comprehension. In both models, more positive beliefs appear associated with a greater likelihood of disclosure, though uncertainty widened under IPW.

A3.2.5 Second-order beliefs and disclosure: regression results

Table A3.6: Second-order beliefs and probability of disclosure

Panel A. Main Effects Model

	OR	95% CI	*p*-value	AME	SE
SOB (ND)	1.33***	1.13, 1.56	< 0.001	0.059***	0.017
Treatment conditions (ref: Control)					
Low Backlash	0.86	0.57, 1.30	0.48	−0.031	0.044
High Backlash	0.96	0.63, 1.45	0.84	−0.009	0.044
Observations	659			659	

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Logistic regression of disclosure on second-order beliefs toward neurodivergence and treatment conditions. Covariates include message framing, age, gender, employment status, mental health, ADHD symptoms, and Go/No-Go task performance (coefficients not shown). Odds ratios (ORs), 95% confidence intervals (CIs), average marginal effects (AMEs), and standard errors (SEs) reported.

Panel B. Interaction model

	OR	95% CI	*p*-value	AME	SE
SOB (ND)	1.26	0.96, 1.68	0.10	0.060***	0.016
Treatment conditions (ref: Control)					
Low Backlash	0.96	0.33, 2.81	0.94	−0.031	0.044
High Backlash	0.54	0.18, 1.66	0.29	−0.010	0.043
Interactions					
Low Backlash × SOB (ND)	0.96	0.66, 1.41	0.84		
High Backlash × SOB (ND)	1.24	0.83, 1.87	0.29		
Observations	659			659	

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Covariates and outcome specification match Panel A, with added interaction terms.

Panel C. Interaction model (IPW)

	OR	95% CI	*p*-value	AME	SE
SOB (ND)	1.20	0.88, 1.62	0.243	0.053**	0.018
Treatment conditions (ref: Control)					
Low Backlash	0.67	0.21, 2.17	0.503	−0.078	0.048
High Backlash	0.51	0.15, 1.74	0.285	−0.030	0.048
Interactions					
Low Backlash × SOB (ND)	1.01	0.67, 1.54	0.950		
High Backlash × SOB (ND)	1.22	0.79, 1.89	0.366		
Observations	659			659	

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: Inverse probability weighted (IPW) logistic regression with the same covariates and interactions as Panel B.

A3.3 Additional results

A3.3.1 Framing effects on rejection (Study 2)

Table A3.7: Effect of message framing on rejection by Player 2 (by subsample)

	(1)	(2)
Panel A: Unweighted logistic regression		
Diagnostic framing (ref: Identity-based)	4.11*	1.37
	[1.45, 13.63]	[0.46, 4.16]
Panel B: IPW-weighted logistic regression		
Diagnostic framing (ref: Identity-based)	4.55*	1.17
	[1.35, 15.31]	[0.39, 3.51]
Controls	Y	Y
Observations	246	258

Significance level: * $p < 0.05$.
Notes: Subsamples correspond to high vs. low backlash groups. Coefficients are estimated from logistic regressions, reported as odds ratios (OR) with 95% confidence intervals in brackets. Panel A presents unweighted estimates; Panel B presents inverse probability weighted (IPW) estimates. Model (1) pertains to Player 2s assigned to the Low Backlash condition. Model (2) pertains to those in the High Backlash condition. Controls: age, gender, ethnicity, employment, education, general and mental health, ADHD, performance BIS (Go/No-Go, N-back), substance use (coefficients not shown).

A3.3.2 Bonus task performance

Table A3.8: Odds of disclosure by baseline BIS performance

	(1)			(2)			(3)		
	OR	95% CI	*p*	OR	95% CI	*p*	OR	95% CI	*p*
Go/No-Go BIS	0.83	0.69, 0.99	0.036*	0.83	0.69, 0.99	0.033*	0.82	0.69, 0.98	0.028*
N-back BIS	1.00	0.84, 1.19	0.975	1.00	0.84, 1.19	0.957	1.00	0.84, 1.19	0.993
Controls		N			Y			Y	
Observations		659			659			659	

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Note: Logistic regression models predicting disclosure. Balanced Integration Scores (BIS) are standardized (z-transformed). Model (1) includes BIS predictors only; Model (2) adds experimental conditions (treatment and framing); Model (3) includes the full set of controls: experimental conditions, age, gender, employment, mental health, and ADHD (coefficients not reported). OR = Odds Ratio; CI = Confidence Interval.

Table A3.9: Effect of disclosure on bonus task performance

	Bonus BIS	Accuracy	Log RT	Miss rate (%)
	(1)	(2)	(3)	(4)
Disclosure (1 = Yes)	1.900***	0.048*	–0.927***	12.46***
	[1.67, 2.13]	[0.01, 0.09]	[–1.04, –0.81]	[8.74, 16.18]
Go/No-Go BIS	0.251***	0.036***	–0.060*	–3.13***
	[0.16, 0.35]	[0.02, 0.05]	[–0.11, –0.01]	[–4.66, –1.59]
N-back BIS	0.391***	0.069***	–0.014	–5.48***
	[0.34, 0.45]	[0.06, 0.08]	[–0.04, 0.01]	[–6.36, –4.60]
Controls	Y	Y	Y	Y
Observations	659	659	659	659

95% Confidence intervals in brackets. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Notes: OLS regressions predicting bonus task performance. Outcome variables reflect post-treatment performance only. "Log RT" is the natural log of average response time (ms). Go/No-Go and N-back BIS are based on pre-treatment tasks. Models include controls for treatment condition, message framing, and bonus task type (coefficients not reported). BIS = Balanced Integration Score (standardized accuracy – standardized RT). "Miss rate" is the percentage of trials with no response.

A3.4 Social preference clustering

A3.4.1 Cluster composition, labels and theoretical mapping

Clusters were derived using K-means clustering on standardized satisfaction ratings from nine distributional choice scenarios (inequity list profiles). Cluster labels in Table A3.10 are assigned based on the best-fitting satisfaction profiles described in Diaz *et al.* (2023) and are illustrated in Figure A3.3. Following Diaz *et al.*'s approach, I used scenario-level satisfaction patterns to infer individual distributional preferences. While their original design featured a broader set of scenarios, I simplified the task to nine conditions to reduce participant burden. For a comprehensive review of pattern nomenclature equivalence and alternative models, see Diaz *et al.* (2023). Note that due to its bimodal structure, Cluster 4 does not map directly onto any of the types proposed by Diaz *et al.* (2023). Instead, it seems to best fit a 'rank sensitive' type with positional concerns, suggesting that what matters is not just how much one earns, but how much one earns relative to others: satisfaction is relatively low except at equality or when the respondent is strongly advantaged. See Chapter 2 for further discussion on positional concerns.

Table A3.10: Cluster descriptions based on satisfaction profiles

Cluster	Label	Description	Diaz *et al.* (2023)
Cluster 1	Guilt	Satisfaction is lowest when the respondent earns more and increases as the partner's earnings rise, peaking at equality. Moderate decline in satisfaction as disadvantageous inequality grows.	F&S guilt
Cluster 2	Envy	Satisfaction is high as long as the respondent is better off, and drops sharply once the partner earns more. Strong aversion to disadvantageous inequality.	F&S envy
Cluster 3	Self-interest	Satisfaction remains consistently high across all allocation scenarios, with little to no concern for relative outcomes.	Self-regarding (F&S no guilt, no envy)
Cluster 4	Positional concerns	Satisfaction ratings are relatively low across all allocation scenarios, with further decline as the partner's payoff increases. Bimodal satisfaction profile: satisfaction peaks when the respondent is either strongly advantaged or at equality.	Capped Relative Income Effect (RIE), mild F&S envy-and-guilt (inequality aversion)

Table A3.11: Cluster sizes and percentages (final sample)

Cluster	n	%
1: Guilt	186	28.2
2: Envy	186	28.2
3: Self-interest	155	23.5
4: Positional concerns	132	20

A3.4.2 Mean satisfaction profiles by cluster

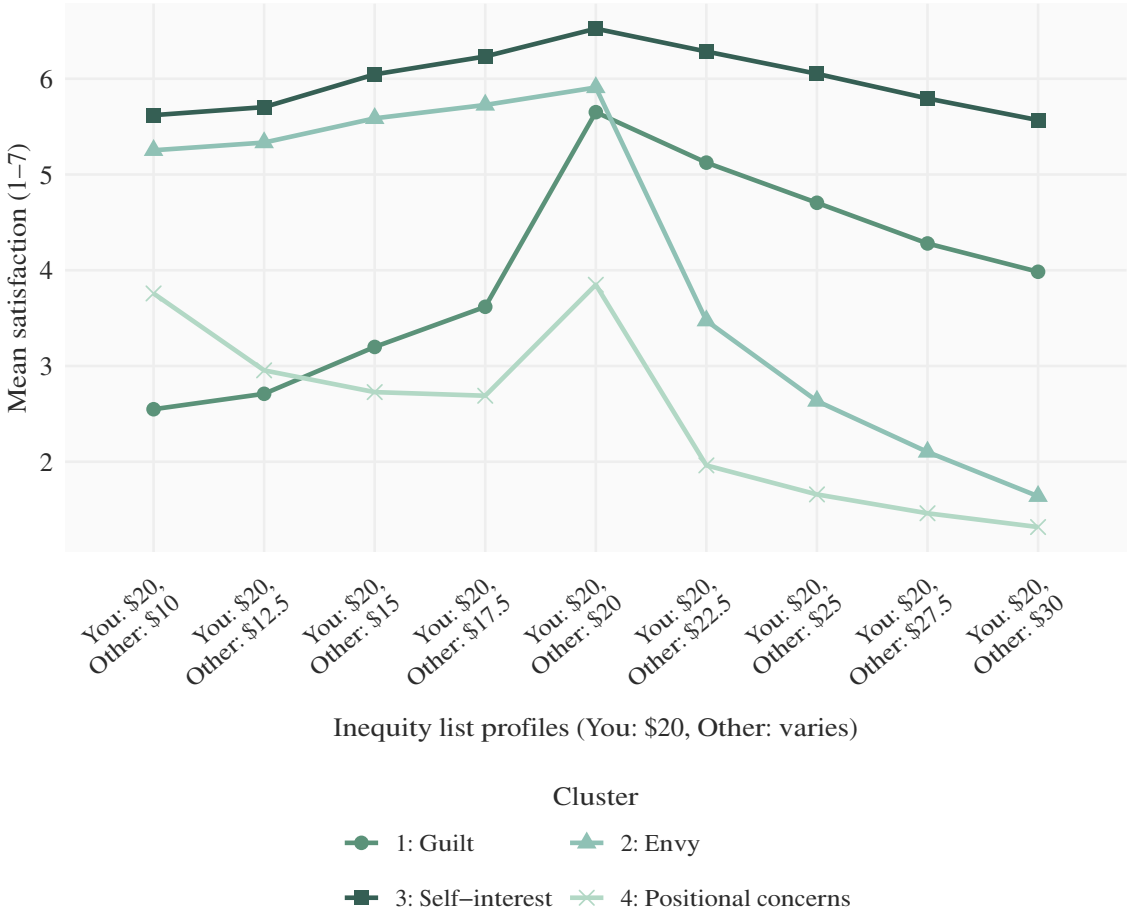


Figure A3.3: Mean satisfaction profiles across distributional choice scenarios

Table A3.12: Mean satisfaction ratings by distributional scenario and cluster

Cluster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
1	2.55	2.71	3.20	3.62	5.65	5.12	4.70	4.28	3.98
2	5.25	5.33	5.59	5.73	5.91	3.47	2.63	2.10	1.64
3	5.62	5.70	6.05	6.23	6.52	6.28	6.05	5.79	5.57
4	3.76	2.95	2.73	2.69	3.85	1.96	1.66	1.46	1.32

Notes: Ratings were based on an inequity list (based on Diaz *et al.*, 2023). "Q1–Q9" correspond to satisfaction with nine allocations in which the participant always received $20, while the partner's amount varied from $10 to $30. Responses were given on a 7-point Likert scale (1 = *Extremely dissatisfied*, 7 = *Extremely satisfied*). Values correspond to those presented in Figure A3.3.

A3.5 Experimental instructions and survey instruments

Part 1: Demographic background and well-being

Table A3.13: Pre-treatment survey questions

Question item	Options and coding
What is your age?	*Open text entry from 18 to 65 years*
What is your gender?	• Female (1) • Male (2) • Non-binary (3)
Which of the following best describes your ethnicity?	• American Indian/Alaska Native (1) • Asian (2) • Black/African American (3) • Hispanic/Latino (4) • Middle Eastern/North African (5) • Native Hawaiian/Pacific Islander (6) • White (7)
What is the highest degree or level of school you have completed?	• Less than secondary education (1) • High school diploma/equivalent (2) • Some college, no degree (3) • Associate's degree (4) • Bachelor's degree (5) • Master's degree (6) • Professional degree (7) • Doctoral degree (8)
What is your current employment status?	• Employed full-time (1) • Employed part-time (2) • Unemployed, seeking work (3) • Not working, not seeking work (4) • Retired (5) • Student (6)
In general, would you say your health is:	• Excellent (1) • Very good (2) • Good (3) • Fair (4) • Poor (5)
In the last 2 weeks, how often were you bothered by:	• Little interest or pleasure (1) • Feeling down/depressed (2) • Feeling nervous/anxious (3) • Unable to control worrying (4) Options: Not at all (0) | Several days (1) | More than half the days (2) | Nearly every day (3)
In the past 2 weeks, how often were you bothered by:	• Trouble finishing tasks (1) • Disorganized with tasks (2) • Forgetting obligations (3) • Avoiding complex tasks (4) • Fidget/squirm when sitting (5) • Overactive, driven by a motor (6) Options: Never (0) | Rarely (1) | Sometimes (2) | Often (3) | Very often (4)
Have you taken anything today (e.g., medication, substances) that might have influenced your focus, attention, or reaction speed?	*Presented after the performance tasks, before the pairing information.* • Yes, I took prescription medication that may affect attention or focus. (1) • Yes, I consumed caffeine or other stimulants (e.g., coffee, energy drinks). (2) • No, I did not take anything that would affect my attention or reaction time. (3) • Prefer not to say. (4)

Part 2: Performance tasks

The Go/No-Go and N-back tasks in Study 1 and Study 2 were programmed in JavaScript and implemented via Qualtrics. Both tasks are well-established paradigms from cognitive psychology (e.g., Criaud and Boulinguez, 2013; Drewe, 1975; Jaeggi *et al.*, 2010), often used to measure response inhibition, impulsivity, and working memory.

Go/No-Go task (Example Round 1). Sequences of stimuli were individually generated in the participant's browser at the start of each round. In the first round, for each of the 35 trials, a random number between 0 and 9 was generated and assigned to either a *Go* or *No-Go* condition, where the color of a stimulus indicated the condition (blue = Go, orange = No-Go). If a Go trial was shown, the participant was instructed to press a button, and withhold otherwise. Figure A3.4 shows the screenshot of an actual example for how stimuli were presented during the experiment. In round 1, the stimulus would appear for 1500 ms, followed by a 650 ms delay before the next stimulus was presented. If no response was given within the 1500 ms window, the trial was coded as a miss (for a Go trial) or a correct rejection (for a No-Go trial). The probability of a Go trial to appear in round 1 was 70%.

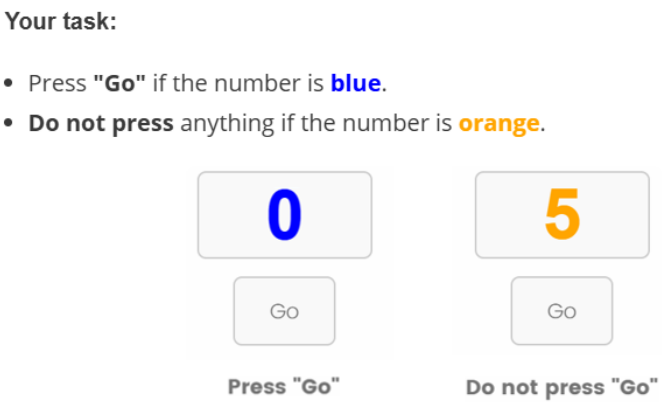


Figure A3.4: Screenshot of Go/No-Go (GNG) instruction

Table A3.14: Design specifications for Go/No-Go (GNG) tasks

Parameter	Practice	Round 1	Round 2	Round 3	Bonus (Extra)	Bonus (Standard)
Total duration limit (ms)	–	60000	60000	60000	75000	60000
Number of trials	12	35	40	45	40	40
Trial duration (ms)	2000	1500	400	350	600	400
Inter-trial delay (ms)	650	650	650	650	650	600
Go/No-Go ratio	70% Go	70% Go	60% Go	50% Go	60% Go	60% Go
Stimuli type	0–9	0–9	0–9	0–9	0–9	0–9
Stimuli colors	Blue/Orange	Blue/Orange	Blue/Orange	Blue/Orange	Blue/Orange	Blue/Orange
Feedback shown	Yes	No	No	No	No	No

N-back task (Example Round 1). For each of the 40 trials, a random stimulus was drawn from a set of 10 uppercase letters (A–J). The probability of a target to appear was 30%. In the 1-back task, targets repeated the letter shown one trial before the actual target would appear; distractors were other randomly generated letters that did not match the target letter. Figure A3.5 shows how a stimulus was presented during the first practice round. The stimulus would appear for 1500 ms, followed by a 600 ms delay before the next was presented. Participants responded using two buttons, which were disabled until a stimulus appeared. If no response was recorded before the 1500 ms timeout, the trial was coded as a miss.

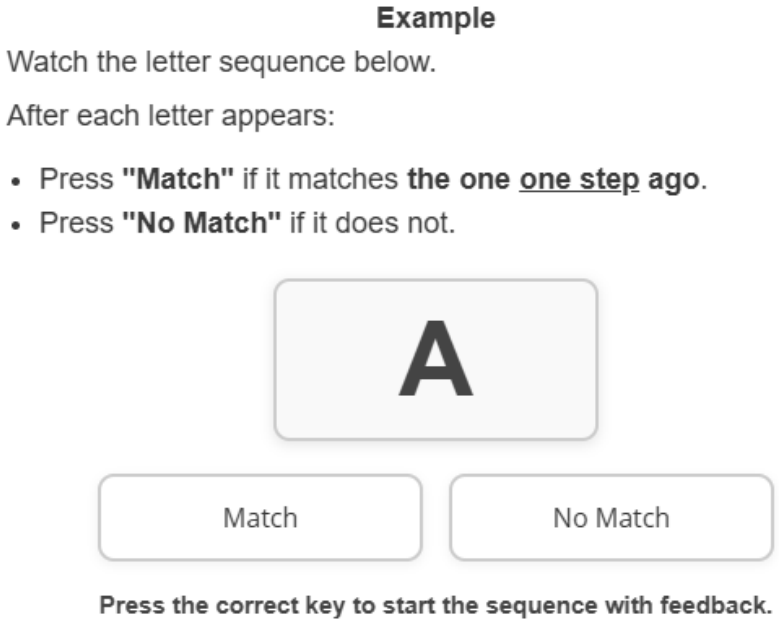


Figure A3.5: Screenshot of N-back practice instruction

Table A3.15: Design specifications for N-back tasks

Parameter	Practice 1	Practice 2	Round 1	Round 2	Round 3	Bonus (Extra)	Bonus (Standard)
N-back level	1	2	1	2	3	2	2
Total duration limit (ms)	20000	20000	60000	65000	65000	80000	65000
Number of trials	10	12	40	50	50	50	50
Trial duration (ms)	2000	2000	1500	1500	1500	1500	1500
Delay (ms)	600	600	600	600	600	650	650
Target probability	40%	40%	30%	25%	25%	25%	25%
Stimuli set	A–E	A–E	A–J	A–J	A–J	A–J	A–J
Feedback shown	Yes	Yes	No	No	No	No	No

Part 3.1: Backlash condition, framing and decision tasks

Table A3.16: Player 1 instructions

Item	Instructions and coding
Instructional screen	This is Part 3 of 4. You will now complete a final task that determines your bonus payment. The bonus can range from \$0.40 to \$1.50, and in rare cases, \$0. The exact calculation will be explained just before the task begins. However, your bonus will also depend on something else: After the experiment ends, you will be matched with another participant. Matches will be made after all responses are collected, based on choices made during the study. Here is what your matched partner will learn or be able to do: Receive some information about you (you will see what this is in a few minutes); Be told that they will earn the same bonus as you; In some conditions, be able to accept or reject the match. You will learn about your specific matching condition on the next page.
Conditional variant	Your partner will be assigned automatically by the computer.
Control	*Displayed only in control group* You will not be able to choose or reject this assignment. Neither will your partner. This means your partner's opinion will not affect whether you remain matched. Next, an update based on your earlier choices will be presented. It relates to your upcoming bonus task.
Low Backlash	*Displayed only in Low Backlash group* Your assigned partner will have the option to accept or reject the match. This means that: If your partner accepts, you will stay matched until the end of the study. If they reject you, their decision will only be implemented in 20% of cases. In the remaining 80%, the match will remain in place as if no rejection occurred. If a rejection is implemented, you will be randomly assigned a new partner. This will not affect your bonus beyond the matching process. Next, an update based on your earlier choices will be presented. It relates to your upcoming bonus task.
High Backlash	*Displayed only in High Backlash group* Your assigned partner will have the option to accept or reject the match. This means that: If your partner accepts, you will stay matched until the end of the study. If they reject you, their decision will only be implemented in 20% of cases. In the remaining 80%, the match will remain in place as if no rejection occurred. If a rejection is implemented, you will not be reassigned. In that case, you will not complete the final bonus task. You will still receive your base participation payment and, if applicable, the earlier performance-based bonus you earned during the first tasks. Next, an update based on your earlier choices will be presented. It relates to your upcoming bonus task.
Eligibility message preceding framing and disclosure decision	Based on your earlier responses, you are eligible to receive extra time in the upcoming task. This extra time may improve your performance by increasing your number of correct responses – and boosting your chances of earning a bonus of up to \$1.50. How is the bonus calculated? Your partner's performance in the final task is compared to others. The amount depends on how well your partner performed: Top 5% = \$1.50; Next 20% = \$1.20; Middle 20% = \$0.90; Next 20% = \$0.65; Bottom 35% = \$0.40. Please note: Choosing extra time comes with one condition. Your assigned partner will receive some information about you. On the next screen, you will see what would be shared. You can then choose to use the extra time or continue without it.
Conditional variant (Framing)	If you choose to receive extra time to boost your performance, the following information about you will be shared with your assigned partner: • You are another study participant who completed the same series of three tasks. • You will complete one additional round of a randomly selected task. Your performance in that round will determine the final bonus for both of you.
Diagnostic	*Additionally displayed only in diagnostic framing* • You reported symptoms that are commonly associated with a clinical ADHD diagnosis.
Identity	*Additionally displayed only in identity-based framing* • You fall within a category reflecting neurodivergent cognitive traits, often described as neurodivergence.
Disclosure decision	Your partner will not learn that you received an advantage. Would you like to receive extra time? • Yes, I choose the time advantage. (1) • No, I continue without the time advantage. (2)

Table A3.17: Player 2 instructions

Item	Instructions and coding
Instructional screen	This is Part 2 of 3. For this part of the study, you have been pre-matched with another participant based on responses collected earlier in the experiment. On the next screen, you will learn more about your matched partner and how your final bonus will be determined. Please read the information carefully – your decision may influence your partner's outcome as well as your own. This decision will be implemented after the study ends, based on all participants' responses.
Profile preceding decision; *identical across treatment cells.*	Your partner's profile: Here is some additional information about the participant who was pre-assigned to you earlier in the study: • Your partner completed the same three tasks you completed earlier. • To qualify for an additional bonus, your partner completed one more task (about 1.5 minutes), similar to the earlier ones. • You will receive the same bonus amount they earned — up to \$1.50, depending on their performance.
Conditional variant	
Diagnostic	*Displayed only in diagnostic framing* • Your partner reported symptoms that are commonly associated with a clinical ADHD diagnosis.
Identity	*Displayed only in identity-based framing* • Your partner falls within a category reflecting neurodivergent cognitive traits, often described as neurodivergence.
Bonus	How is the bonus calculated? Your partner's performance in the final task is compared to others. The amount depends on how well your partner performed: Top 5% = \$1.50; Next 20% = \$1.20; Middle 20% = \$0.90; Next 20% = \$0.65; Bottom 35% = \$0.40.
Conditional variant	Your partner will be assigned automatically by the computer.
Control	*No choice. Proceed immediately to the comprehension check.*
Low Backlash	*Displayed only in Low Backlash group* Please note: If you choose to reject this partner and want to be reassigned, your decision will only be implemented in 20% of cases. In the other 80%, you will remain matched with the same participant. If your rejection is implemented, you will be reassigned to a new partner. You will not receive any information about this new match but will receive the bonus they earned. Your original match will also be reassigned to a new partner. Your decision will not affect their outcome or the bonus they receive.
High Backlash	*Displayed only in High Backlash group* Please note: If you choose to reject this partner and want to be reassigned, your decision will only be implemented in 20% of cases. In the other 80%, you will remain matched with the same participant. If your rejection is implemented, you will be reassigned to a new partner. You will not receive any information about this new match but will receive the bonus they earned. Your original match will not receive the associated bonus payment, even though they already completed the final bonus task.
Rejection decision	Would you like to stay matched with this participant? • Yes, stay matched; receive the same bonus they earned. (1) • No, be reassigned; implemented in 20% of cases. (2)

Part 3.2: Comprehension and manipulation checks

Table A3.18: Post-treatment comprehension checks

Question item	Options and coding
Study 1 only	
How is your bonus amount determined?	• All three tasks contribute to the bonus payment. (1) • One randomly selected task will determine the bonus payment for both you and your assigned partner. (2) • Only your partner's performance determines the bonus payment. (3)
What will your partner learn about you if you choose the time advantage?	• Your partner will receive general information about you based on the health survey but will NOT know that you received extra time. (4) • Your assigned partner will NOT receive any information about you. (5) • Your partner will learn that you received extra time and will be told why. (6)
Study 2 only	
Whose performance determines your final bonus?	• My own performance determines the bonus. (1) • The bonus is based equally on my and my partner's performances. (2) • My partner's performance determines the bonus for both of us. (3)
What information does your partner receive about you?	• My partner will receive general information about me based on the health survey. (4) • My assigned partner will not receive any information about me. (5) • My partner learns about my demographic background. (6)
According to the message you just read, what did your assigned partner reveal?	• Nothing about their cognitive traits. (1) • They disclosed traits associated with neurodivergence. (2) • They disclosed symptoms consistent with an ADHD diagnosis. (3) • I'm not sure / I didn't read the message. (4)

Part 4: Post-treatment stated satisfaction, fairness and second-order belief elicitation

Stated satisfaction approach (Diaz *et al.*, 2023). Participants were asked to evaluate how satisfied they would feel with various income allocations. In each scenario, their own payment was fixed at $20, while the payment to another (unspecified) person varied from $10 to $30, in increments of $2.50. This task was identical across studies. Each scenario was rated on a 7-point Likert scale ranging from 1 = *Extremely dissatisfied* to 7 = *Extremely satisfied*.

Table A3.19: Post-treatment survey questions

Question item	Options and coding
Fairness perception (within-subject randomization)	
How fair do you find the following scenario? Your assigned partner receives an advantage (e.g., extra time) based on their health condition.	• Completely fair (1) • Acceptable (2) • Unfair (3) • Very unfair (4)
How fair do you find the following scenario? Your assigned partner receives an advantage (e.g., extra time) based on luck.	• Completely fair (1) • Acceptable (2) • Unfair (3) • Very unfair (4)
Second-order beliefs (between-subject randomization)	
How do you think most participants in this study would perceive someone when deciding whether to remain matched with them — if they knew the person had one of the following attributes?	• Has a physical disability (e.g., uses a wheelchair, limited mobility) (1) • Has a low-income background (2) • Has a history of mental health challenges (e.g., depression, anxiety) (3) • Is neurodivergent (e.g., ADHD, autism) (4) Options: Very negative (1) | Somewhat negative (2) | Neutral (3) | Somewhat positive (4) | Very positive (5)
How do you think most people in the general population would perceive someone when deciding whether to work with them — if they knew the person had one of the following attributes?	• Has a physical disability (e.g., uses a wheelchair, limited mobility) (1) • Has a low-income background (2) • Has a history of mental health challenges (e.g., depression, anxiety) (3) • Is neurodivergent (e.g., ADHD, autism) (4) Options: Very negative (1) | Somewhat negative (2) | Neutral (3) | Somewhat positive (4) | Very positive (5)

Conclusion

Conclusion

This dissertation examines how social interactions and incentives guide individual behavior in health-related decisions. Across three chapters, I show that what people do depends not just on private preferences or monetary incentives, but on what they see others do, what they believe others expect, and how they expect to be evaluated by others. Social signals, such as norms, comparisons, and reputational cues, are central to modeling choice behavior. Whether a person adopts a new behavior, asks for support, or discloses sensitive information often hinges on how that action will be interpreted by others.

Health decisions are especially responsive to these cues. Many are visible, emotionally charged, and made under uncertainty and incomplete information. Chapter 1 documents how different motives, ranging from norms to social learning, channel peer effects in weight-related behavior among young people, with evidence pointing to social norms as the main driver. Chapter 2 shifts the focus to positional concerns. While positional concerns also strongly rely on social comparison, they originate less from a desire to belong than from being motivated by status. Using hypothetical trade-off decisions, I show that concerns about status, and specifically envy, drive preferences over health outcomes, even when absolute gains are available. Chapter 3 adds a reputational layer, testing how people decide to disclose a stigmatized trait when doing so may involve social (and economic) costs. However, while economic risk and message framing had no measurable effect, disclosure was predicted by fairness beliefs and second-order expectations about how others would respond.

This last result is telling: Participants from a minoritized group (i.e., people who self-identified as having ADHD) disclosed at higher rates, possibly not because they faced lower risks, but because they expected others to be more accepting. These beliefs were not shared by their non-neurodivergent partners. That gap in second-order beliefs (what people believe others believe) could explain why disclosure remained high even under the threat of rejection and forfeiture of a bonus. The mechanism driving disclosure behavior is not just a cost-benefit calculation; it is about how people interpret what they believe they signal to others (e.g., "If I disclose this, will others think I am able to deliver or fear I might underperform?") in addition to what they expect others to do in response (e.g., accept or reject). People act on what they believe the signal means, and what they expect it will trigger.

This dissertation demonstrates that behavior in health contexts is structured by social information, and that this holds across a range of methods, including meta-regression, survey experiments, and incentivized experimental tasks. Such social information may take the shape of norms and

comparisons, but also of beliefs about legitimacy and fairness. Across three chapters, I identify and unpack distinct mechanisms (conformity, status, stigma), and show how they operate across different contexts. The objective has been to not just document peer effects, but to clarify how they work, and to offer straightforward implications for policy design: Many health interventions focus on incentives or on increasing access to health services and care, as if people's behaviors operated in isolation. Particularly with regard to behavioral change in health, we have observed that once pecuniary rewards come to an end, behaviors often revert to how they used to be (Michaelsen and Esch, 2023; Vlaev *et al.*, 2019). However, if decisions are influenced by what others do or expect, policy must also target social beliefs. Interventions that make norms visible, reduce stigma, or signal fairness or legitimacy can change behavior, even when financial incentives do not. For example, although policymakers often use financial incentives to promote health behaviors, actual evidence on their effectiveness and sustained behavioral change is mixed and inconclusive (see, for example, Mantzari *et al.*, 2015; De Walque, 2020).

That said, there is still much that remains to be understood. For example, future work should focus on the study of how people form second-order beliefs, how stable such beliefs are, and whether they can be shifted at scale. When we think about "what others believe", these "others" are rarely a single abstract entity, and we usually think about a particular reference group (such as friends, colleagues, or neighbors). Naturally, people pay most attention to the beliefs of those closest to them, because they matter for belonging and social acceptance. This implies that identity and group membership determine whose opinions and actions are monitored, and that, in order to understand how second-order beliefs are formed, there is also a need to understand the social identities they relate to. From a methodological point of view, combining experimental methods with digital behavioral data, or even neuroimaging methods (where feasible), could significantly deepen our understanding of how social information and incentives operate and influence decision-making in real time. With regard to the latter, for example, functional near-infrared spectroscopy (fNIRS) has already proven insightful in social interaction research, demonstrating how certain brain regions associated with social cognition respond to observability (Krishnan-Barman *et al.*, 2023). The premise of mapping the neural correlates of how we perceive, interpret, and respond to social information makes fNIRS an auspicious addition to the behavioral economics and social research toolbox. It is particularly well suited to settings where individuals must navigate social norms, interpret signals, or make decisions shaped not only by economic considerations but also by the cognitive and social processing of information. Thus, utilizing fNIRS or similar neuroimaging tools

in behavioral economics experiments to study identity signaling, exposure to peers, and prevalent norms could provide meaningful insights into how social influences are internalized.

Social influences play a central role in human decision-making. They guide people's understanding of available options and the risks or benefits they associate with those. Whether someone thinks a health behavior is "normal", risky, or acceptable often depends on what people observe their peers do or think. Additionally, health decisions (like smoking, disclosing a diagnosis, or getting vaccinated) are often not purely private and are shaped by social norms and potential judgment from others. This dissertation takes a step toward understanding how norms, peers, and reputations guide behavior, and how policy might best respond. By integrating these findings into public health strategies, we may help design interventions that advance both effectiveness and fairness, however small this contribution to the creation of a more equitable society may be. Ultimately, it is one that, in the Leibnizian sense, moves us closer to the best of all possible worlds.

References

De Walque, Damien. 2020. "The use of financial incentives to prevent unhealthy behaviors: a review." *Social Science & Medicine*, 261: 113236.

Krishnan-Barman, Sujatha, Uzair Hakim, Marchella Smith, Ilias Tachtsidis, Paola Pinti, and Antonia F de C Hamilton. 2023. "Brain mechanisms of social signalling in live social interactions with autistic and neurotypical adults." *Scientific Reports*, 13(1): 18850.

Mantzari, Eleni, Florian Vogt, Ian Shemilt, Yinghui Wei, Julian PT Higgins, and Theresa M Marteau. 2015. "Personal financial incentives for changing habitual health-related behaviors: a systematic review and meta-analysis." *Preventive medicine*, 75: 75–85.

Michaelsen, Maren M, and Tobias Esch. 2023. "Understanding health behavior change by motivation and reward mechanisms: a review of the literature." *Frontiers in Behavioral Neuroscience*, 17: 1151918.

Vlaev, Ivo, Dominic King, Ara Darzi, and Paul Dolan. 2019. "Changing health behaviors using financial incentives: a review from behavioral economics." *BMC public health*, 19(1): 1059.

List of Figures

List of Tables