# Decision-making in Social Systems:

# Injustice, Inequality, and Ignorance

Jolyon Joseph Miles-Wilson

Thesis submitted for the degree of Doctor of Philosophy

School of Psychology

University of Sussex

August, 2021

## Declaration

This thesis is presented in an article format wherein each chapter follows a layout appropriate for publication in peer-reviewed journals (though none of the papers have yet been submitted for publication). The first section of the thesis provides a general introduction to the research topic and the final section considers the findings of the presented experiments in the context of this broader literature.

The experiments were designed with considerable guidance from my supervisor, Dr Daniel Campbell-Meiklejohn. I was solely responsible for creating all experimental materials (including experimental tasks), recruiting participants, administering experiments, analysing data, and authoring the manuscripts presented herein. Dr Campbell-Meiklejohn provided feedback at analysis and write-up stages.

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Jolyon Joseph Miles-Wilson

19th August, 2021

## Acknowledgments

Many thanks go to my supervisor Dr Daniel Campbell-Meiklejohn for the opportunity to embark on this formative journey and the advice and guidance he has provided along the way. Thank you to everyone in the School of Psychology that I have had the pleasure to know. I am especially grateful for the companionship of the friends I have made: Bence, Francesco, Helena, James, Jamie, Jo, Jukka, Lina, Pedro, Petra, Selin, and Sumeyra. A big thank you to Clare Holmes, who inspired me to pursue Psychology all the way back in college when I was 18 years old. Thank you to all the participants and project contributors that helped me with this research.

A huge thank you to my family – Mum, Dad, Endora – who have patiently endured my infantile regression back in the family home whilst I completed this thesis, and who have supported me in so many ways over the years. Without them I surely would not have arrived at this point. Thank you to Ludovica, who held my hand.

**Note to the reader**

For ease of navigation, the reader is encouraged to make use of the ‘Bookmarks’ navigation pane (usually in an expanding toolbar on the left of the PDF viewer interface), which provides the document outline and contains hyperlinks to sections and subsections. In-text references to Tables and Figures are also hyperlinked to their corresponding Table or Figure for quick cross-referencing.

## Thesis Summary

University of Sussex

Jolyon Joseph Miles-Wilson

PhD Psychology

## Decision-making in Social Systems: Injustice, Inequality, and Ignorance

I present three studies exploring decision-making in artificial social systems that provide insight on the situational factors and individual differences relevant to political engagement. Studies 1 and 2 aimed to provide insight, using a value-expectancy approach, into the dynamics of individual and group opposition to systemically unfair social contexts. Participants were placed in live interactive groups that highlighted a "class difference" between "Elite" and non-Elite participants. Non-Elite participants made incentive-based decisions to support or challenge the system. In Study 1, I found that individuals were more likely to challenge the system when acting in a group compared to when acting alone. This result was accompanied by greater feelings of efficacy when in a group, but only when participants strongly identified with one another. In Study 2, I showed that efforts to challenge the system were undermined by the opportunity to freeride only when the class difference was not emphasised; when the difference between Elites and Non-Elites was salient, participants were no less likely to challenge when freeriding was possible compared to when it was not. In both studies, the availability of coordinating information reduced the relevance of social identity in moderating feelings of efficacy. These findings add to the literature on collective action by experimentally demonstrating when mobilisation is more likely. Study 3 aimed to explore the extent to which individuals are motivated to discover whether their immediate social context is fair or unfair, and what individual difference traits predict these decisions. Participants completed a task in pairs and were subsequently made aware of the possibility they had been paid differently to their partner. Participants could discover whether payments were unfair or remain in ignorance. The results showed that several individual differences were important for this decision. The most compelling was System Justification, which had a complex influence on participants' comfort not knowing and whether they chose to seek information. Taken together, the thesis contributes empirical data to an understanding of the attitudes and behaviour of constituents of (potentially) unjust social systems.

**Contents**

## List of Tables

## List of Figures

# General introduction

As members of a highly social species, much of what we do involves a social dimension. This is no less true of some of the most considerable challenges we face in the 21$^{st}$ Century. Borne from a dispositional sense of morality (Tomasello, 2016), human sociality – that is, the tendency for individuals to form interdependent group relationships (as evidenced, for example, by the formation of societies) – permits sophisticated and coordinated group action that provides better outcomes than could be achieved by any individual alone. Unfortunately, though it is true that humans are well-equipped to deal with social coordination challenges, the limits of our sociality are at the root of many of the issues we currently face. Being characteristically social and cooperative as a species does not equate to universally positive society. By the same token that effective cooperation can challenge and overcome social problems, failures to cooperate can create and sustain them. Furthermore, group affiliation and social hierarchy mean that successful cooperation within one group may spell trouble for other groups (Pratto et al., 2006; Sidanius & Pratto, 1999).

Social inequality is perhaps the most pernicious modern challenge that can be both caused and solved by social minds. The tendency for humans to organise into group-based social hierarchies creates a context in which members of dominant groups have better opportunities and prospects than non-dominant groups (Pratto & Stewart, 2011; Sidanius & Pratto, 1999). Such hierarchies are sustained because dominant groups are motivated to maintain their privileged position and non-dominant groups may lack the resources to change it (Klandermans, 1984). This can lead to the perpetuation of inequality whereby a lack of resources and a deflated sense of efficacy are persistent barriers to change and, in light of this, members of disadvantaged groups may rather be motivated to legitimise the inequality in order to maintain a sense of wellbeing (Jost et al., 2003, 2004). Thus, cooperation in advantaged groups and a lack of cooperation in disadvantaged groups – both strongly social in nature – can contribute to the persistence of social inequality.

Social inequality is particularly problematic because it sets the stage for multiple forms of disadvantage that affect every aspect life. In addition to the proximate malignant effects of discrimination, systemic inequality contributes to a host of complicating issues which disproportionately affect disadvantaged groups. For example, minority communities are disproportionately impacted by the ill effects associated with the environment. African Americans are exposed to 38% more nitrogen dioxide (an air pollutant) than Caucasian Americans, and this is estimated to be responsible for around 7,000 additional deaths per year (Clark et al., 2014). Furthermore, African Americans are more likely than Caucasian Americans to live in close proximity to an industrial facility (Perlin et al., 2001). The fallout from extreme weather events has also been found to disproportionately affect minority communities. Systematic differences in the amount of emergency aid provided to White and non-White communities following natural disasters are shown to increase wealth inequality; whereas White individuals living in areas severely affected by natural disasters gained wealth over a period of fourteen years, Black, Latino, and Asian individuals living in severely affected areas lost wealth (Howell & Elliott, 2019). Changes to oceans as a result of the changing climate may reduce the prevalence of certain sea-dwelling species, which in turn disproportionately affects members of minority communities, such as the Quinault Indian Nation, whose culture and livelihood are closely linked with these species (Crosman et al., 2019; Lynn et al., 2014; Weatherdon et al., 2016).

This "environmental injustice" is mirrored by other health inequalities made even more evident by the present pandemic. Patel et al. (2020) argue that economically-disadvantaged individuals may be at higher risk from COVID-19 due to overcrowded housing (Cardoso et al., 2004), fewer opportunities to work from home, and higher incidence of hypertension and diabetes (*Health Survey for England 2005: The Health of Older People*, 2007; Marmot et al., 2010), which have been linked to an increased likelihood of being admitted to intensive care or dying as a result of COVID-19 (Guan et al., 2020). The effects of economic inequality on health are also currently playing out at the level of nations; with several of the richest countries having passed the 50% vaccinated mark (and hoarding more vaccines than required; Dyer, 2020), at the time of writing many poorer nations have vaccinated less than one per cent of their population (*Coronavirus Vaccinations - Statistics and Research*, 2021). Health inequality is not unique to the pandemic, however. Differences in life expectancy and health are consistently documented between the most and least deprived areas in England (*Health Profile for England*, 2017, 2018, 2019). In nations with limited or no public healthcare service (e.g., the USA), the problem is further compounded by limited access to healthcare amongst residents of poorer counties (Shaw et al., 2016).

It is clear, then, that social inequality systematically disadvantages individuals in multiple and highly consequential ways. It is therefore highly desirable (especially for members of disadvantaged groups) to reduce and ultimately remove inequality from society. This is of course easier said than done, since inequality emerges in the highly complex context of the social world. However, the power for change can be achieved through a better understanding of how social inequality persists and how individuals can work collectively against it. This has been an increasingly active area of research in social psychology. Collective action is defined as any action by representatives of a group that is aimed at improving that group's status (Wright et al., 1990). Researchers of collective action have explored the complex ways in which groups of individuals are more and less likely to challenge social inequality. A number of models of collective action have been developed which aim to identify the situational and personal conditions that make cooperation to effect social change – particularly amongst members of disadvantaged groups – more likely. In recent years this has led to the convergence of distinct perspectives into integrated accounts. A particularly influential one – especially in terms of its stimulation of further research – is the Social Identity Model of Collective Action (SIMCA; Van Zomeren et al., 2008), which integrates perspectives from Social Identity Theory (Tajfel, 1978; Tajfel & Turner, 2004), Relative Deprivation Theory (Runciman, 1966), and Resource Mobilisation Theory (Klandermans, 1984; McCarthy & Zald, 1977). In essence, this model places social identity as a direct causal predictor of willingness to engage in collective action and an indirect predictor via its influence on feelings of injustice and efficacy. This integration of perspectives, derived from a meta-analysis of the broad literature on collective action, captures elements of the complex social dynamics at work in contexts of systemic inequality, and subsequent research has further refined these ideas (e.g., Thomas et al., 2009; van Zomeren et al., 2010; van Zomeren, Leach, et al., 2012; van Zomeren, Postmes, et al., 2012). Thomas et al. (2009) refined the model by suggesting two-way relationships such that subjective experiences of injustice and group efficacy can influence an individual's social identity. Van Zomeren et al. (2010) presented experimental evidence supporting the notion that efficacy could causally precede feelings of identity in collective action, consistent with both Thomas et al.'s (2009) suggestions and the earlier Dual Pathway model (Sturmer & Simon, 2004).

Studies 1 and 2 of this thesis aimed to contribute to this literature by creating interactive incentivised social systems in the lab. These systems were characterised by a stark social inequality insofar as most participants were disadvantaged and forced to sacrifice some of their wealth or risk losing everything, whilst there was a small Elite class of participants who benefited disproportionately from the unfairness of the system. This approach allowed me to explore two key questions. In Study 1, I explored what influence interdependency amongst the disadvantaged had on their experiences and decisions. This was achieved by making favourable financial outcomes either contingent on chance or contingent on the decisions on other group members. In a follow-up experiment, I explored whether and how the ability to communicate intentions to one another influenced this process. The results of Study 1 expand on the theoretical insights of the aforementioned integrative models (van Zomeren et al., 2008; Thomas et al., 2009) by demonstrating a role of feelings of shared identity as a coordinating tool. The results showed that decisions to challenge the system were more likely when decisions were made collectively, and revealed complex interactions between feelings of identity, efficacy, and decisions to challenge. Specifically, when communication was not possible, efficacy beliefs appeared to be inferred from feelings of shared identity. In contrast, when communication was possible, identification had no moderating effect on the how efficacy beliefs influenced decisions. This pattern of results suggests that, in the absence of concrete information, social identity served a coordinating role by helping individuals gauge how efficacious a decision to challenge would be. Thus, the findings corroborate previous assertions of the central importance of social identity in moderating feelings of efficacy (van Zomeren et al, 2008; Thomas et al., 2009), but highlight that this influence is context-sensitive and that in some cases strong identification may not be strictly necessary for individuals and groups to feel efficacious in challenging existing social structures.

In Study 2, I aimed to further emulate real-world decision-making by introducing the possibility of "freeriding" whereby members of the disadvantaged group could benefit from positive social outcomes without having to take a personal financial risk. Freeriding is a recurring challenge to collective action efforts because it provides an incentive to not contribute to the collective goal whilst still benefiting from others' efforts (Olson, 1965). As a result, cooperation in scenarios where it is possible to freeride is typically lower than in scenarios where it is not possible to freeride (Thöni & Volk, 2018; Boone et al., 2010). In Study 2, by comparing subjective experiences and decisions under conditions where freeriding was and was not possible, I quantified the potential impact of the threat of freeriding amongst members of a disadvantaged group striving for better social outcomes. As in Study 1, in Study 2 I also compared conditions under which participants could communicate intentions with conditions where they could not. The results of this study suggested that in a context with no social inequality, freeriding did indeed reduce the likelihood of challenging the system, consistent with the literature documenting the deleterious effects of freeriding on group cooperation (Fishbacher et al., 2001, Boone et al. 2010). However, in contexts of social inequality, the ability to freeride did not decrease cooperation, and when it was possible for participants to communicate in these scenarios, the possibility of freeriding even *increased* the likelihood of challenging the system. The findings of Study 2 therefore extend previous findings by providing evidence that in social dilemma-like scenarios, the demobilising effects of freeriding may be mitigated and even reversed if decision-makers perceive a strong sense of social injustice and are able to coordinate their efforts by communicating with one another.

In addition to understanding the conditions under which inequality is unlikely to be tolerated by the disadvantaged, an intriguing question concerns what motivates individuals to seek or avoid information that may reveal inequality. This is important because inequality is likely to persist if individuals are not aware of its presence. Decades of research has provided insight into why those disadvantaged by inequality might accept or even justify it. Research suggests that it may be preferable to sustain uncertainty if there is a suspicion that information-seeking will lead to an inconvenient truth. For example, by avoiding information relating to how wealth is distributed, individuals are more likely to make self-interested decisions, possibly because, in the absence of certain information, they can maintain a belief that their decision was not selfish (Dana et al., 2007) or, alternatively, because sustained uncertainty reduces clarity on what is an appropriate decision (Spiekermann & Weiss, 2016).

Whether an individual is motivated to seek equality information is likely to be at least partially explained by elements of their personality. Perhaps most pertinent to the context is the extent to which individuals tend to justify existing social structures. System Justification Theory argues that individuals can be motivated to justify the way things are because doing so has palliative effects and helps to preserve wellbeing by addressing existential, epistemic, and relational needs (Jost, 2019; Jost & Banaji, 1994; Jost & Hunyady, 2003, 2005; Kay et al., 2009). That is, individuals' needs relating to security, certainty, and interpersonal relationships may sometimes be satisfied by accepting, rather than rejecting, the social status quo. Whilst this research has focused on attitudes toward existing or known inequality, an interesting question concerns how individuals approach unestablished or unknown inequality.

A handful of other personality characteristics may be important for whether individuals seek or avoid equality information. The Just World hypothesis (from which System Justification Theory takes influence) argues that individuals are motivated to believe that people tend to get what they deserve so that they can maintain a Belief in a Just World (BJW) and that the world is stable and predictable (Lerner, 1980). Research suggests that high BJW is negatively related to justice-seeking behaviours (Beierlein et al., 2011), suggesting that such individuals may be expected to be less likely to seek equality information. Research on Social Value Orientation (SVO) – which describes the extent to which an individual is concerned with others' outcomes relative to their own (Liebrand, 1984; Messick & McClintock, 1968; Murphy & Ackermann, 2014; Van Lange et al., 1997) – suggests that deviation from pure SVO-measured individualism is associated with greater consideration of others' outcomes (Fiedler et al., 2013) and that pro-social individuals tend to engage in more information processing (Haruno & Frith, 2010). Dispositional optimism – measured using Life Orientation scale (Carver et al., 2010) – is associated with increased engagement with an issue (Nes et al., 2005), suggesting optimists may be likely to seek equality information. Finally, the extent to which individuals are able to tolerate uncertainty (Intolerance of Uncertainty; Koerner & Dugas, 2008) is, of course, likely to be important for whether they seek to resolve uncertainty regarding equality; those less tolerant of uncertainty tend to seek information to resolve it (Rosen et al., 2007).

In Study 3, I investigated how individuals approach equality information from the perspectives of both those potentially disadvantaged and those potentially advantaged by the inequality, and how the above-mentioned personality traits relate to these decisions. I created an artificial system where participants completed work and were either paid the same or differently for their work. Participants were made aware of this possible inequality and whether it would advantage or disadvantage them, and could choose whether to find out for certain. I explored how individual difference measurements (such as the tendency to justify the status quo and pro-sociality) influenced the perceptions and information-seeking behaviour of these potentially advantaged and disadvantaged participants. Doing so provided insight not only into the possible dispositional traits relevant for equality information-seeking, but importantly highlighted how social status (i.e., whether a participant was potentially advantaged or disadvantaged) moderated the influence of these dispositions on perceptions and information-seeking. Approximately half of participants chose to find out whether the social system was fair or unfair. Several of the above-mentioned individual difference measures were important for predicting whether a person would find out, and the predictive power of these characteristics depended on whether the participant was potentially advantaged or disadvantaged by the inequality of the system. Potentially disadvantaged individuals were more likely to find out if they were strong system justifiers, and less likely to find out if they had a pro-social SVO. In contrast, potentially advantaged individuals were less likely to find out if they were strong system justifiers, and more likely to find out if they had a pro-social SVO. The results corroborate previous findings (e.g., Dana et al., 2007, Thunström et al., 2016) by implicating a strategic motivation to avoid information. In addition, the findings add to the existing System

Justification literature by highlighting how relative social status moderates the way in which justifying ideologies manifest in behaviour.

I present each study with a corresponding consideration of its relevance and contribution to the existing literature. In the final section of this thesis, I consider the findings more broadly with respect to the issue of social inequality. In combination, the research presented provides experimental insight into a complex social phenomenon, and I hope that its findings can be helpful for guiding future research and action addressing social inequality.

## References

Beierlein, C., Werner, C. S., Preiser, S., & Wermuth, S. (2011). Are just-world beliefs compatible with justifying inequality? collective political efficacy as a moderator. *Social Justice Research*, *24*(3), 278-296. https://doi.org/10.1007/s11211-011-0139-2

Boone, C., Declerck, C., & Kiyonari, T. (2010). Inducing cooperative behavior among proselfs versus prosocials: the moderating role of incentives and trust. *Journal of Conflict Resolution*, *54*(5), 799–824. https://doi.org/10.1177/0022002710372329

Cardoso, M. R. A., Cousens, S. N., de Góes Siqueira, L. F., Alves, F. M., & D'Angelo, L. A. V. (2004). Crowding: Risk factor or protective factor for lower respiratory disease in young children? *BMC Public Health*, *4*, 19. https://doi.org/10.1186/1471-2458-4-19

Carver, C. S., Scheier, M. F., & Segerstrom, S. C. (2010). Optimism. *Clinical Psychology Review*, *30*(7), 879–889. 10.1016/j.cpr.2010.01.006

Clark, L. P., Millet, D. B., & Marshall, J. D. (2014). National patterns in environmental injustice and inequality: outdoor no2 air pollution in the united states. *PLOS ONE*, *9*(4), e94431. https://doi.org/10.1371/journal.pone.0094431

*Coronavirus (COVID-19) Vaccinations—Statistics and Research*. (2021). Our World in Data. https://ourworldindata.org/covid-vaccinations

Craig, R., & Mindell, J. (Eds.). (2007). *Health Survey for England 2005: The health of older people.* https://files.digital.nhs.uk/publicationimport/pub01xxx/pub01184/heal-surv-heal-old-peo-eng-2005-rep-v5.pdf

Crosman, K., Petrou, E., Rudd, M., & Tillotson, M. (2019). Clam hunger and the changing ocean: Characterizing social and ecological risks to the Quinault razor clam fishery using participatory modeling. *Ecology and Society*, *24*(2). https://doi.org/10.5751/ES-10928-240216

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80. https://doi.org/10.1007/s00199-006-0153-z

Dyer, O. (2020). Covid-19: Many poor countries will see almost no vaccine next year, aid groups warn. *BMJ*, *371*, m4809. https://doi.org/10.1136/bmj.m4809

Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social Value Orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes*, *120*(2), 272–284. https://doi.org/10.1016/j.obhdp.2012.07.002

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9

Guan, W., Liang, W., Zhao, Y., Liang, H., Chen, Z., Li, Y., Liu, X., Chen, R., Tang, C., Wang, T., Ou, C., Li, L., Chen, P., Sang, L., Wang, W., Li, J., Li, C., Ou, L., Cheng, B., … He, J. (2020). Comorbidity and its impact on 1590 patients with COVID-19 in China:

A nationwide analysis. *European Respiratory Journal*, *55*(5): 2000547. https://doi.org/10.1183/13993003.00547-2020

Haruno, M., & Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*(2), 160–161. https://doi.org/10.1038/nn.2468

*Health profile for England*. (2017, 2018, 2019). GOV.UK. https://www.gov.uk/government/collections/health-profile-for-england

Howell, J., & Elliott, J. R. (2019). Damages done: the longitudinal impacts of natural hazards on wealth inequality in the United States. *Social Problems*, *66*(3), 448–467. https://doi.org/10.1093/socpro/spy016

Jost, J., & Hunyady, O. (2003). The psychology of system justification and the palliative function of ideology. *European Review of Social Psychology*, *13*(1), 111–153. https://doi.org/10.1080/10463280240000046

Jost, J. T. (2019). A quarter century of system justification theory: Questions, answers, criticisms, and societal applications. *British Journal of Social Psychology*, *58*(2), 263–314. https://doi.org/10.1111/bjso.12297

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1–27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(6), 881–919. https://doi.org/10.1111/j.1467-9221.2004.00402.x

Jost, J. T., & Hunyady, O. (2005). Antecedents and consequences of system-justifying ideologies. *Current Directions in Psychological Science*, *14*(5), 260–265. https://doi.org/10.1111/j.0963-7214.2005.00377.x

Jost, J. T., Pelham, B. W., Sheldon, O., & Sullivan, B. N. (2003). Social inequality and the reduction of ideological dissonance on behalf of the system: Evidence of enhanced system justification among the disadvantaged. *European Journal of Social Psychology*, *33*(1), 13–36. https://doi.org/10.1002/ejsp.127

Kay, A. C., Gaucher, D., Peach, J. M., Laurin, K., Friesen, J., Zanna, M. P., & Spencer, S. J. (2009). Inequality, discrimination, and the power of the status quo: Direct evidence for a motivation to see the way things are as the way they should be. *Journal of Personality and Social Psychology*, *97*(3), 421–434. http://dx.doi.org/10.1037/a0015997

Klandermans, B. (1984). Mobilization and participation: Social-Psychological Expansions of Resource Mobilization Theory. *American Sociological Review*, *49*(5), 583-600. https://doi.org/10.2307/2095417

Koerner, N., & Dugas, M. J. (2008). An investigation of appraisals in individuals vulnerable to excessive worry: The role of intolerance of uncertainty. *Cognitive Therapy and Research*, *32*(5), 619–638. https://doi.org/10.1007/s10608-007-9125-2

Lerner, M. J. (1980). The Belief in a Just World. In M. J. Lerner (Ed.), *The Belief in a Just World: A Fundamental Delusion* (pp. 9–30). Springer US. https://doi.org/10.1007/978-1-4899-0448-5_2

Liebrand, W. B. G. (1984). The effect of social motives, communication and group size on behaviour in an N-person multi-stage mixed-motive game. *European Journal of Social Psychology*, *14*(3), 239–264. https://doi.org/10.1002/ejsp.2420140302

Lynn, K., Daigle, J., Hoffman, J., Lake, F., Michelle, N., Ranco, D., Viles, C., Voggesser, G., & Williams, P. (2014). The impacts of climate change on tribal traditional foods. In J. K. Maldonado, B. Colombi, & R. Pandya (Eds.), *Climate Change and Indigenous Peoples in the United States: Impacts, Experiences and Actions* (pp. 37–48). Springer International Publishing. https://doi.org/10.1007/978-3-319-05266-3_4

Marmot, M., Allen, J., Goldblatt, P., Boyce, T., McNeish, D., Grady, M., & Geddes, I. (2010). *Fair society healthy lives. The Marmot Review*. Institute of Health Equity. https://www.instituteofhealthequity.org/resources-reports/fair-society-healthy-lives-the-marmot-review/fair-society-healthy-lives-full-report-pdf.pdf

McCarthy, J. D., & Zald, M. N. (1977). Resource mobilization and social movements: A partial theory. *American Journal of Sociology*, *82*(6), 1212–1241. https://doi.org/10.1086/226464

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, *4*(1), 1–25. https://doi.org/10.1016/0022-1031(68)90046-2

Murphy, R. O., & Ackermann, K. A. (2014). Social Value Orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, *18*(1), 13–41. https://doi.org/10.1177/1088868313501745

Nes, L. S., Segerstrom, S. C., & Sephton, S. E. (2005). Engagement and arousal: Optimism's effects during a brief stressor. *Personality and Social Psychology Bulletin*, *31*(1), 111–120. https://doi.org/10.1177/0146167204271319

Olson, M. (1965). *The logic of collective action*. Harvard University Press.

Perlin, S. A., Wong, D., & Sexton, K. (2001). Residential proximity to industrial sources of air pollution: Interrelationships among race, poverty, and age. *Journal of the Air & Waste Management Association*, *51*(3), 406–421. https://doi.org/10.1080/10473289.2001.10464271

Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology*, *17*(1), 271–320. https://doi.org/10.1080/10463280601055772

Pratto, F., & Stewart, A. L. (2011). Social Dominance Theory. In D. J. Christie (Ed.), *The Encyclopedia of Peace Psychology*, 1-4. Blackwell Publishing. https://doi.org/10.1002/9780470672532.wbepp253

Rosen, N. O., Knäuper, B., & Sammut, J. (2007). Do individual differences in intolerance of uncertainty affect health monitoring? *Psychology & Health*, *22*(4), 413–430. https://doi.org/10.1080/14768320600941038

Runciman, W. G. (1966). *Relative deprivation and social justice: A study of attitudes to social inequality in twentieth-century England*. Routledge & Kegan Paul.

Shaw, K. M., Theis, K. A., Self-Brown, S., Roblin, D. W., & Barker, L. (2016). Chronic disease disparities by county economic status and metropolitan classification, Behavioral Risk Factor Surveillance System, 2013. *Preventing Chronic Disease*, *13:* 160088. https://doi.org/10.5888/pcd13.160088

Sidanius, J., & Pratto, F. (1999). *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge University Press. https://doi.org/10.1017/CBO9781139175043

Spiekermann, K., & Weiss, A. (2016). Objective and subjective compliance: A norm-based explanation of ‘moral wiggle room’. *Games and Economic Behavior*, *96*, 170–183. https://doi.org/10.1016/j.geb.2015.11.007

Sturmer, S., & Simon, B. (2004). Collective action: Towards a dual-pathway model. *European Review of Social Psychology*, *15*(1), 59–99. https://doi.org/10.1080/10463280340000117

Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. Academic Press.

Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of intergroup behavior. In J. T. Jost & J. Sidanius (Eds.), *Political Psychology* (pp. 276–293). Psychology Press. https://doi.org/10.4324/9780203505984-16

Thomas, E. F., McGarty, C., & Mavor, K. I. (2009). Aligning identities, emotions, and beliefs to create commitment to sustainable social and political action. *Personality and Social Psychology Review*, *13*(3), 194–218. https://doi.org/10.1177/1088868309341563

Thöni, C., & Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters, 171*, 37–40. https://doi.org/10.1016/j.econlet.2018.06.022

Thunström, L., Nordström, J., Shogren, J. F., Ehmke, M., & van’t Veld, K. (2016). Strategic self-ignorance. *Journal of Risk and Uncertainty*, *52*(2), 117–136. https://doi.org/10.1007/s11166-016-9236-9

Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.

Van Lange, P. A., Otten, W., De Bruin, E. M., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, *73*(4), 733–746. https://doi.org/10.1037//0022-3514.73.4.733

van Zomeren, M., Leach, C. W., & Spears, R. (2010). Does group efficacy increase group identification? Resolving their paradoxical relationship. *Journal of Experimental Social Psychology*, *46*(6), 1055–1060. https://doi.org/10.1016/j.jesp.2010.05.006

van Zomeren, M., Leach, C. W., & Spears, R. (2012). Protesters as "passionate economists": a dynamic dual pathway model of approach coping with collective disadvantage. *Personality and Social Psychology Review*, *16*(2), 180–199. https://doi.org/10.1177/1088868311430835

van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin*, *134*(4), 504–535. https://doi.org/10.1037/0033-2909.134.4.504

van Zomeren, M., Postmes, T., & Spears, R. (2012). On conviction's collective consequences: Integrating moral conviction with the social identity model of collective action. *British Journal of Social Psychology*, *51*(1), 52–71. https://doi.org/10.1111/j.2044-8309.2010.02000.x

Weatherdon, L. V., Magnan, A. K., Rogers, A. D., Sumaila, U. R., & Cheung, W. W. L. (2016). Observed and projected impacts of climate change on marine fisheries, aquaculture, coastal tourism, and human health: An update. *Frontiers in Marine Science*, *3*. https://doi.org/10.3389/fmars.2016.00048

Wright, S. C., Taylor, D. M., & Moghaddam, F. M. (1990). Responding to membership in a disadvantaged group: From acceptance to collective protest. *Journal of Personality and Social Psychology*, *58*(6), 994–1003. https://doi.org/10.1037/0022-3514.58.6.994

# Study 1. The Strength of Groups: How group decision-making and social hierarchy influence decisions to challenge unfair social systems

## Abstract

An extremely productive multi-disciplinary literature has emerged that aims to explain why and how individuals engage in coordinated efforts that come at a cost to the individual but benefit the collective. This literature encompasses many theoretical viewpoints and has encouraged an integrative approach, giving rise to multiple models that aim to explain the phenomenon of collective action. The present study aimed to contribute to this literature and help inform theory by providing empirical data from a controlled incentivised decision environment designed to emulate real-world economic disadvantage. This was achieved by creating unfair social systems involving real participants that were either advantaged or disadvantaged by the prescribed social system and measuring the subjective experiences and system-challenging behaviour of the disadvantaged. Across two experiments I implemented a 2 x 3 within-subjects factorial design to elicit experiences of identification, injustice, and efficacy. Ninety-one (~70% female, ~91% undergraduate, of which ~65% were psychology students) participants were rendered "disadvantaged" in the experiment and made decisions either individually or as part of a group (group versus individual manipulation, GVI) in systems where the inequality either benefited a "Legitimate" or "Illegitimate" advantaged participant, or where there was no individual that benefited from the inequality ("Elite" manipulation). Through interactively-linked computers, Disadvantaged participants were required to decide to either support the system and accept lower payoffs or to take a personal risk to challenge the system and receive fairer payoffs. In Experiment 1, participants had no information about the likelihood of success, whilst in Experiment 2 a different sample of eighty-seven participants (~62% female, ~93% undergraduate, of which 90% were psychology students) were able to communicate intentions to one another before deciding in the group scenarios, and in the individual scenarios received analogous information about the likelihood of success in the form of coin tosses. Generalised linear mixed model analysis showed positive effects of GVI on decisions to challenge the system and context-specific effects of Elite. Linear mixed model analysis indicated that efficacy beliefs (i.e., expectations relating to the probability of success) and social identification were sensitive to changes in context and were important in guiding decisions. Feelings of efficacy increased the probability of challenging the system in groups when identification (i.e., feelings of association between group members) was high, but only in Experiment 1 where the extra coordinating information was not available. The results add to the existing literature providing an empirical demonstration of how social identification may serve as a coordinating tool for gauging group efficacy in the absence of explicit efficacy information. By extension, the findings suggest that under circumstances where explicit information is available, identification may be less important in shaping feelings of efficacy than has previously been suggested.

## Introduction

Collective action is defined as action taken by representatives of a group that is aimed at improving that group's status (Wright et al., 1990). The topic has been avidly researched for over 50 years and in more recent decades has been the interest of theorists from multiple perspectives. Indeed, this widespread interest has given rise to an "age of integration" (van Zomeren, 2013), and there exist more than a few models that seek to integrate disparate theories into unified frameworks. In a meta-analysis of 182 studies on the topic, Van Zomeren et al. (2008) identified three of the most prominent social-psychological factors that predict collective action for social change: social identity, perceived injustice, and perceived efficacy. Their Social Identity Model of Collective Action (SIMCA) integrated the approaches of Social Identity Theory (Tajfel, 1978; Tajfel & Turner, 2004), Relative Deprivation Theory (Runciman, 1966), and Resource Mobilisation Theory (Klandermans, 1984; McCarthy & Zald,

1977) in an effort to comprehensively explain the motivating factors predicting collective action. According to SIMCA, social identity is central to predicting collective action. The degree to which individuals identify with their social group is argued to be a predictor both directly and indirectly via its influence on feelings of injustice and group efficacy; that is, feelings of shared identity can influence perceptions of injustice and how effective individuals feel their group will be in engaging in collective action to address the situation. This model has been highly influential in stimulating new research questions and more refined accounts of the complexity of decision-making in social dilemma-like scenarios.

The Encapsulation Model of Social Identity in Collective Action (EMSICA; Thomas et al., 2009) provided a similar account to SIMCA but was distinct in positing that the relations between social identity, efficacy, and injustice are bi-directional. According to this model, perceptions of group efficacy and injustice can feed back into and even causally precede the formation of social identity. A slightly earlier model was offered by Sturmer and Simon (2004), who conceived two pathways involved in collective action: one involving the calculation of costs and benefits and one involving identification with a social group. The latter pathway itself is suggested to comprise distinct stages of appraisal (Lazarus, 1991) such as shared feelings of deprivation and attribution of blame for this deprivation (Simon & Klandermans, 2001).

The various models, though independently conceived, together offer considerable insight into the complexities of social decision-making in collective action. The different lines of research tend to converge on the importance of social identity, perceptions of injustice, and feelings of efficacy. I briefly elaborate on each of these concepts below.

**Social identity**

The premises of the highly influential Social Identity Theory (SIT; Tajfel, 1978; Tajfel & Turner, 1979) are adopted in most research invoking a role of identity in collective action.

This is because SIT provides a framework for explaining why individuals would be motivated to mobilise (i.e., engage in action). It posits that individuals are motivated to maintain positive social identities and that this can be achieved in several ways. When individuals' social identity comes under threat (for example, by experiencing prejudice), there are at least three factors that influence how they can act to maintain a positive social identity. One possibility is simply to change group membership; if the boundaries of group membership are permeable, individuals can regain esteem by moving out of the threatened group. If, however, group boundaries are impermeable (as is often the case), it is suggested that individuals assess the legitimacy and stability of the intergroup status differential (i.e., the pre-potence of one group over another) and act accordingly to preserve identity-related self-esteem. If the state of affairs is perceived to be legitimate and stable, it is possible that the disadvantaged can preserve a positive social identity by justifying the status differential (this phenomenon, known as System Justification, is considered closely in Study 3). If the situation is perceived as illegitimate and unstable, SIT argues that disadvantaged individuals will be motivated to contest the status differential to re-establish a positive group identity.

Research consistently finds support for the notion that social identity is important for collective action efforts. Experimentally, findings indicate that identification with a group increases group-members' intentions to engage in collective action. For example, Masson et al. (2016) found that experimentally increased identification with an ingroup increased intentions to engage in environmentally sustainable behaviour. Ingroup identification has also been shown to increase cooperation in social and resource dilemma games and to increase cooperation amongst group-members (Brewer & Kramer, 1986; De Cremer & Van Vugt, 2002). Qualitative work likewise provides compelling and detailed accounts of social identity processes in collective action, including how participation leads to empowerment, psychological change, and the formation of social identities (Drury & Reicher, 2000, 2005, 2009; Vestergren et al., 2018).

Whilst ingroup identification is demonstrably important for encouraging collective cooperation, some theorists suggest that a more specific form of identity is particularly potent in galvanising individuals in collective efforts. Simon and Klandermans (2001) conceptualised a *politicised* collective identity as an identity that *motivates* group members to act on an issue with the intention of eliciting change. Collective identity politicises when (i) there is a shared feeling of group-based deprivation which is (ii) attributable to an adversary/adversarial group, and (iii) which is acted upon so that members of broader society are forced to take a stance on the issue. For example, an individual who identifies as a member of the Lesbian, Gay, Bisexual, Transgender, Queer/Questioning+ (LGBTQ+) community possesses a collective social identity that per se is not politicised, but which can become politicised in situations where group members are motivated by this identity to engage in actions for change, such as LGBTQ+ rights demonstrations. This conceptualisation thus emphasises the notion that collective identities can become more or less relevant depending on contextual variables (Turner et al., 1987). In a series of studies investigating different social movement contexts (e.g., US/German gay movement, US Fat Acceptance Movement), Sturmer and Simon (2004) showed that identification with the social movement (i.e., politicised collective identity) was a more important predictor of intentions to engage and of actual engagement than identification with the relevant disadvantaged group. Van Zomeren et al.'s (2008) meta-analytic data similarly demonstrated that politicised identities were more strongly associated with collective action than non-politicised identities, and in a similar vein Akfirat et al. (2020) more recently showed in the context of digitally-mediated collective action that identification with emergent groups was more predictive of collective action than identification with pre-existing groups. This evidence lends weight to the notion that politicised social identities are particularly important for social mobilisation. It should be noted here that *social* and *collective* identity are suggested to overlap and yet to be distinct constructs at different levels of analysis (Klandermans, 2014; Snow & Corrigall-Brown, 2015). Klandermans (2014) argues that *social identity* refers to the social cognition of an individual relating to their membership of one or more groups, whilst *collective identity* relates to "cognitions shared by members of a single group about the group of which they are a member" (p. 3). In the present study, I made the assumption that social identity was incorporated into collective identity because only one identity was salient and relevant for decisions in the experimental context (i.e., the collective identity; see Method for more details).

**Injustice**

Implicit in Social Identity Theory is the notion that one's group's status is evaluated relative to other groups. This relative evaluation can lead one to infer whether one's social group is better or worse off compared to another social group. Insofar as individuals are motivated to maintain a positive self-esteem by reducing negative evaluations of one's self and one's group, if an individual perceives their group to be disadvantaged relative to another group they may be motivated address this imbalance by taking action. This tenet is made explicit in Relative Deprivation Theory (Runciman, 1966). According to this theory, individuals make interpersonal or intergroup comparisons to subjectively appraise whether they are personally or collectively disadvantaged, and whether this disadvantage is fair. Important to note is that only *perceptions* of disadvantage and fairness are necessary. Indeed, research suggests that in some cases objective inequality (e.g., unequal sharing of money) can be perceived as fair or justified (see Jost, 2019). Once systematic injustice has been perceived, individuals can be motivated to address the negative appraisal of their group by acting.

Runciman (1966) suggested a distinction between perceptions of individual-based relative deprivation and group-based relative deprivation. Individual-based deprivation is related to self-relevant appraisals (e.g., self-esteem), whilst group-based deprivation is associated more with group-relevant appraisals (e.g., ethnic identity; Osborne, Sibley, & Sengupta, 2015), and these differing types of appraisals are found to have implications for social behaviour. Group-based relative deprivation in particular is found to encourage collective actions such as protests, whilst individual deprivation does not (Olson et al., 1995; Osborne & Sibley, 2013). Similarly, group-, but not individual-, based relative deprivation is associated with more strongly prejudiced views toward immigrants (Pettigrew et al., 2008).

**Efficacy**

Collective efficacy is the extent to which a group member believes their group is able to achieve its goals (Bandura, 2000; Mummendey et al., 1999). A stronger belief in the potential success of group action is an important predictor of whether individuals will engage in that group action (Cohen-Chen & Van Zomeren, 2018; Hornsey et al., 2006; Mummendey et al., 1999). The concept of efficacy in the context of social mobilisation has been heavily influenced by Klandermans (1984), who emphasised that the “expected value” of outcomes is a critically important consideration when contemplating social action. In simple terms, according to this argument, expected value is derived by weighting the value of an outcome (e.g., successful social change) against beliefs in being able to achieve the outcome. The result of this cognitive process then guides the individual’s decision.

The aim of the present study was to add to this literature using empirical data from a controlled incentivised decision environment designed to emulate real-world economic disadvantage. Using design principles from behavioural economics, I created a series of incentivised social systems in which some participants were disadvantaged and stood to leave the experiment with considerably lower payments, but could decide to challenge this system for the possibility of fairer payments. Behavioural economics provides a useful framework for exploring the motivations involved in social decision-making because it enables the researcher to infer the expected utility of decision options (i.e., the value of options once subjective preferences have been taken into account) by specifying an objective incentive structure and observing whether and how decisions deviate from this structure. This allowed me to explore whether the social decisions in the present experiment were based solely on the presented value of the outcomes or deviated from the value-based predictions by incorporating situation-specific preferences. Aspects of the systems were manipulated to create a 2 x 3 factor design; participants made decisions either individually or as part of a group in systems where the inequality either benefited a "Legitimate" or "Illegitimate" advantaged participant, or where there was no individual that benefited from the inequality. I measured the subjective experiences of identification, fairness, and efficacy to explore how these variables – which have been identified in the literature as centrally important for collective action behaviour – were influenced by the context and how they in turn influenced decisions. Thus, by drawing on both behavioural economic and social psychological methods, I was able to explore the subjective elements of social mobilisation in a controlled context in which the value of decisions could also be quantified.

# Experiment 1

## Method

### Overview of the experiment

Participants took part in the experiment at the Behavioural Laboratory at the University of Sussex Behavioural Economics Department. This laboratory contains 24 computers that are linked by server to create an interactive decision environment via the use of the Zurich Toolbox for Ready-made Economic Experiments (zTree version 3.6.7; Fischbacher, 2007). Participants completed several tasks in this environment (detailed below). The experimental tasks (also termed ‘games’) were designed to model unfair social systems in which participants were part of a disadvantaged “citizenry” who suffered from the systemic unfairness. In each experimental session, two participants were assigned to an advantaged role for all games; these participants did not make any decisions but presided over the system in certain conditions such that they benefited from the systemic unfairness. Participants were told that these “Elite” positions could be attained through excellent performance in an addition task (detailed below). After the addition task, however, participants had an opportunity to “steal” a position from one of the top-scorers in the addition task. One participant was randomly selected from all the participants who opted to “steal” and the Elite position was assigned to them. There were therefore two Elites: one who was the top-scorer on the addition task and one who successfully “stole” the position from the second-best scorer (termed in this paper “Legitimate” and “Illegitimate” Elite, respectively). Note these terms were not used in the participant-facing materials. Instead, I chose arbitrary names for the two types of participant to avoid exogenous social labels. The Elite individuals were also referred to as “Triangles” and the non-Elites were referred to as “Circles”. To distinguish between the two types of Triangle, I used “the Triangle who scored the most in the addition task” for the Legitimate Elite, and “the Triangle who stole the position from one of the top two scorers" for the "Illegitimate Elite". The participant whose Elite position was stolen participated as a member of the citizenry, however their data was excluded from analysis as it was considered likely that their decisions would be biased by their experience of being displaced by the Illegitimate Elite.

Participants took part in six games in a 2 (group versus individual [GVI]) x 3 (No Elite vs Legitimate Elite vs Illegitimate Elite) within-subjects design; according to the GVI manipulation, participants' decisions either affected only their own outcome (individual scenario) or contributed to the entire group's chances of better outcomes. For each GVI condition there were three different Elite conditions that determined the beneficiary of the inequality; the Legitimate Elite could benefit, the Illegitimate Elite could benefit, or no Elite could benefit. In each game, participants were endowed with 100 coins (equivalent to £8.00). Participants had to make a choice between paying 50 coins (i.e., half their endowment) to the System and keeping their remaining 50 coins for certain or to challenging the System by not paying for a chance of keeping all 100 coins. In the conditions involving an Elite, the payments of participants would go directly to the Elite, whereas payments in the No Elite condition would not go to anyone. Decisions were incentivised; participants were aware that the result from one game would be randomly selected at the end of the experiment and they would be paid according to their monetary outcome in that game.

**Participants**

One-hundred participants were recruited by canvassing for interest on campus at the University of Sussex, through posters placed around campus, and through the School of Psychology's online subject database. Of the 87 participants who provided demographic information, 61 were female (70%), 79 were undergraduate students (91%), and of these 52 (65%) were psychology students. Participant ages ranged from 18 to 30 years old ($M$ = 20, $SD$

= 2.09). Participants were screened before participating to ensure they met our criteria for inclusion in the experiment: fluent/native English speakers, not studying an economic- or business-related discipline, and no history of gambling addiction. These criteria were specified because of the high importance of understanding the instructions, avoiding over-rational thinking, and protecting potentially vulnerable individuals, respectively. Participants were paid a minimum of £8.00 for their participation, but could leave with more money depending on the outcome of the games (see Materials and Procedure below). Participants were aware at the time of recruitment of the minimum they could receive and the potential for receiving more.

Note that no formal procedure was undertaken to decide on the size of the sample. Rather, sample size was determined by what was feasible in the time available to run the experiment; because the experiment involved groups of individuals interacting and each testing session required a minimum number of participants to go ahead, recruitment was quite challenging, resulting in only 100 participants recruited over a period of 3 months. The final sample of 100 participants was therefore not an a priori target but the number of participants that it was possible to recruit in the testing period. Post-hoc consideration suggests that at least 110 participants are required for 80% power in a repeated measures design including an interaction (Brysbaert, 2019). Thus, because the sample in the present study falls short of this number, it is likely the power achieved was less than 80%.

## Materials and Procedure

### *Overall procedure*

Upon arrival participants were directed into personal computer booths in the behavioural economics lab and directed to read instructions and provide consent. Contact between participants at this stage was minimised by presenting a PowerPoint slide on a screen giving directions and asking participants to remain in their booths and wait for the experiment to begin. I aimed to recruit 19 participants per session, equating to two Elites and a group of 17 non-Elites. Due to the challenges of recruiting this number of participants and the risk of no-shows, I allowed flexibility in the design so that the games would work over a range of group sizes.

Once all participants were believed to be present, I first provided a general explanation of what to expect in the experiment. This included explanation of the terminology used in the experiment, the general procedure (explained below), and the difference between Elite and non-Elite participants and how these roles were assigned. These instructions and all subsequent instructions were presented using animated PowerPoint presentations displayed on a screen that was visible to all participants. Questions were invited at several stages to allow participants to clarify their understanding. Following the general instructions, participants completed the real-effort task (described below) to determine their roles for the duration of the experiment.

I next explained the risk preference task (described below) and participants were invited to ask questions again before completing this task. As part of the instructions, I provided a demonstration of how pie charts could be used as a way of visualising the relative probabilities of each event. Participants were also provided with a printout that depicted as pie charts the range of probabilities that would be encountered in the experiment and were told they could refer to this at any point in the experiment. Following this, the experimental paradigms were administered. For each paradigm I provided instructions. Following the instructions, participants completed a series of comprehension questions via zTree designed to check their understanding of the paradigm. Participants could only progress by answering correctly and were encouraged to ask for help if required. This ensured that any misunderstandings were addressed and participants had a clear conception of the scenario and how decisions affected their own and others' outcomes. This exhaustive comprehension check was necessary because of the number of conditions and the complexity of the rules of each. In this way I could be confident that differences observed between paradigms could be attributed to the differing contingencies and not because of poor understanding.

Once all participants had finished the comprehension questions, the paradigm itself was run. This consisted of the two phases of the games detailed below. The process was repeated for each game. The games were grouped by the GVI factor and administered in a counter-balanced order such that individual and group scenarios were presented first an approximately equal number of times across sessions, and within the GVI grouping the order of presentation of Elite conditions was counter-balanced so that they were each administered first, second, and third an approximately equal number of times across sessions. This was done so that participants did not become overwhelmed with constantly changing scenarios and instructions could be delivered more succinctly. By grouping all three Elite scenarios within the GVI factor, it was possible to present, for example, the individual scenario with the Illegitimate Elite directly after the individual scenario with the Legitimate Elite by simply explaining that the scenario was the same except that the Elite type had changed.

Once all paradigms were complete, participants completed a short demographic questionnaire and were paid for their participation according to the random payoff selected by the program in zTree. The experiment took approximately 2 hours to complete (note that the experimental session included an additional three paradigms that are the subject of Study 2 of this thesis).

***Real-Effort task***

To determine roles in the experiment (see Procedure below) participants completed a real-effort addition task. This comprised of 10 randomly generated two-digit addition problems that participants were required to solve as accurately and quickly as possible. The two-top scorers were provisionally assigned to the Elite roles. The Elite roles were desirable because the Elites in the experiment stood to leave the experiment with considerably larger payments than the citizenry (see ‘The Games’ below). Upon completion of the real-effort task, participants who were not one of the two provisional Elites (i.e., who were not top scorers) were offered the opportunity to ‘steal’ the position from the second-highest scorer. One participant was randomly selected from those who chose the steal option, and this participant became the “Illegitimate” Elite for the session.

***Risk Preference task***

To calculate a measure of participants’ risk preferences in the decision space of the experiment, I constructed a non-social (i.e., involving only the decision-maker) task following the same format as the social tasks. In this task, each trial participants made a decision between keeping 50 coins for certain or choosing a risky option of keeping 100 coins with probability $P$ with a chance of keeping 0 coins with probability 1 - $P$. The task was comprised of 17 trials, covering each 0.05 increment of $P$ ranging from 0.05 to 0.85 (participants were unaware of the exact number of trials). The order of trials was randomised. The procedure for calculating individual risk preference based on responses in this task is described in the Results section.

***The Games***

Six games were described and presented as social systems in which participants made decisions. The two above-mentioned factors GVI and Elite determined the structure of each game: participants made decisions individually or as a group (GVI) and No Elite, the Legitimate Elite, or the Illegitimate Elite received payments (Elite). Each game consisted of two phases. Phase 1 presented the System (i.e., specifying the GVI and Elite condition) and participants were asked to decide either to pay 50 of their 100-coin endowment to the System, or to challenge the system by not paying for a chance of keeping all 100 coins. In Phase 2, participants were presented with a series of questions about the scenario. These were presented one after another on the participants' computers, in the following order. All games were one-shot decisions (i.e., only one decision per paradigm) to avoid the possibility of 'bet-hedging' within each game (for example, by making alternating decisions across a series of trials). No feedback about the outcome of decisions was provided so that knowledge of past outcomes could not influence decisions. To incentivise participants, the outcome of one game in the experimental session was randomly chosen at the end and converted to Great British Pounds and paid to participants in addition to their showup fee.

**Phase 1.**

***The Individual Games***. In the individual games, participants made decisions alone; they were told that their decisions could only affect their own outcomes. A choice to pay 50 (also referred to as "System Support") resulted in the participant keeping the remaining 50 for certain. A choice to pay 0 (also referred to as "System Challenge") introduced the possibility of 'overturning' the system, in which case a participant would successfully avoid paying and keep all 100 of their coins. However, paying 0 could also result in being unsuccessful and punished, in which case all 100 coins would be taken by the System. In the No Elite scenarios, participants were told that any money they pay or lose goes to the System; nobody receives it. For the Legitimate and Illegitimate Elite conditions, participants were told that any money they pay or lose goes to the Elite. Participants were told that the chance of the System being overturned was determined by a randomly selected number between 0 and the size of the group minus one (e.g., in a group of 16, the random number would range from 0 to 15). The highest random number represented an 85% chance of the System being overturned, and a random number of 0 corresponded to a chance equivalent to 85 divided by the size of the group (i.e., 85/group size). The range of the random number is specified in this way to provide a non-social analogue to the group scenario (see below). ***Table 1.1*** shows the expected value associated with each choice, assuming a group size of 17.

**Table 1.1.** *Expected payoffs for player i based on random number in the individual games.*

| Random number | Probability of overturn if player *i* challenges | System Challenge expected value | System Support expected value |
|---|---|---|---|
| 0 | 0.05 | 5 | 50 |
| 1 | 0.1 | 10 | 50 |
| 2 | 0.15 | 15 | 50 |
| 3 | 0.2 | 20 | 50 |
| 4 | 0.25 | 25 | 50 |
| 5 | 0.3 | 30 | 50 |
| 6 | 0.35 | 35 | 50 |
| 7 | 0.4 | 40 | 50 |
| 8 | 0.45 | 45 | 50 |
| 9 | 0.5 | 50 | 50 |
| 10 | 0.55 | 55 | 50 |
| 11 | 0.6 | 60 | 50 |
| 12 | 0.65 | 65 | 50 |
| 13 | 0.7 | 70 | 50 |
| 14 | 0.75 | 75 | 50 |
| 15 | 0.8 | 80 | 50 |
| 16 | 0.85 | 85 | 50 |

***The Group Games***. In the group games, participants made decisions together; they were told that their decisions could affect one another. Specifically, they were instructed that each circle who challenged the system by paying 0 would increase the likelihood of the System being overturned. To keep the minimum and maximum probabilities of overturn constant across sessions, the likelihood amount 'contributed' by each challenger was calculated by dividing the maximum possible likelihood (85) by the size of the group. Participants were told that the resulting figure, rounded to the nearest whole number, was the approximate amount that each challenger would increase the chance of overturn by approximately this rounded amount. The expected value associated with each choice (again assuming group size 17) is shown in ***Table 1.2***. As can be seen by comparing ***Table 1.1*** and ***Table 1.2***, the payoff structures for the individual and group games are identical. If a participant decides to pay 0 in the group game, and no others pay 0, the expected value of that decision would be 5 coins (row 1 in ***Table***

***1.2***). Likewise, the expected value of challenging in the individual game with random number 0 would be 5 coins. Thus, the number-overturn probability correspondence was identical across the two games, as were the resulting payoffs. This means that, in terms of rational thought, behaviour resulting from beliefs about numbers should be the same in both individual and group games.

**Table 1.2.** *Expected payoffs for player i based on number of other (j) players challenging the system*

| Number of other (*j*) challengers | Probability of overturn if player *i* challenges | System challenge expected value | System support expected value |
|---|---|---|---|
| 0 | 0.05 | 5 | 50 |
| 1 | 0.1 | 10 | 50 |
| 2 | 0.15 | 15 | 50 |
| 3 | 0.2 | 20 | 50 |
| 4 | 0.25 | 25 | 50 |
| 5 | 0.3 | 30 | 50 |
| 6 | 0.35 | 35 | 50 |
| 7 | 0.4 | 40 | 50 |
| 8 | 0.45 | 45 | 50 |
| 9 | 0.5 | 50 | 50 |
| 10 | 0.55 | 55 | 50 |
| 11 | 0.6 | 60 | 50 |
| 12 | 0.65 | 65 | 50 |
| 13 | 0.7 | 70 | 50 |
| 14 | 0.75 | 75 | 50 |
| 15 | 0.8 | 80 | 50 |
| 16 | 0.85 | 85 | 50 |

***Figure 1.1*** presents the payoff structure for the individual and group scenarios (assuming 17 participants) graphically. For any group size, there is a value of the random number/number of others challenging below which supporting the system yields greater expected value and above which challenging the system yields greater value. In the example of a group size of 17, this value is 9.

Note that in Elite conditions (and at the start of the experiment) the Elite's payoffs were made clear to emphasise the unfair wealth distribution. Elites, like non-Elites, started each game with a 100-coin endowment. In addition, if the system was not overturned, Elites received 50 coins from each non-Elite who paid 50, and 100 coins from each non-Elite who tried to avoid paying. This meant it was possible for Elites to receive considerably more money than non-Elites. For example, in a group of 17, if five non-Elites challenged and 12 supported and the system was not overturned, the Elite would receive 1,200 coins (100 [endowment] + 5 x 100 [from challengers] + 12 x 50 [from supporters]), which would equate to £96.00 (contrast with the £8.00 maximum payoff that non-Elites could receive). If the system was overturned, the Elite would not receive 100 from challengers but would still keep the 50 from supporters. Thus, if in the above scenario the system was overturned, the Elite would receive 700 coins (100 [endowment] + 12 x 50 [from supporters]), which would equate to £56.00.

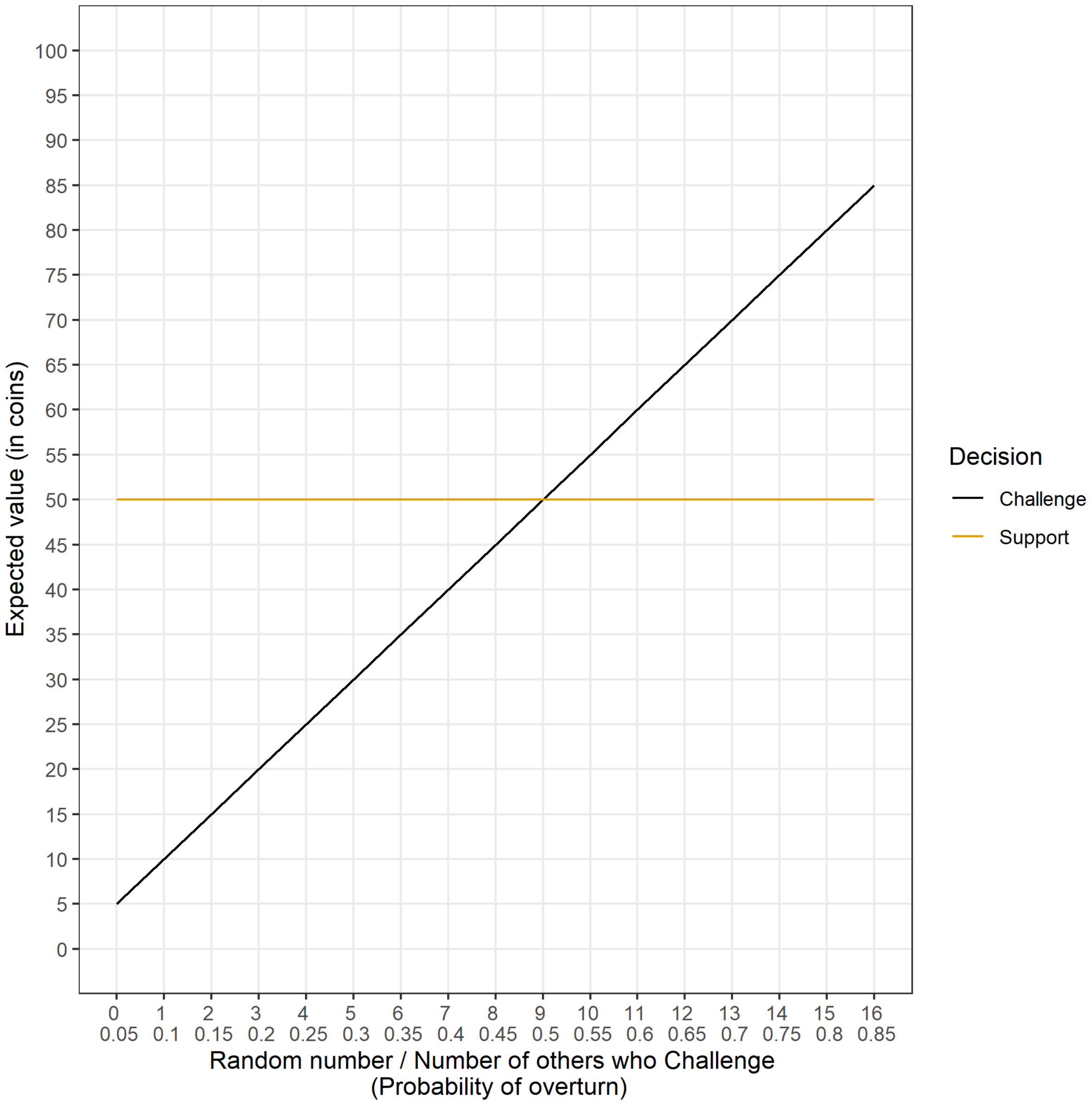


*Figure 1.1.* Expected value of each decision for player i based on the random number (individual scenarios) or the number of others who challenge (group scenarios).

**Phase 2**. Following the decision phase of each game, participants provided further ratings about their experience in the game.

***Identification***. A screen was presented with numbers 1 to 7, each with a radio button beneath. "Fully disagree" was placed at the left extreme (1) and "Fully agree" at the right (7). The text read "Indicate, by clicking one of the circles, the degree to which you agree with the following statement: In the round just played, I identified with the other circles." Participants could only select one option. I chose to measure identification in this way because of the simplicity with which it could be deployed across multiple paradigms. Though it might not capture identification in all its qualitative depth, this single-item measure has been shown to be valid and reliable across a variety of contexts (Postmes et al., 2013).

***Number belief.*** For the group scenario, the screen presented two horizontal lines with number increments of 1 from 0 to the size of the group minus one. The top line was labelled "Number of other circles I think decided to pay 50 coins" and the bottom line was labelled "Number of other circles I think decided to pay 0 coins". Participants were asked to "indicate using the sliders how many other circles you think decided to pay 50 coins and how any other circles you think decided to pay 0." The screen also explained that the total number of other circles was equal to the size of the group, so the two numbers indicated on the slider should sum to the value of the size of the group minus one. Participants could drag the sliders to the desired point. The slider not dragged automatically adjusted its value to add up to the size of the group minus one. So, if a participant in a group size of 10 selected 4 on one slider, the other slider would automatically adjust to value 5. For the individual game, there was only one slider on screen, and participants were asked to indicate what random number they believe occurred. The random number ranged from 0 to the size of the group minus one.

Because Number belief by proxy represented the belief in the likelihood of overturning the system, it is also referred to as "efficacy" throughout this paper and Study 2.

***Fairness.*** On a 1-7 scale as described for identification, participants were asked to "indicate, by clicking one of the radio buttons, the degree to which you agree with the following statement: In the round just played, I felt the system was fair".

Once all experimental paradigms were completed, participants were asked to complete a short demographic questionnaire and were paid for their participation.

## Results

Analyses were conducted using RStudio. The majority of analyses were achieved using the 'lme4' package (version 1.1.26; Bates et al., 2015). I also made extensive use of 'ggplot2' (version 3.3.3; Wickham, 2016) and 'ggeffects' (version 1.0.1; Lüdecke, 2018) for visualisation, and 'emmeans' (version 1.5.4; Lenth et al., 2021) for post-hoc analysis. For each model presented I ran a series of diagnostic tests to check that underlying model assumptions were met. In the interests of space, I only report where diagnostic tests flagged potential issues. For all analyses I used the 'bobyqa' optimizer and set the maximum number of iterations to 200,000.

### Computed variables

#### *Number belief*

Number belief was expressed as a proportion of group size minus one. For example, for a given decision made as part of a group with nine other circles, if a participant believed that 5 other circles would pay 0, their number belief would be transformed into a proportion like so: 5 / 9 = 0.56. This value was then multiplied by 100, so that an increment of 1 corresponded to a proportion change of 1% (this was done so that odds ratios were more readily interpretable;

without doing so would mean that an increment of 1 would correspond to a proportion change of 100%).

***Risk preference***

A measure of risk preference was constructed based on responses in the Risk preference task. I opted to construct this measure instead of using existing risk preference indexes (e.g., Eckel & Grossman, 2008) because decisions in the Risk preference task are procedurally very similar to the decisions made in the experimental games and thus preferences for risky versus safe options in this task are a more appropriate estimate of baseline risk preference in this decision context. To calculate Risk preference, participants' responses in the Risk preference task were coded according the degree of risk seeking or risk avoidance they represented. First, decisions to choose the risky 100 option were coded as 'risky' and decisions to choose the certain 50 option were coded as 'non-risky'. Risk-seeking was quantified as

$$risk\ seeking = 1 - P(chosen\ risky\ option) - 0.1$$

The extra subtraction of 0.1 was included so that the maximum and minimum possible values for risk-seeking were the same as those for risk avoidance. Risk avoidance was quantified as

$$risk\ avoidance = 1 - (1 - P(non - chosen\ risky\ option))$$

To clarify, consider a choice between the certain option and a more rewarding but risky option with probability 0.7. If the risky option is chosen, risk-seeking for this choice would be quantified as 1 – 0.7 – 0.1 = 0.2, reflecting the relatively low risk associated with the choice (recall the maximum probability of a risky option is always 0.85). If the non-risky option is chosen, risk-avoidance would instead be quantified as 1 – (1 – 0.7) = 0.7, reflecting the relatively risk avoidant nature of the choice. To create the final measure of Risk preference, the values for risk-seeking and risk-avoidance for each choice were summed and the total risk avoidance score was subtracted from the total risk-seeking score:

$$risk\ preference = (\sum risk\ seeking) - (\sum risk\ avoidance)$$

In this way, individuals scoring above 0 are relatively risk-seeking and individuals scoring below zero are relatively risk-avoidant.

**System Challenge decisions**

I first investigated what effect playing in a group versus playing individually (i.e., GVI) and what type of Elite (i.e., No Elite versus Legitimate Elite versus Illegitimate Elite) had on decisions to challenge the System using Generalised Linear Mixed Modelling (GLMM). The binary dependent variable was termed “System Challenge” and its levels were coded “Support” (i.e., pay 50; reference level) and “Challenge” (i.e., pay 0 to the system). First, as above, I specified a saturated model with GVI, Elite, and a GVI x Elite interaction term predicting System Challenge. Risk preference and group size were included as covariates. In this saturated model intercepts were set to vary randomly by participant, and random slopes were specified for both GVI and Elite. From this saturated start point, I reduced the complexity of the model to allow it to converge, starting with the random effects explaining the least variance in the outcome variable. Only one step was required; removing Elite as a random effect allowed the model to converge.

After controlling for Risk preference and Group size, GVI was a significant positive predictor of SC (*odds ratio (OR)* = 8.92, $z$ = 4.49, $p$ < .001); the odds of a circle deciding to challenge the system were 8.92 times higher in the group scenario compared to the individual scenario. No significant effect of Elite was observed. All levels of the GVI x Elite interaction were non-significant. ***Table 1.3*** shows the estimates for all parameters. ***Figure 1.2*** depicts the effects graphically.

**Table 1.3.** *Generalised linear mixed effects model predicting System Challenge*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 1.97 | 0.31 – 12.51 | 0.471 |
| Risk preference | 1.47 | 1.23 – 1.77 | **< 0.001** |
| Group size | 0.92 | 0.81 – 1.06 | 0.252 |
| GVI† | 8.92 | 3.43 – 23.18 | **< 0.001** |
| Legitimate elite†† | 1.90 | 0.90 – 4.03 | 0.095 |
| Illegitimate elite†† | 1.44 | 0.68 – 3.06 | 0.342 |
| Legit. x GVI | 0.41 | 0.13 – 1.23 | 0.112 |
| Illegit. x GVI | 0.44 | 0.14 – 1.38 | 0.160 |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 1.89 | | |
| $\tau_{11\ \text{Subject.GVI}}$ | 3.29 | | |
| $\rho_{01\ \text{Subject}}$ | 0.02 | | |
| ICC | 0.52 | | |
| $N_{\text{Subject}}$ | 91 | | |
| Observations | 533 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.197 / 0.613 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

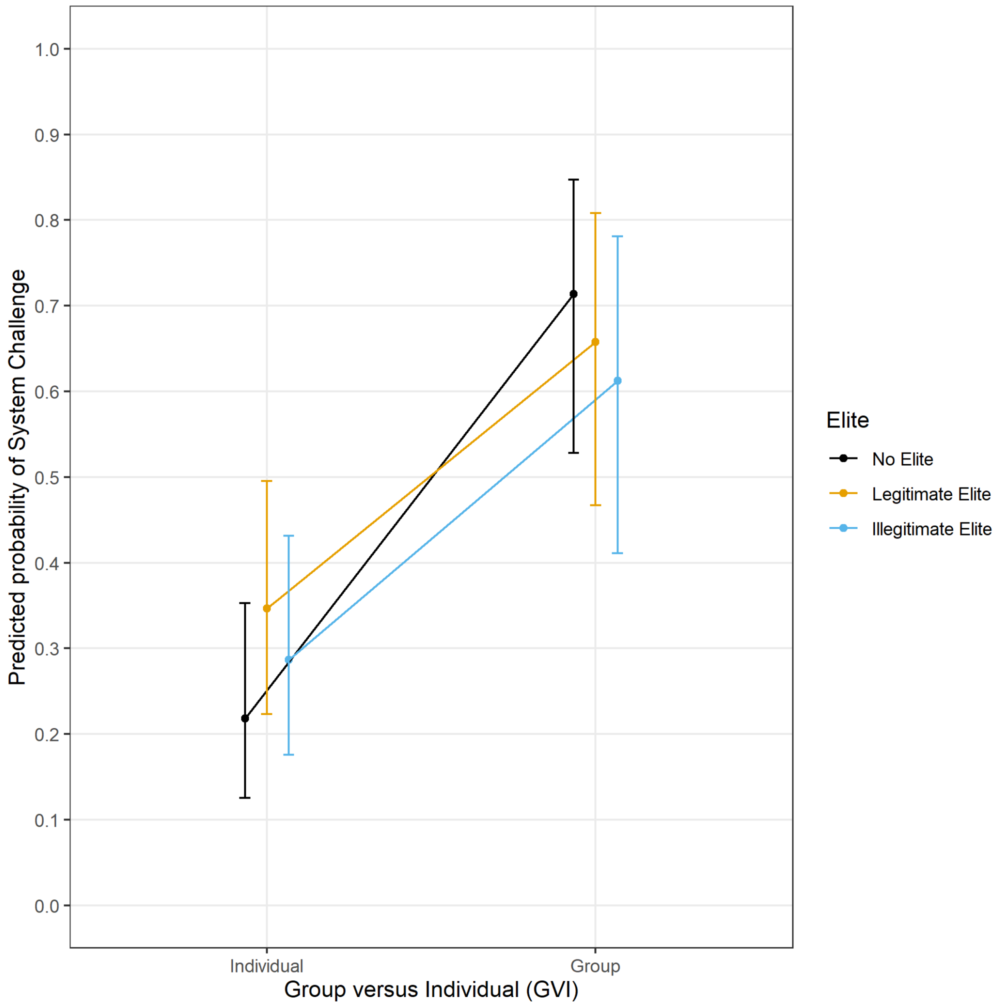


*Figure 1.2*. Predicted probability of a participant challenging the system as a function of GVI and Elite conditions. Only the main effect of GVI was significant. Error bars represent the 95% confidence interval.

Post-hoc contrasts showed that in all Elite conditions, participants were significantly more likely to challenge the system in the group scenarios compared to the individual scenarios (***Table 1.4***). There were no differences between Elite conditions for either individual or group scenarios (***Table 1.5***).

**Table 1.4.** *Post-hoc contrasts comparing Group versus Individual scenarios for each Elite condition*

| *Contrast* | *Elite* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| Individual - Group | No Elite | -2.188 | 0.488 | Inf | -4.487 | **0.000** |
| Individual - Group | Legitimate Elite | -1.286 | 0.458 | Inf | -2.809 | **0.005** |
| Individual - Group | Illegitimate Elite | -1.367 | 0.477 | Inf | -2.866 | **0.004** |

**Table 1.5.** *Post-hoc contrasts comparing Elite conditions for Individual and Group scenarios*

| *Contrast* | *GVI* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legitimate Elite | Individual | -0.642 | 0.384 | Inf | -1.672 | 0.216 |
| No Elite - Illegitimate Elite | Individual | -0.364 | 0.384 | Inf | -0.950 | 0.609 |
| Legitimate Elite - Illegitimate Elite | Individual | 0.278 | 0.374 | Inf | 0.743 | 0.738 |
| No Elite - Legitimate Elite | Group | 0.260 | 0.417 | Inf | 0.622 | 0.808 |
| No Elite - Illegitimate Elite | Group | 0.456 | 0.439 | Inf | 1.040 | 0.552 |
| Legitimate Elite - Illegitimate Elite | Group | 0.197 | 0.433 | Inf | 0.454 | 0.893 |

**Subjective experience of the System**

I next investigated the influence of GVI and Elite on the ratings of Identification, Number belief, and Fairness. For each analysis, I first specified a saturated linear mixed effects model (LMM) including GVI (levels: Individual [reference], Group), Elite (levels: No Elite [reference], Legitimate Elite, Illegitimate Elite), and a GVI x Elite interaction term, allowing correlated random slopes for GVI and Elite. The reference group for the GVI term was 'Individual'. The reference group for the Elite term was 'No Elite.' In all models I included a term to control for group size. I then removed parameters in order of complexity, starting with the random slopes, until the model successfully converged.

### *Fairness*

System Fairness was positively predicted by GVI ($b$ = 0.57, $t$(194.87) = 3.65, $p$ < .001); circles rated the system as fairer when they played as a group compared to alone. Both Legitimate ($b$ = -0.77, $t$(146.74) = -4.4, $p$ < .001) and Illegitimate ($b$ = -0.91, $t$(133.84) = -5.02, $p$ < .001) Elite conditions negatively predicted Fairness, compared to the Elite absent condition; when in a system with either the Legitimate or Illegitimate Elite, circles rated the system as less fair than when they were in a system with no triangle present (***Figure 1.3***). All combinations of the GVI x Elite interaction were non-significant (see ***Table 1.6***).

**Table 1.6.** *Linear mixed effects model predicting Fairness*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 3.42 | 3.04 – 3.79 | **< 0.001** |
| Group size | -0.02 | -0.12 – 0.07 | 0.616 |
| GVI† | 0.57 | 0.26 – 0.88 | **< 0.001** |
| Legitimate Elite†† | -0.77 | -1.11 – -0.43 | **< 0.001** |
| Illegitimate Elite†† | -0.91 | -1.27 – -0.56 | **< 0.001** |
| Legit. x GVI | -0.04 | -0.37 – 0.28 | 0.793 |
| Illegit. x GVI | -0.15 | -0.49 – 0.19 | 0.402 |
| **Random Effects** | | | |
| $\sigma^2$ | 0.64 | | |
| $\tau_{00}$ Subject | 2.65 | | |
| $\tau_{11}$ Subject.GVI | 0.95 | | |
| $\tau_{11}$ Subject.LegitimateElite | 1.50 | | |
| $\tau_{11}$ Subject.IllegitimateElite | 1.73 | | |
| $\rho_{01}$ | -0.22 | | |
| | -0.51 | | |
| | -0.60 | | |
| ICC | 0.78 | | |
| N Subject | 91 | | |
| Observations | 533 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.081 / 0.798 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

Post-hoc Tukey contrasts revealed that both systems with a Legitimate Elite ($z$ = -4.4, $p < .001$) and systems with an Illegitimate Elite ($z$ = -5.02, $p < .001$) were rated as less fair than systems with No Elite. There was no difference in Fairness ratings between Legitimate and Illegitimate elite conditions ($z$ = -1.14, $p$ = 0.487), indicating that participants considered the Legitimate and Illegitimate elite conditions equally less fair than the elite absent condition.

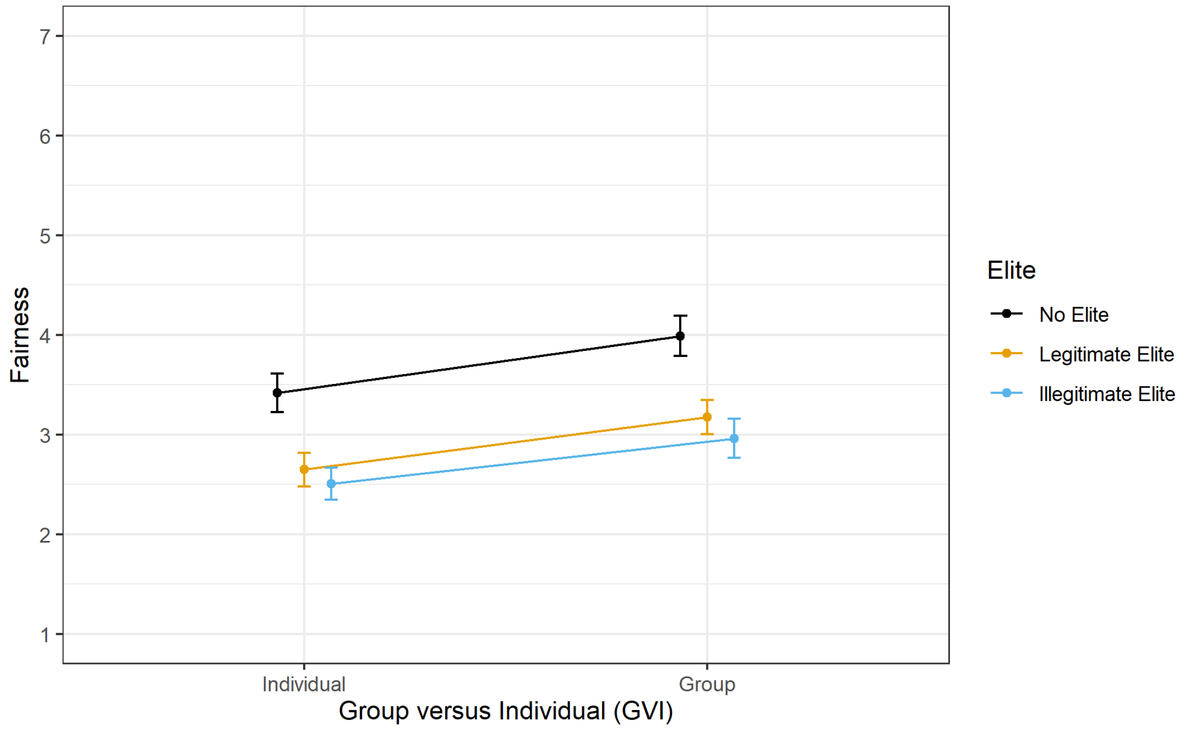


*Figure 1.3*. Mean Fairness ratings in each GVI and Elite condition. The main effects of both GVI and Elite were significant. Error bars represent within-subject standard error of the mean.

### ***Identification***

As expected, Identification was positively predicted by GVI ($b$ = 0.95, $t$(196.68) = 4.13, $p$ = < .001); circles identified more with one another when they played as a group compared to when they played alone. The Elite manipulation and the Elite x GVI interaction had no influence on Identification ratings (see ***Table 1.7***; ***Figure 1.4***). No other terms in the model had an influence on Identification ratings.

**Table 1.7.** *Linear mixed effects model predicting Identification*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 3.23 | 2.83 – 3.64 | **< 0.001** |
| Group size | 0.02 | -0.08 – 0.12 | 0.713 |
| GVI† | 0.95 | 0.50 – 1.39 | **< 0.001** |
| Legitimate Elite†† | 0.25 | -0.07 – 0.58 | 0.127 |
| Illegitimate Elite†† | -0.13 | -0.46 – 0.19 | 0.425 |
| Legit. x GVI | -0.12 | -0.58 – 0.34 | 0.605 |
| Illegit. x GVI | 0.36 | -0.11 – 0.83 | 0.131 |
| **Random Effects** | | | |
| $\sigma^2$ | 1.25 | | |
| $\tau_{00}$ Subject | 2.62 | | |
| $\tau_{11}$ Subject.GVI | 2.27 | | |
| $\rho_{01}$ Subject | -0.53 | | |
| ICC | 0.66 | | |
| N Subject | 91 | | |
| Observations | 533 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.070 / 0.687 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

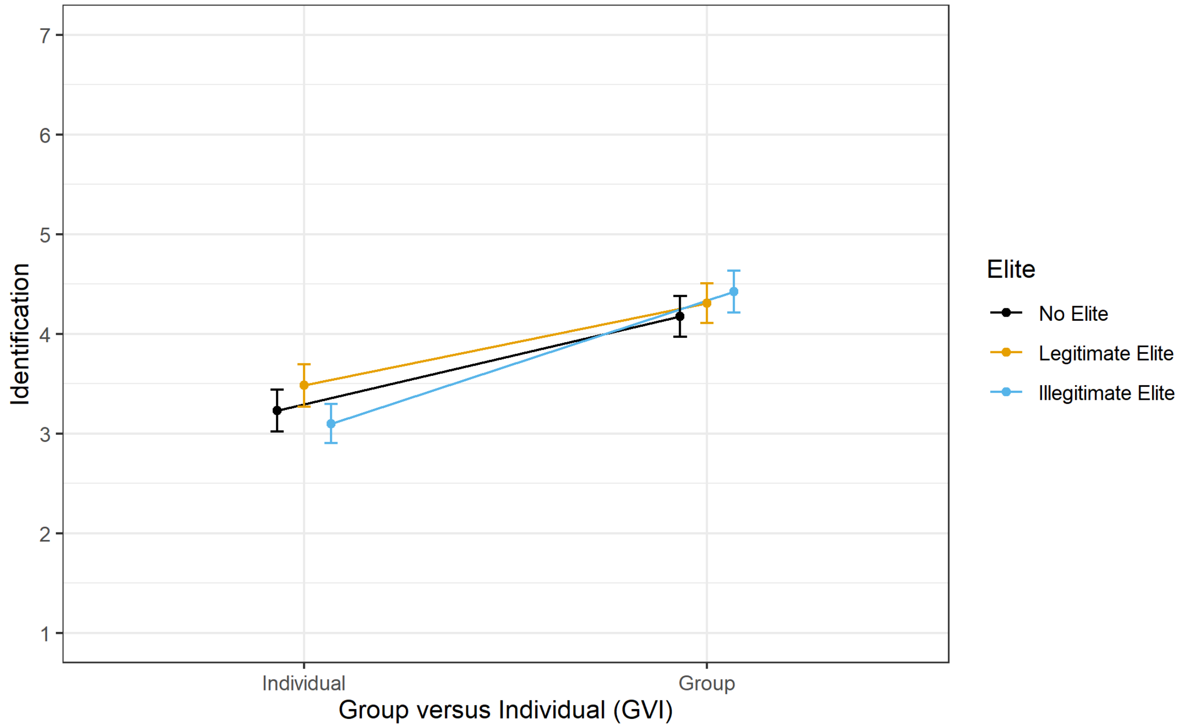


*Figure 1.4.* Mean Identification ratings in each GVI and Elite condition. Only the main effect of GVI was significant. Error bars represent standard error of the mean.

### *Number belief*

A LMM was created to regress the GVI and Elite manipulations onto Number belief. The analysis was conducted in the same way as described for the previous models. Number belief was positively predicted by GVI ($b$ = 0.11, $t$(256.66) = 3.08, $p$ = 0.002); when playing as part of a group circles believed that the number of others paying 0 (i.e., challenging the system) was significantly higher compared to their beliefs about what random number occurred in the individual scenarios. Elite and all levels of the GVI x Elite interaction were non-significant (see ***Table 1.8***; ***Figure 1.5***).

**Table 1.8.** *Linear mixed effects model predicting Number belief*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.46 | 0.41 – 0.51 | < **0.001** |
| Group size | -0.00 | -0.01 – 0.01 | 0.879 |
| GVI† | 0.11 | 0.04 – 0.17 | **0.002** |
| Legitimate Elite†† | 0.02 | -0.04 – 0.07 | 0.592 |
| Illegitimate Elite†† | 0.02 | -0.04 – 0.07 | 0.548 |
| Legit. x GVI | -0.06 | -0.14 – 0.02 | 0.169 |
| Illegit. x GVI | 0.01 | -0.07 – 0.09 | 0.791 |
| **Random Effects** | | | |
| $\sigma^2$ | 0.04 | | |
| $\tau_{00}$ Subject | 0.01 | | |
| $\tau_{11}$ Subject.GVI | 0.03 | | |
| $\rho_{01}$ Subject | -0.22 | | |
| ICC | 0.40 | | |
| N Subject | 91 | | |
| Observations | 533 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.038 / 0.420 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

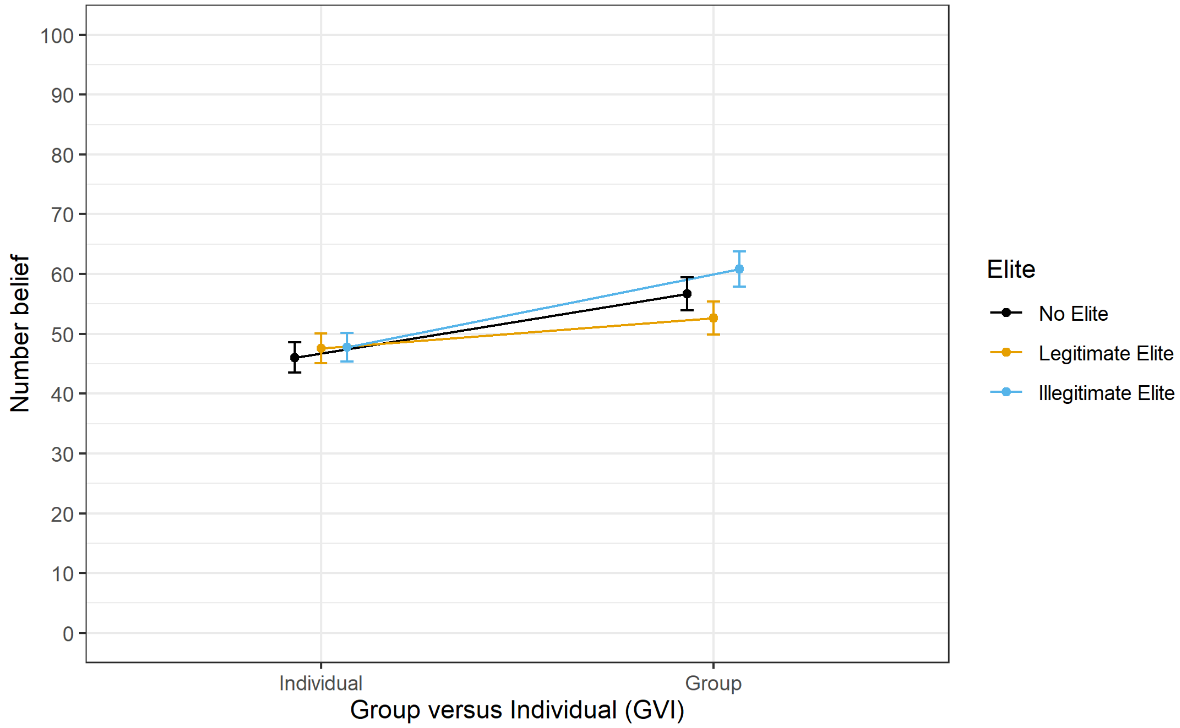


*Figure 1.5*. Mean Number belief ratings in each GVI and Elite condition. Only the main effect of GVI was significant. Error bars represent within-subject standard error of the mean.

## Influence of subjective experiences on System Challenge decisions

To investigate the possible motivations behind decisions to challenge the system, I created further models predicting System Challenge from each of Number belief, Identification, and Fairness and their interactions with GVI and Elite. I did this by first adding each variable as a main effect and testing for improvement over the above base model. For each model, I then tested for the interaction of each variable with GVI and Elite. The results of these steps for each variable are described below.

### *Number belief*

The modelling procedure showed that a model including a Number belief x GVI interaction was a better fit than a model with just the main effect of Number belief, $\chi^2(1, 91) = 7.26$, $p = .007$, which itself was a better fit than the base model, $\chi^2(1, 91) = 93.92$, $p < .001$ (***Table 1.9***).

**Table 1.9.** *Generalised linear mixed effects models predicting System Challenge from GVI, Elite, and Number belief*

| | **Step 0** | | | **Step 1** | | | **Step 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* | *Odds Ratios* | *CI (95%)* | *p* | *Odds Ratios* | *CI (95%)* | *p* |
| Intercept | 1.97 | 0.31 – 12.51 | 0.471 | 0.07 | 0.01 – 0.57 | **0.013** | 0.22 | 0.03 – 1.81 | 0.158 |
| Group size | 0.92 | 0.81 – 1.06 | 0.252 | 0.93 | 0.81 – 1.07 | 0.317 | 0.93 | 0.80 – 1.07 | 0.287 |
| Risk preference | 1.47 | 1.23 – 1.77 | < **0.001** | 1.44 | 1.18 – 1.75 | < **0.001** | 1.47 | 1.22 – 1.78 | < **0.001** |
| GVI† | 8.92 | 3.43 – 23.18 | < **0.001** | 7.13 | 2.59 – 19.61 | < **0.001** | 0.89 | 0.15 – 5.38 | 0.895 |
| Legitimate elite†† | 1.90 | 0.90 – 4.03 | 0.095 | 2.05 | 0.85 – 4.97 | 0.111 | 1.91 | 0.84 – 4.33 | 0.122 |
| Illegitimate elite†† | 1.44 | 0.68 – 3.06 | 0.342 | 1.41 | 0.58 – 3.39 | 0.449 | 1.37 | 0.61 – 3.10 | 0.444 |
| Legit. x GVI | 0.41 | 0.13 – 1.23 | 0.112 | 0.50 | 0.14 – 1.70 | 0.263 | 0.56 | 0.16 – 1.94 | 0.357 |
| Illegit. x GVI | 0.44 | 0.14 – 1.38 | 0.160 | 0.35 | 0.10 – 1.25 | 0.107 | 0.29 | 0.08 – 1.07 | 0.063 |
| Number belief | | | | 1.06 | 1.05 – 1.08 | < **0.001** | 1.04 | 1.03 – 1.06 | < **0.001** |
| GVI x Number belief | | | | | | | 1.04 | 1.01 – 1.07 | **0.010** |

**Random Effects**

| | | | |
|---|---|---|---|
| $\sigma^2$ | 3.29 | 3.29 | 3.29 |
| $\tau_{00}$ | 1.89 $_{\text{Subject}}$ | 3.27 $_{\text{Subject}}$ | 2.24 $_{\text{Subject}}$ |
| $\tau_{11}$ | 3.29 $_{\text{Subject.GVI}}$ | 2.22 $_{\text{Subject.GVI}}$ | 2.21 $_{\text{Subject.GVI}}$ |
| $\rho_{01}$ | 0.02 $_{\text{Subject}}$ | -0.55 $_{\text{Subject}}$ | -0.16 $_{\text{Subject}}$ |
| ICC | 0.52 | 0.47 | 0.47 |
| N | 91 $_{\text{Subject}}$ | 91 $_{\text{Subject}}$ | 91 $_{\text{Subject}}$ |
| Observations | 533 | 533 | 533 |
| Marginal $R^2$ / Conditional $R^2$ | 0.197 / 0.613 | 0.423 / 0.694 | 0.458 / 0.715 |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

The final model indicated that Number belief was significantly predictive of System Challenge in both Individual and Group scenarios, but that it had a stronger influence when participants played in groups ($trend_{Individual}$ = .04, $z$ = 4.71, $p$ < .001; $trend_{Group}$ = .08, $z$ = 6.32, $p$ < .001). ***Figure 1.6*** shows this effect.

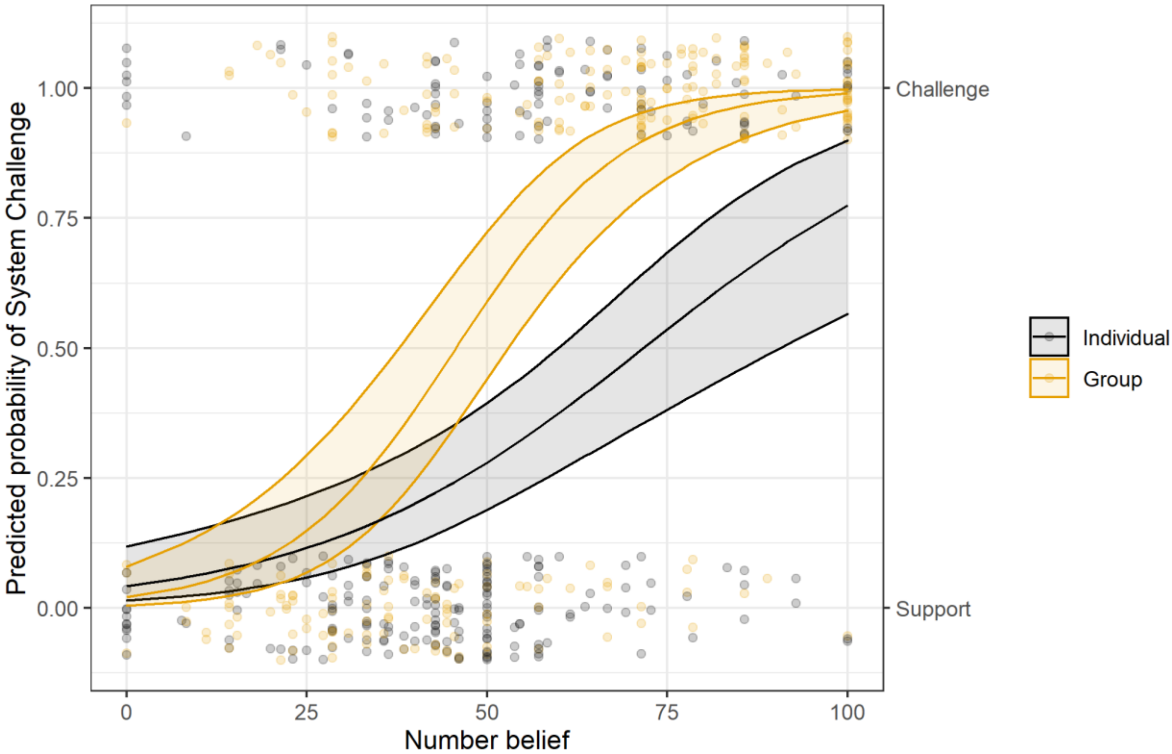


*Figure 1.6*. Predicted probabilities of System Challenge based on Number belief and Group versus Individual scenarios. Shaded areas represent the 95% confidence interval. Data points (jittered vertically) correspond to the right vertical axis and depict the raw decisions.

The finding that Number belief was more influential in Group compared to Individual scenarios warranted further explanation. Because Number belief corresponded to the believed probability that the System would be overturned (see Methods section), in terms of expected value it should be equally predictive in both scenarios. That is, a Number belief of 50 in the Individual scenario represents the same belief in the probability of overturn as a Number belief of 50 in the Group scenario. Thus, if decisions were based solely on the expected value of Number belief, they should have been equally influenced by it in individual and group scenarios. The fact that Number belief was more influential in group scenarios suggests that processes beyond expected value were at work.

A prime candidate for such a process was Identification. I therefore tested whether Identification moderated the GVI x Number belief interaction by adding Identification and its two-way interactions with GVI and Number belief as well as the three-way GVI x Number belief x Identification interaction. To avoid scaling issues, the Number belief variable was divided by 10 so that its range (0-10) was on a similar scale to that of Identification (1-7). With Number belief defined in this way, an increment change of 1 represented a proportion change in Number belief of 10%.

The resulting model showed that Identification did indeed moderate the interaction of GVI with Number belief (***Table 1.10***). Post-hoc tests revealed that there was no difference in the influence of Number belief between Individual and Group scenarios for low levels of Identification (Identification = 1-3), but a difference was evident for moderate and high levels (Identification = 4-7), with higher levels of Identification associated with larger differences (***Table 1.11***). Trend analysis showed that all trends were significant apart from the Group trend at Identification = 1 and the Individual trend at Identification = 7 (***Table 1.12***). ***Figure 1.7*** shows the three-way interaction observed (see Supplementary materials for animated gif version that aids viewing). This finding suggests that greater Identification endowed the expected value of Number belief with greater utility in group scenarios compared to individual scenarios.

**Table 1.10.** *Generalised linear mixed effects model predicting System Challenge from GVI, Elite, Number belief, Identification, and the interactions between GVI, Number belief and Identification*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.07 | 0.00 – 1.00 | 0.050 |
| Group size | 0.91 | 0.79 – 1.05 | 0.218 |
| Risk preference | 1.45 | 1.20 – 1.75 | **< 0.001** |
| GVI† | 9.12 | 0.27 – 310.14 | 0.219 |
| Legitimate elite†† | 1.76 | 0.77 – 4.03 | 0.181 |
| Illegitimate elite†† | 1.44 | 0.63 – 3.28 | 0.392 |
| Number belief ⁕ | 1.63 | 1.16 – 2.31 | **0.005** |
| Identification | 1.45 | 0.87 – 2.42 | 0.155 |
| Legit. x GVI | 0.47 | 0.13 – 1.71 | 0.249 |
| Illegit. x GVI | 0.21 | 0.05 – 0.85 | **0.029** |
| GVI x Number belief | 0.72 | 0.39 – 1.33 | 0.297 |
| GVI x Identification | 0.57 | 0.25 – 1.29 | 0.179 |
| Number belief x Identification | 0.98 | 0.90 – 1.07 | 0.702 |
| GVI x Number belief x Identification | 1.19 | 1.02 – 1.40 | **0.027** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00}$ Subject | 2.22 | | |
| $\tau_{11}$ Subject.GVI | 2.39 | | |

| | |
|---|---|
| $\rho_{01\ \text{Subject}}$ | -0.24 |
| ICC | 0.46 |
| $N_{\text{Subject}}$ | 91 |
| Observations | 533 |
| Marginal $R^2$ / Conditional $R^2$ | 0.577 / 0.774 |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

⁕ Variable scaled by dividing by 10

**Table 1.11.** *Post-hoc estimations of the trend of Number belief for each GVI by Identification combination*

| *Identification* | *GVI* | *Number belief trend* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| 1 | Individual | 0.474 | 0.141 | Inf | 3.372 | **0.001** |
| 1 | Group | 0.327 | 0.205 | Inf | 1.596 | 0.110 |
| 2 | Individual | 0.457 | 0.111 | Inf | 4.105 | **0.000** |
| 2 | Group | 0.488 | 0.163 | Inf | 2.985 | **0.003** |
| 3 | Individual | 0.440 | 0.095 | Inf | 4.623 | **0.000** |
| 3 | Group | 0.648 | 0.142 | Inf | 4.550 | **0.000** |
| 4 | Individual | 0.422 | 0.099 | Inf | 4.284 | **0.000** |
| 4 | Group | 0.808 | 0.151 | Inf | 5.351 | **0.000** |
| 5 | Individual | 0.405 | 0.120 | Inf | 3.374 | **0.001** |
| 5 | Group | 0.969 | 0.185 | Inf | 5.231 | **0.000** |
| 6 | Individual | 0.388 | 0.152 | Inf | 2.550 | **0.011** |
| 6 | Group | 1.129 | 0.234 | Inf | 4.828 | **0.000** |
| 7 | Individual | 0.371 | 0.190 | Inf | 1.958 | 0.050 |
| 7 | Group | 1.289 | 0.290 | Inf | 4.448 | **0.000** |

**Table 1.12.** *Post-hoc contrasts of the influence of Number belief across Individual and Group scenarios*

| *Identification* | *Contrast* | *Estimate* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| 1 | Individual - Group | 0.147 | 0.247 | Inf | 0.594 | 0.552 |
| 2 | Individual - Group | -0.031 | 0.195 | Inf | -0.157 | 0.875 |
| 3 | Individual - Group | -0.208 | 0.168 | Inf | -1.240 | 0.215 |
| 4 | Individual - Group | -0.386 | 0.176 | Inf | -2.187 | **0.029** |
| 5 | Individual - Group | -0.563 | 0.216 | Inf | -2.602 | **0.009** |
| 6 | Individual - Group | -0.741 | 0.275 | Inf | -2.697 | **0.007** |
| 7 | Individual - Group | -0.918 | 0.342 | Inf | -2.685 | **0.007** |

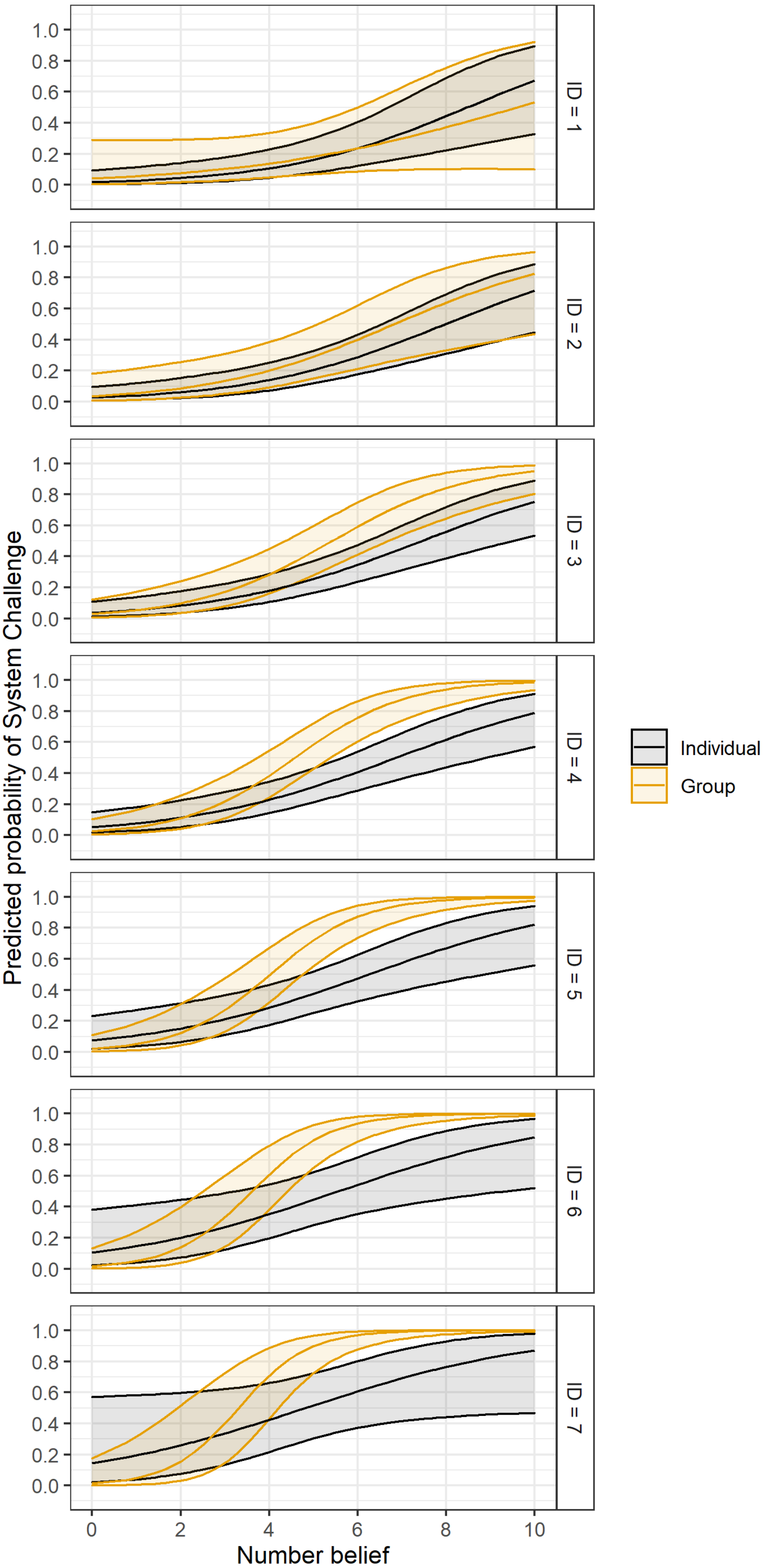


*Figure 1.7.* The three-way interaction of GVI, Number belief, and Identification predicting decisions to Challenge. A significant difference in the influence of Number belief between Individual and Group scenarios emerged at Identification = 4 and above. Shaded areas represent the 95% confidence interval.

***Identification***

The best-fitting Identification model contained a three-way interaction between Identification, GVI, and Elite. This was a better fit than the model with the GVI x Identification interaction, $\chi^2(4, 91) = 18.05$, $p = .001$, and the model with the Elite x Identification interaction, $\chi^2(3) = 12.82$ , $p = .005$, which themselves were better fits than the main effect model (GVI x Identification versus main effect model: $\chi^2(1, 91) = 3.90$, $p = .048$; Elite x Identification versus main effect model: $\chi^2(2, 91) = 9.13$, $p = .010$), which was a better fit than the base model, $\chi^2(1, 91) = 28.17$, $p < .001$. ***Table 1.13*** shows the model construction from base model to the GVI x Identification model. ***Table 1.14*** shows the progression from base model to the Elite x Identification model. ***Table 1.15*** shows the final model with the three-way interaction. ***Figure 1.8*** depicts the three-way interaction. In Individual scenarios, Identification was only predictive of System Challenge when there was a Legitimate Elite present (left column, middle row). In Group scenarios, Identification predicted Challenge when there was a Legitimate or Illegitimate Elite but not when there was no Elite (compare bottom two rows to top row in right column). Thus, it appeared that identification was primarily influential in the group scenarios when an Elite was present. ***Table 1.16***, ***Table 1.17***, *and* ***Table 1.18*** present the results of post-hoc comparisons.

***Fairness***

The final Fairness model was not an improvement over the base model, $\chi^2(1, 91) = 0.010$, $p = 0.922$).

**Table 1.13.** *Generalised linear mixed effects model predicting System Challenge from GVI, Elite, and Identification and its interaction with GVI*

| | **Step 0** | | | **Step 1** | | | **Step 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* | *Odds Ratios* | *CI (95%)* | *p* | *Odds Ratios* | *CI (95%)* | *p* |
| Intercept | 1.97 | 0.31 – 12.51 | 0.471 | 0.35 | 0.05 – 2.58 | 0.306 | 0.67 | 0.10 – 4.55 | 0.678 |
| Group size | 0.92 | 0.81 – 1.06 | 0.252 | 0.92 | 0.80 – 1.06 | 0.239 | 0.92 | 0.81 – 1.05 | 0.204 |
| Risk preference | 1.47 | 1.23 – 1.77 | **< 0.001** | 1.39 | 1.15 – 1.67 | **0.001** | 1.40 | 1.17 – 1.66 | **< 0.001** |
| GVI† | 8.92 | 3.43 – 23.18 | **< 0.001** | 6.35 | 2.54 – 15.87 | **< 0.001** | 1.84 | 0.41 – 8.20 | 0.425 |
| Legitimate elite†† | 1.90 | 0.90 – 4.03 | 0.095 | 1.79 | 0.81 – 3.93 | 0.148 | 1.77 | 0.83 – 3.77 | 0.141 |
| Illegitimate elite†† | 1.44 | 0.68 – 3.06 | 0.342 | 1.60 | 0.73 – 3.52 | 0.241 | 1.52 | 0.71 – 3.25 | 0.282 |
| Legit. X GVI | 0.41 | 0.13 – 1.23 | 0.112 | 0.41 | 0.13 – 1.26 | 0.120 | 0.40 | 0.13 – 1.23 | 0.111 |
| Illegit. X GVI | 0.44 | 0.14 – 1.38 | 0.160 | 0.37 | 0.11 – 1.16 | 0.088 | 0.36 | 0.11 – 1.14 | 0.081 |
| Identification | | | | 1.57 | 1.31 – 1.88 | **< 0.001** | 1.35 | 1.10 – 1.67 | **0.004** |
| GVI x Identification | | | | | | | 1.37 | 1.00 – 1.88 | **0.049** |

**Random Effects**

| | | | |
|---|---|---|---|
| $\sigma^2$ | 3.29 | 3.29 | 3.29 |
| $\tau_{00}$ | 1.89 $_{\text{Subject}}$ | 2.52 $_{\text{Subject}}$ | 1.92 $_{\text{Subject}}$ |
| $\tau_{11}$ | 3.29 $_{\text{Subject.GVI}}$ | 1.97 $_{\text{Subject.GVI}}$ | 1.92 $_{\text{Subject.GVI}}$ |
| $\rho_{01}$ | 0.02 $_{\text{Subject}}$ | -0.33 $_{\text{Subject}}$ | -0.16 $_{\text{Subject}}$ |
| ICC | 0.52 | 0.46 | 0.44 |
| N | 91 $_{\text{Subject}}$ | 91 $_{\text{Subject}}$ | 91 $_{\text{Subject}}$ |
| Observations | 533 | 533 | 533 |
| Marginal $R^2$ / Conditional $R^2$ | 0.197 / 0.613 | 0.280 / 0.608 | 0.292 / 0.601 |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

**Table 1.14.** *Generalised linear mixed effects model predicting System Challenge from GVI, Elite, and Identification and its interaction with Elite*

| | **Step 0** | | | **Step 1** | | | **Step 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* | *Odds Ratios* | *CI (95%)* | *p* | *Odds Ratios* | *CI (95%)* | *p* |
| Intercept | 1.97 | 0.31 – 12.51 | 0.471 | 0.35 | 0.05 – 2.58 | 0.306 | 0.87 | 0.11 – 7.19 | 0.899 |
| Group size | 0.92 | 0.81 – 1.06 | 0.252 | 0.92 | 0.80 – 1.06 | 0.239 | 0.91 | 0.79 – 1.05 | 0.196 |
| Risk preference | 1.47 | 1.23 – 1.77 | < **0.001** | 1.39 | 1.15 – 1.67 | **0.001** | 1.39 | 1.15 – 1.68 | **0.001** |
| GVI | 8.92 | 3.43 – 23.18 | < **0.001** | 6.35 | 2.54 – 15.87 | < **0.001** | 7.21 | 2.89 – 17.99 | < **0.001** |
| Legitimate elite | 1.90 | 0.90 – 4.03 | 0.095 | 1.79 | 0.81 – 3.93 | 0.148 | 0.32 | 0.08 – 1.29 | 0.109 |
| Illegitimate elite | 1.44 | 0.68 – 3.06 | 0.342 | 1.60 | 0.73 – 3.52 | 0.241 | 0.64 | 0.17 – 2.46 | 0.519 |
| Legit. x GVI | 0.41 | 0.13 – 1.23 | 0.112 | 0.41 | 0.13 – 1.26 | 0.120 | 0.33 | 0.10 – 1.05 | 0.061 |
| Illegit. x GVI | 0.44 | 0.14 – 1.38 | 0.160 | 0.37 | 0.11 – 1.16 | 0.088 | 0.32 | 0.10 – 1.03 | 0.057 |
| Identification | | | | 1.57 | 1.31 – 1.88 | < **0.001** | 1.25 | 0.98 – 1.59 | 0.070 |
| Legit. x Identification | | | | | | | 1.61 | 1.17 – 2.22 | **0.003** |
| Illegit. x Identification | | | | | | | 1.30 | 0.94 – 1.80 | 0.112 |

| **Random Effects** | | | |
|---|---|---|---|
| $\sigma^2$ | 3.29 | 3.29 | 3.29 |
| $\tau_{00}$ | 1.89 $_{\text{Subject}}$ | 2.52 $_{\text{Subject}}$ | 2.80 $_{\text{Subject}}$ |
| $\tau_{11}$ | 3.29 $_{\text{Subject.GVI}}$ | 1.97 $_{\text{Subject.GVI}}$ | 2.01 $_{\text{Subject.GVI}}$ |
| $\rho_{01}$ | 0.02 $_{\text{Subject}}$ | -0.33 $_{\text{Subject}}$ | -0.41 $_{\text{Subject}}$ |
| ICC | 0.52 | 0.46 | 0.46 |
| N | 91 $_{\text{Subject}}$ | 91 $_{\text{Subject}}$ | 91 $_{\text{Subject}}$ |
| Observations | 533 | 533 | 533 |
| Marginal $R^2$ / Conditional $R^2$ | 0.197 / 0.613 | 0.280 / 0.608 | 0.295 / 0.620 |

*Note*. † Reference group = Individual scenario; Treatment group = Group scenario. †† Reference group = No Elite

**Table 1.15.** *Generalised linear mixed effects model predicting System Challenge from GVI, Elite, and Identification and its interaction with GVI and Elite*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.90 | 0.11 – 7.63 | 0.924 |
| Group size | 0.91 | 0.79 – 1.04 | 0.155 |
| Risk preference | 1.43 | 1.20 – 1.71 | **< 0.001** |
| GVI† | 11.70 | 1.62 – 84.35 | **0.015** |
| Legitimate elite†† | 1.00 | 0.20 – 5.07 | 0.996 |
| Illegitimate elite†† | 1.97 | 0.40 – 9.68 | 0.406 |
| Identification | 1.32 | 0.96 – 1.81 | 0.089 |
| Legit. X GVI | 0.01 | 0.00 – 0.24 | **0.003** |
| Illegit. X GVI | 0.01 | 0.00 – 0.24 | **0.003** |
| GVI x Identification | 0.87 | 0.55 – 1.37 | 0.548 |
| Legit. X Identification | 1.17 | 0.79 – 1.74 | 0.442 |
| Illegit. x Identification | 0.92 | 0.61 – 1.40 | 0.702 |
| GVI x Legit. x Identification | 2.32 | 1.17 – 4.58 | **0.016** |
| GVI x Illegit. x Identification | 2.30 | 1.17 – 4.55 | **0.016** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 1.97 | | |
| $\tau_{11\ \text{Subject.GVI}}$ | 1.88 | | |

| | |
|---|---|
| $\rho_{01}$ Subject | -0.11 |
| ICC | 0.45 |
| N Subject | 91 |
| Observations | 533 |
| Marginal $R^2$ / Conditional $R^2$ | 0.338 / 0.636 |

*Note*. † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

**Table 1.16.** *Post-hoc contrasts of the final Identification model comparing the influence of Identification across GVI scenarios in each Elite condition*

| *Contrast* | *Elite* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| Individual – Group | No Elite | 0.139 | 0.232 | Inf | 0.601 | 0.548 |
| Individual – Group | Legit. Elite | -0.701 | 0.286 | Inf | -2.447 | **0.014** |
| Individual – Group | Illegit. Elite | -0.695 | 0.279 | Inf | -2.492 | **0.013** |

**Table 1.17.** *Post-hoc contrasts of the final Identification model comparing the influence of Identification across Elite conditions in each GVI scenario*

| *Contrast* | *GVI* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| No Elite – Legitimate Elite | Individual | -0.156 | 0.203 | Inf | -0.769 | 0.722 |
| No Elite – Illegitimate Elite | Individual | 0.081 | 0.211 | Inf | 0.382 | 0.923 |
| Legitimate Elite – Illegitimate Elite | Individual | 0.237 | 0.206 | Inf | 1.149 | 0.484 |
| No Elite – Legitimate Elite | Group | -0.996 | 0.284 | Inf | -3.507 | **0.001** |
| No Elite – Illegitimate Elite | Group | -0.753 | 0.274 | Inf | -2.752 | **0.016** |
| Legitimate Elite – Illegitimate Elite | Group | 0.242 | 0.306 | Inf | 0.792 | 0.708 |

**Table 1.18.** *Post-hoc estimations of the trend of Identification in each GVI x Elite combination*

| *GVI* | *Elite* | *Identification trend* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| Individual | No Elite | 0.276 | 0.162 | Inf | 1.700 | 0.089 |
| Group | No Elite | 0.136 | 0.171 | Inf | 0.797 | 0.426 |
| Individual | Legitimate Elite | 0.431 | 0.156 | Inf | 2.769 | **0.006** |
| Group | Legitimate Elite | 1.132 | 0.247 | Inf | 4.588 | < **0.001** |
| Individual | Illegitimate Elite | 0.195 | 0.164 | Inf | 1.189 | 0.234 |
| Group | Illegitimate Elite | 0.890 | 0.231 | Inf | 3.850 | < **0.001** |

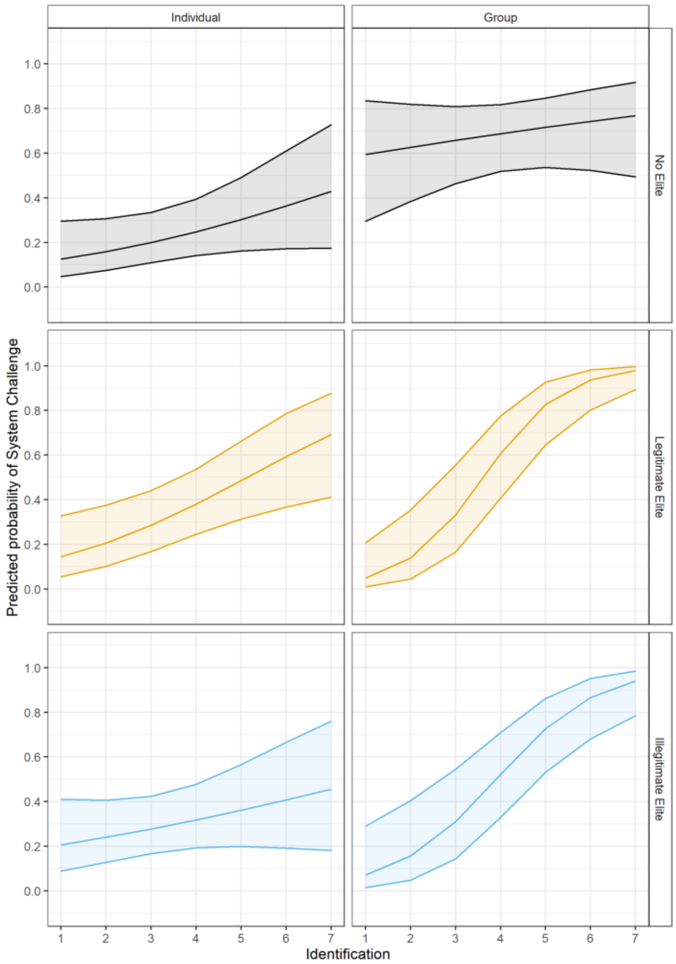


*Figure 1.8.* Predicted probability of Challenge based on Identification in each GVI scenario and Elite condition. Shaded areas represent the 95% confidence interval.

## Interim Summary

In Experiment 1, I investigated how contextual factors influenced decisions to support or challenge a social system that unfairly disadvantaged participants. I found that collective decision-making strongly increased decisions to challenge the system compared to individual decision-making. Making decisions collectively made feelings of efficacy more influential in guiding decisions, and this appeared to depend on how strongly identified with one another participants felt. Feelings of identification directly influenced decisions in groups when either a Legitimate or Illegitimate Elite was present and individually in the presence of the Legitimate Elite, providing evidence to suggest that when inter-group boundaries were salient, identity processes were more relevant in guiding behaviour. Contrary to expectations, overall there was little evidence to show that Legitimate Elites were considered more favourably than Illegitimate Elites. Systems with either Elite were considered less fair than systems with No Elite, but no difference in fairness ratings was observed between the two. In addition, the presence and type of Elite had no direct influence on the likelihood of challenging the system, suggesting that this factor was not particularly important for participants' decisions.

## Experiment 2

Experiment 1 considered situations in which participants were unable to communicate or use other coordinating information to guide decision-making. In Experiment 2, with a different sample of participants, I administered the same series of paradigms but with the addition of extra coordinating information. In group scenarios, participants were able to virtually communicate an intention and learn one another's intentions before making a final decision. In the individual scenarios, analogous non-social information about the likelihood of system overturn was provided by telling participants about virtual coin tosses that would determine the likelihood of system overturn. In this way, Experiment 2 sought to build on Experiment 1 by examining decision-making in the same scenarios when extra coordinating information was available.

### Method

**Participants**

One-hundred-and-ten participants took part in the experiment. Excluding the Elite and displaced Elites this amounted to 87 participants. Of these, 54 (62%) were female, 11 (13%) were male, 2 (2%) preferred not to say, one (1%) was 'other' and 19 (22%) were not defined (because the experimenter erroneously omitted the gender questionnaire for some participants). Of the 76 participants who provided the information, 71 (93%) were undergraduates and of these 64 (90%) were Psychology students.

As in Experiment 1, no formal procedure was adopted to determine sample size. Instead, because my intention was to compare decision-making in Experiment 2 to decision-making in Experiment 2, I aimed to recruit approximately the same number of participants in Experiment 2 as I did in Experiment 1. As considered in the Methods section of Experiment 1, this likely meant that the observed power for some effects was lower than 80%.

## Materials and procedure

Most of the materials and procedure of Experiment 2 were identical to those of Experiment 1. I will therefore only detail the elements that differed and refer the reader to the Methods section of Experiment 1 for the details that are the same.

As in Experiment 1, participants completed a total of six games, one for each combination of the 2 (GVI) x 3 (Elite) factor design. However, here the group games (always administered first) were split into two parts: an intention phase and a decision phase. The intention phase presented the scenario in the same way as in Experiment 1 but, instead of making a decision immediately, participants were asked to communicate an intention virtually by selecting 'I intend to Pay 50' or 'I intend to Pay 0'. It was made clear to participants that their expressed intentions were non-binding and they could change their mind in the decision phase if they wanted to. These intentions were carried forward to the decision phase, where the scenario was presented again, along with the intention information, which was presented as two sentences – one describing how many participants expressed an intention to pay 50, and one describing how many participants expressed an intention to pay 0. The participants' own intention was made clear in these sentences; depending on their intention, one sentence would read "You and [x] others expressed an intention to [pay 50/pay0]" whilst the other sentence read "[x] others expressed an intention to [pay 50/pay 0]". Participants completed the intentions phase of all group games before proceeding to the decision phase.

To create a non-social analogue of intentions for the individual scenarios I implemented a coin-tossing analogy. Participants were told that there were group size-minus-one virtual coins (not to be confused with the payment coins) in the game, a subset of which would be tossed, and that the likelihood of system overturn was determined by the number of coins landing on heads. Participants would receive this information on the decision screen: "There are [group size - 1] coins. [x] will be tossed. [group size – 1 – x] will not be tossed and will definitely be tails. Each coin that lands of heads increases the chance of overturn by approximately [85 / (group size – 1)] %."

The intentions communicated by participants for each Elite condition of the group games were carried forward to the individual paradigms and determined how many coins would be tossed (this is why the intention phase of the group games were administered first; note this was not explicitly stated to participants). In this way, the same "intentions" value was communicated in the individual and group scenarios for each Elite condition. For example, if 10 out of 17 participants expressed an intention to challenge in the Legitimate Elite condition of the group scenario, the decision phase of the group game would report that 10 participants expressed an intention to challenge, and the decision phase of the individual Legitimate Elite condition would state that 10 coins would be tossed and the remaining seven coins would not be tossed and would definitely be tails. In this way, an approximate analogue of intentions could be applied to the non-social individual scenarios. The analogy captured the possibility of a 'Pay 0' *intention* swapping to a 'Pay 50' *decision* and modelled this with 50% probability (i.e., a coin toss). It also assumed that no 'Pay 50' intention (not challenge the system) would convert to a 'Pay 0' decision (challenge the system). This of course could happen (and indeed one would hope that collective action would involve converting non-participants to active participants), but it was reasoned that since a rational actor would not express a non-cooperative intention that would undermine group efficacy and then subsequently choose to cooperate, this was a reasonable assumption for the individual scenarios. (However, there is a limitation in this assumption since it does not capture the possibility that non-cooperation intentions could convert to cooperation decisions if expectations of success increased upon learning intentions.)

As in Experiment 1, the order of administration of the games within the GVI factor was counter-balanced so that each Elite condition was presented first, second, and third an approximately equal number of times across sessions.

# Results

## Computed variables

In addition to the variables computed in Experiment 1, I also created the variable "Communicated information". This variable represented the number of other group members expressing an intention to challenge the system (in Group scenarios) or the number of coins that were to be tossed (in Individual scenarios) as a proportion of the total number of Group members (and thus total number of coins; see Methods) present. This provided a value ranging from 0 (no others intended to challenge/no coins would be tossed) to 1 (all others intended to challenge/all coins would be tossed). As for Number belief, this value was multiplied by 100 so that an increment of 1 corresponded to a proportion change of 1% (this was done so that odds ratios were more readily interpretable and so that the scale for Communicated information was identical to that of Number belief).

## System Challenge decisions

As in Experiment 1, I first tested what effect the manipulations of GVI and Elite had on decisions to Challenge the System with GLMM analysis. After controlling for Risk preference and Group size, GVI was a significant positive predictor of System Challenge (*log odds* = 1.52, *odds ratio (OR)* = 4.57, $z$ = 3.5, $p$ = < .001); the odds of a circle deciding to challenge the system were 4.57 times higher in the group scenario compared to the individual scenario. In contrast to Experiment 1, in this experiment where communication was possible participants were significantly less likely to challenge the system in the presence of either a Legitimate (*log odds* = -1.06, *OR* = 0.35, $z$ = -2.78, $p$ = 0.005) or Illegitimate (*log odds* = -1.06, *OR* = 0.35, $z$ = -2.78, $p$ = 0.005) Elite. All levels of the GVI x Elite interaction were non-significant. Note that Risk preference was also a significant predictor; more risk-seeking individuals were more likely to challenge the system. ***Table 1.19*** shows the estimates for all parameters.

**Table 1.19.** *Generalised linear mixed effects model predicting System Challenge*

| *Predictors* | *Odds Ratios* | *CI* | *p* |
|---|---|---|---|
| Intercept | 0.54 | 0.08 – 3.69 | 0.533 |
| Risk preference | 1.26 | 1.10 – 1.44 | **0.001** |
| Group size | 1.08 | 0.93 – 1.24 | 0.300 |
| GVI† | 4.57 | 1.95 – 10.70 | < **0.001** |
| Legitimate Elite†† | 0.35 | 0.16 – 0.73 | **0.005** |
| Illegitimate Elite†† | 0.35 | 0.16 – 0.73 | **0.005** |
| GVI x Legit. Elite | 2.89 | 0.97 – 8.63 | 0.057 |
| GVI x Illegit. Elite | 2.10 | 0.71 – 6.21 | 0.182 |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00}$ Subject | 1.15 | | |
| $\tau_{11}$ Subject.GVI | 2.05 | | |
| $\rho_{01}$ Subject | -0.07 | | |
| ICC | 0.39 | | |
| N Subject | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.235 / 0.530 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

Post-hoc tests showed that the odds of challenging the system were significantly higher in groups for all Elite conditions, and were significantly lower in the Legitimate and Illegitimate compared to no Elite conditions when playing individually (***Table 1.20*** & ***Table 1.21***, ***Figure 1.9***).

**Table 1.20.** *Post-hoc contrasts comparing System Challenge across Elite conditions in Individual and Group scenarios*

| *Contrast* | *GVI* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legitimate Elite | Individual | 1.06 | 0.38 | Inf | 2.78 | **0.02** |
| No Elite - Illegitimate Elite | Individual | 1.06 | 0.38 | Inf | 2.78 | **0.02** |
| Legitimate Elite - Illegitimate Elite | Individual | -0.00 | 0.39 | Inf | -0.00 | 1.00 |
| No Elite - Legitimate Elite | Group | -0.00 | 0.41 | Inf | -0.00 | 1.00 |
| No Elite - Illegitimate Elite | Group | 0.32 | 0.40 | Inf | 0.80 | 0.70 |
| Legitimate Elite - Illegitimate Elite | Group | 0.32 | 0.40 | Inf | 0.80 | 0.70 |

**Table 1.21.** *Post-hoc contrasts comparing System Challenge across Individual and Group scenarios in each Elite condition*

| *Contrast* | *Elite* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| Individual - Group | No Elite | -1.52 | 0.43 | Inf | -3.50 | **<0.01** |
| Individual - Group | Legitimate Elite | -2.58 | 0.47 | Inf | -5.53 | **<0.01** |
| Individual - Group | Illegitimate Elite | -2.26 | 0.45 | Inf | -4.98 | **<0.01** |

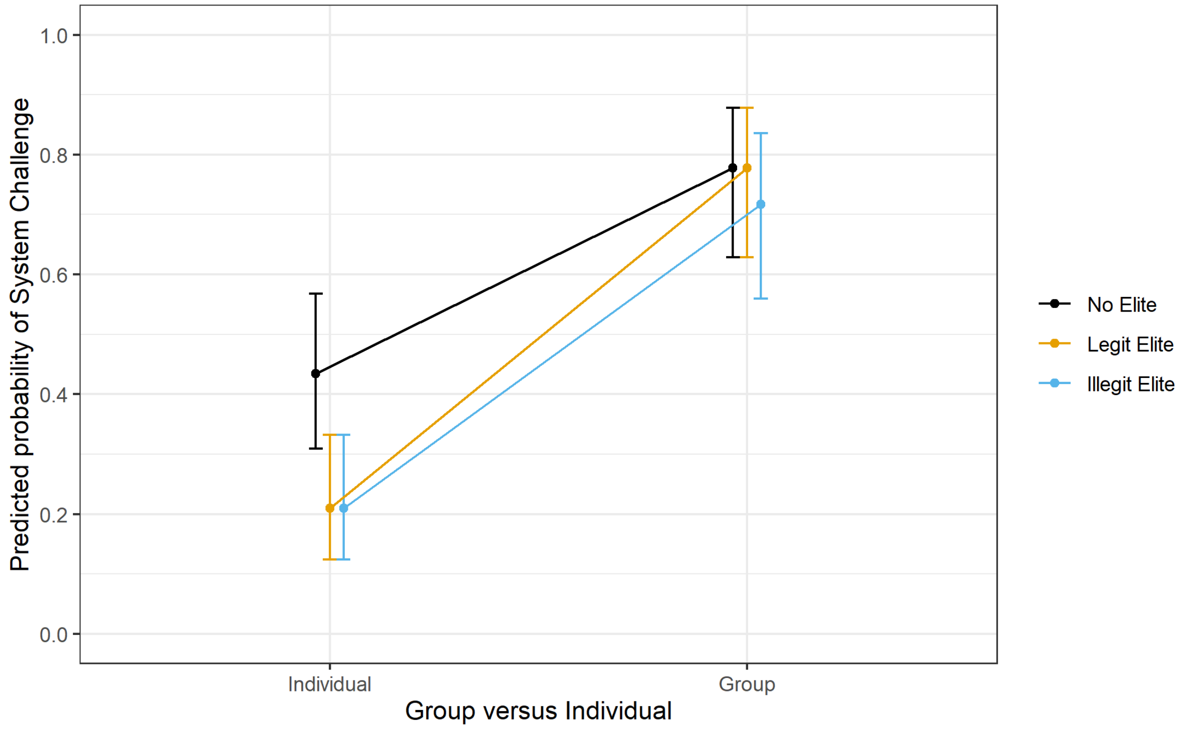


*Figure 1.9.* Predicted probability of a participant challenging the system as a function of GVI and Elite. Error bars represent the 95% confidence interval.

## Subjective experience of the System

I next checked what effect the experimental manipulations had on participants' experience of the system using linear mixed effects models predicting each of Fairness, Identification, and Number belief. For each analysis, I first specified a saturated model including GVI, Elite, and a GVI x Elite interaction term, allowing random slopes for GVI and Elite. I then removed parameters in order of complexity, starting with the random slopes, until the model successfully converged.

### *Fairness*

Consistent with Experiment 1, both Legitimate ($b$ = -0.45, $t$(348) = -2.99, $p$ = 0.003) and Illegitimate ($b$ = -0.68, $t$(348) = -4.52, $p$ < .001) Elite conditions negatively predicted Fairness, compared to the elite absent condition; when in a system with either the Legitimate or Illegitimate triangle, circles rated the system as less fair than when they were in a system with no triangle present. Contrary to the findings of Experiment 1, GVI was not found to have any influence on ratings of Fairness, and all combinations of the GVI x Elite interaction were non-significant (see ***Table 1.22***).

**Table 1.22.** *Linear mixed effects model predicting Fairness*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 3.43 | 3.04 – 3.81 | < **0.001** |
| Group size | 0.01 | -0.11 – 0.14 | 0.822 |
| GVI† | 0.01 | -0.33 – 0.35 | 0.947 |
| Legitimate elite†† | -0.45 | -0.74 – -0.15 | **0.003** |
| Illegitimate elite†† | -0.68 | -0.98 – -0.37 | < **0.001** |
| Legit. x GVI | 0.28 | -0.12 – 0.67 | 0.172 |
| Illegit. x GVI | 0.30 | -0.10 – 0.69 | 0.139 |
| **Random Effects** | | | |
| $\sigma^2$ | 0.89 | | |
| $\tau_{00}$ Subject | 2.54 | | |
| $\tau_{11}$ Subject.GVI | 0.87 | | |
| $\tau_{11}$ Subject.LegitimateElite | 0.18 | | |
| $\tau_{11}$ Subject.IllegitimateElite | 0.34 | | |
| $\rho_{01}$ | -0.13 | | |
| | -0.29 | | |
| | -0.43 | | |
| N Subject | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.067 / NA | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

Post-hoc contrasts revealed that in Individual scenarios both systems with a Legitimate elite and systems with an Illegitimate elite were rated as less fair than systems with no elite, and there was no difference in Fairness ratings between Legitimate and Illegitimate elite conditions. In Group scenarios the only significant difference was between the Illegitimate and No elite conditions (***Table 1.23; Figure 1.10***).

**Table 1.23.** *Post-hoc contrasts comparing Fairness across Elite conditions and Individual and Group scenarios*

| *contrast* | *GVI* | *Estimate* | *SE* | *df* | *t ratio* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legitimate Elite | Individual | 0.45 | 0.15 | 352.05 | 2.97 | **0.01** |
| No Elite - Illegitimate Elite | Individual | 0.68 | 0.15 | 352.05 | 4.50 | **0.00** |
| Legitimate Elite - Illegitimate Elite | Individual | 0.23 | 0.15 | 352.05 | 1.52 | 0.28 |
| No Elite - Legitimate Elite | Group | 0.17 | 0.15 | 352.05 | 1.14 | 0.49 |
| No Elite - Illegitimate Elite | Group | 0.38 | 0.15 | 352.05 | 2.52 | **0.03** |
| Legitimate Elite - Illegitimate Elite | Group | 0.21 | 0.15 | 352.05 | 1.37 | 0.36 |

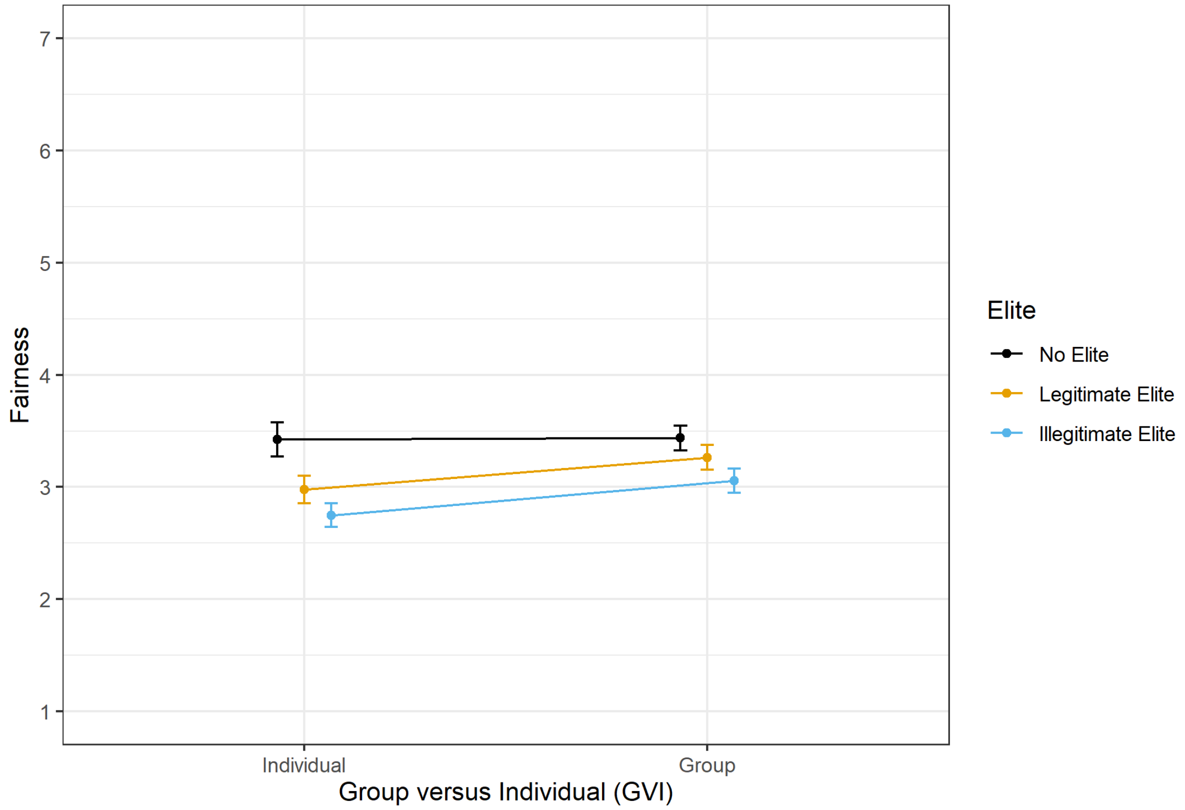


*Figure 1.10.* Average rated system Fairness as a function of Group versus Individual scenarios and Elite conditions. Error bars represent within-subject standard error.

### *Identification*

Consistent with Experiment 1, Identification was positively predicted by GVI ($b$ = 1.62, $t(186.62) = 7.08$, $p < .001$); circles identified more with one another when they played as a group compared to when they played alone (***Figure 1.11***). The Elite manipulation and the Elite x GVI interaction had no influence on Identification ratings (see ***Table 1.24***). No other terms in the model had an influence on Identification ratings.

**Table 1.24.** *Linear mixed effects model predicting Identification*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 3.13 | 2.76 – 3.50 | < **0.001** |
| Group size | -0.03 | -0.11 – 0.05 | 0.475 |
| GVI† | 1.62 | 1.17 – 2.07 | < **0.001** |
| Legitimate elite†† | 0.03 | -0.29 – 0.35 | 0.833 |
| Illegitimate elite†† | -0.15 | -0.47 – 0.17 | 0.359 |
| Legit. x GVI | 0.01 | -0.44 – 0.46 | 0.960 |
| Illegit. x GVI | 0.03 | -0.42 – 0.49 | 0.881 |
| **Random Effects** | | | |
| $\sigma^2$ | 1.16 | | |
| $\tau_{00}$ Subject | 1.92 | | |
| $\tau_{11}$ Subject.GVI | 2.24 | | |
| $\rho_{01}$ Subject | -0.77 | | |
| ICC | 0.56 | | |
| N Subject | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.207 / 0.648 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

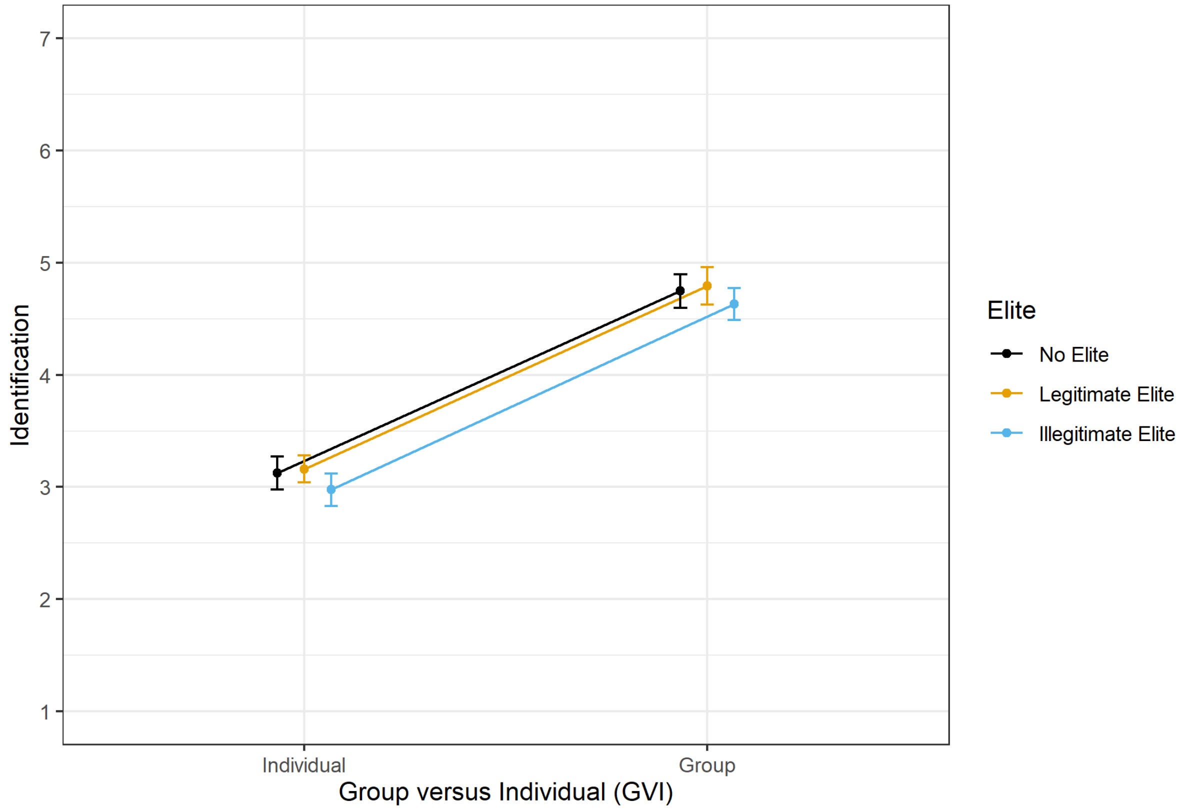


*Figure 1.11.* Average rated Identification as a function of Group versus Individual scenarios and Elite conditions. Error bars represent within-subject standard error.

### *Number belief*

Number belief was positively predicted by GVI ($b$ = 0.15, $t$(233.97) = 4.14, $p$ < .001) and negatively predicted by both Legitimate ($b$ = -0.09, $t$(348) = -3.27, $p$ = .001) and Illegitimate ($b$ = -0.11, $t$(348) = -3.81, $p$ < .001) Elite scenarios. The Illegitimate Elite x Group interaction term was significant ($b$ = 0.1, $t$(348) = 2.45, $p$ = .015) and indicated that the difference in Number belief between the individual and group scenarios was greatest when there was an Illegitimate Elite in the system (***Figure 1.12***). These findings therefore differed to those of Experiment 1, where Number belief was predicted only by the GVI manipulation. The group size term was also significant ($b$ = -0.02, $t$(87) = -3, $p$ = .003) indicating that in group sizes larger than the mean group size number belief was significantly lower (***Table 1.25***).

**Table 1.25.** *Linear mixed effects model predicting Number belief*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.46 | 0.42 – 0.51 | < **0.001** |
| Group size | -0.02 | -0.03 – -0.01 | **0.003** |
| GVI† | 0.15 | 0.08 – 0.21 | < **0.001** |
| Legitimate elite†† | -0.09 | -0.15 – -0.04 | **0.001** |
| Illegitimate elite†† | -0.11 | -0.16 – -0.05 | < **0.001** |
| Legit. x GVI | 0.06 | -0.01 – 0.14 | 0.111 |
| Illegit. x GVI | 0.10 | 0.02 – 0.18 | **0.014** |
| **Random Effects** | | | |
| $\sigma^2$ | 0.03 | | |
| $\tau_{00}$ Subject | 0.02 | | |
| $\tau_{11}$ Subject.GVI | 0.04 | | |
| $\rho_{01}$ Subject | -0.71 | | |
| ICC | 0.35 | | |
| N Subject | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.192 / 0.479 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

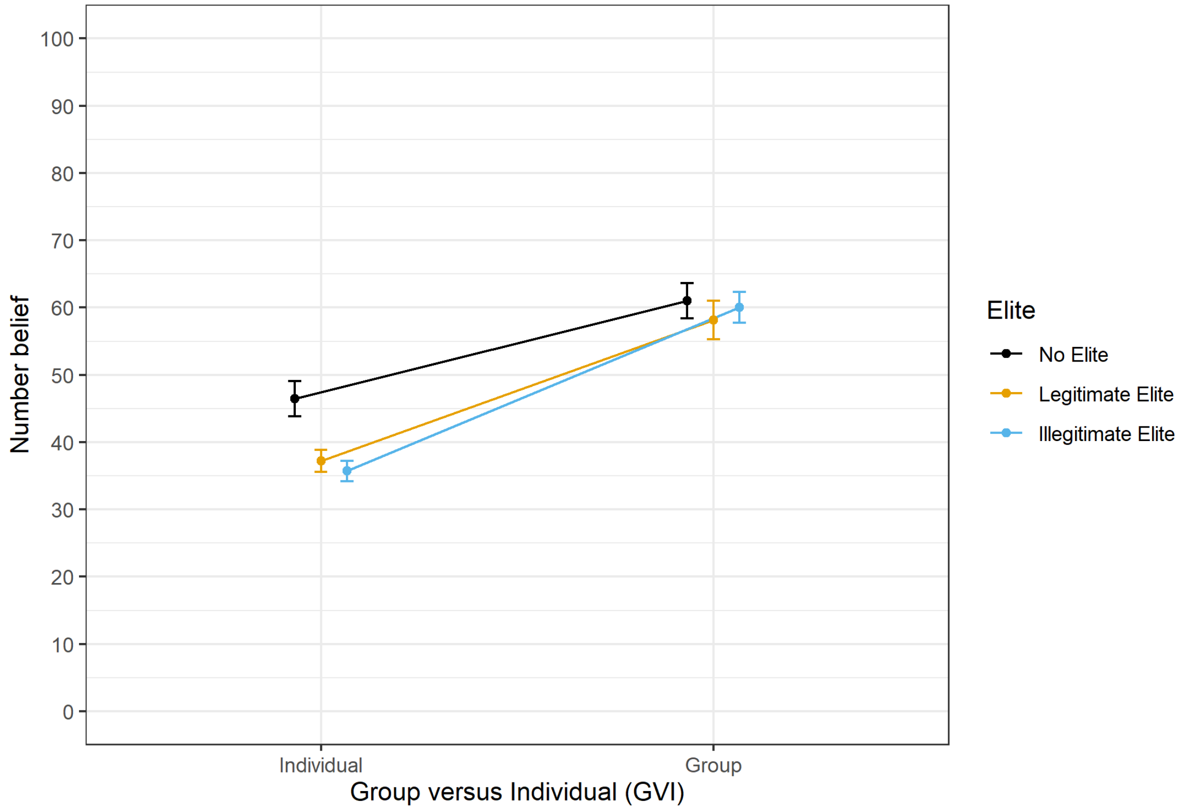


*Figure 1.12.* Average Number belief ratings as a function of Group versus Individual scenarios and Elite conditions. Error bars represent within-subject standard error.

Post-hoc tests showed that Number belief was significantly lower in Illegitimate and Legitimate Elite scenarios compared to the No Elite scenario when playing individually but not when playing as a group (***Table 1.26***), and that in all Elite conditions Number belief was higher when playing in a group compared to playing alone (***Table 1.27***).

**Table 1.26.** *Post-hoc contrasts comparing Number belief across Elite conditions in Individual and Group scenarios*

| *Contrast* | *GVI* | *Estimate* | *SE* | *df* | *t ratio* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legitimate Elite | Individual | 0.092 | 0.028 | 352.047 | 3.251 | **0.004** |
| No Elite - Illegitimate Elite | Individual | 0.107 | 0.028 | 352.047 | 3.786 | **0.001** |
| Legitimate Elite - Illegitimate Elite | Individual | 0.015 | 0.028 | 352.047 | 0.535 | 0.854 |

| | | | | | | |
|---|---|---|---|---|---|---|
| No Elite - Legitimate Elite | Group | 0.029 | 0.028 | 352.047 | 1.007 | 0.573 |
| No Elite - Illegitimate Elite | Group | 0.010 | 0.028 | 352.047 | 0.336 | 0.940 |
| Legitimate Elite - Illegitimate Elite | Group | -0.019 | 0.028 | 352.047 | -0.671 | 0.780 |

**Table 1.27.** *Post-hoc contrasts comparing Number belief across Individual and Group scenarios in each Elite condition*

| *Contrast* | *Elite* | *Estimate* | *SE* | *df* | *t ratio* | *p* |
|---|---|---|---|---|---|---|
| Individual - Group | No Elite | -0.146 | 0.035 | 236.686 | -4.117 | < **0.001** |
| Individual - Group | Legitimate Elite | -0.209 | 0.035 | 236.686 | -5.918 | < **0.001** |
| Individual - Group | Illegitimate Elite | -0.243 | 0.035 | 236.686 | -6.885 | < **0.001** |

**Influence of subjective experiences on System Challenge decisions**

Next, I explored what influence Number belief, Identification, and Fairness had on decisions to challenge the System by creating separate models containing each predictor and its interaction with GVI and Elite. This procedure produced models showing significant main effects of Number belief (***Table 1.28***), Identification (***Table 1.29***) and Fairness (***Table 1.30***), and significant interactions of GVI and Identification (***Table 1.29***) and of Elite and Fairness (***Table 1.30***).

***Number belief***

The Number belief model showed a significant main effect of Number belief. In contrast to Experiment 1, the GVI x Number belief interaction was non-significant (***Table 1.28***). Thus, contrary to Experiment 1, here feelings of efficacy were equally influential in both group and individual scenarios. This suggests that when coordinating information was available, there was no extra utility conferred to feelings of efficacy when making decisions collectively compared to alone.

**Table 1.28.** *Generalised linear mixed effects model predicting System Challenge from GVI, Elite, and Number belief*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.02 | 0.00 – 0.29 | **0.005** |
| Risk preference | 1.33 | 1.13 – 1.56 | **0.001** |
| Group size | 1.24 | 1.03 – 1.50 | **0.023** |
| GVI$^{\dagger}$ | 0.81 | 0.12 – 5.47 | 0.825 |
| Legitimate elite$^{\dagger\dagger}$ | 0.47 | 0.21 – 1.05 | 0.066 |
| Illegitimate elite$^{\dagger\dagger}$ | 0.50 | 0.22 – 1.11 | 0.088 |
| Number belief | 1.04 | 1.02 – 1.06 | **< 0.001** |
| Legit. x GVI | 2.71 | 0.75 – 9.83 | 0.129 |
| Illegit. x GVI | 1.24 | 0.35 – 4.35 | 0.738 |
| GVI x Number belief | 1.03 | 1.00 – 1.07 | 0.073 |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 1.53 | | |
| $\tau_{11\ \text{Subject.GVI}}$ | 5.52 | | |
| $\rho_{01\ \text{Subject}}$ | -0.11 | | |
| ICC | 0.55 | | |

| | |
|---|---|
| $N_{Subject}$ | 87 |
| Observations | 522 |
| Marginal $R^2$ / Conditional $R^2$ | 0.394 / 0.726 |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

### *Identification*

The Identification model contained a significant main effect of Identification and a significant GVI x Identification interaction (***Table 1.29***). Post-hoc tests showed that the trend of Identification was significant in both Individual ($trend_{Individual}$ = .30, $z$ = 2.69, $p$ = .007) and Group ($trend_{Group}$ = .97, $z$ = 5.12, $p$ < .001) scenarios, but was stronger in Group compared to Individual scenarios ($b$ = .68, $z$ = 3.07, $p$ = .002). ***Figure 1.13*** shows this relationship. This finding indicates that feelings of Identification were more relevant for decisions in the Group scenarios than in the Individual scenarios.  In contrast to Experiment 1, no three-way interaction between Identification, GVI and Elite was observed.

**Table 1.29.** *Generalised linear mixed effects model predicting System Challenge from GVI, Elite, and Identification*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.15 | 0.02 – 1.09 | 0.061 |
| Risk preference | 1.22 | 1.06 – 1.40 | **0.005** |
| Group size | 1.10 | 0.96 – 1.27 | 0.171 |
| GVI† | 0.15 | 0.02 – 0.96 | **0.046** |
| Legitimate elite†† | 0.34 | 0.16 – 0.72 | **0.005** |
| Illegitimate elite†† | 0.36 | 0.17 – 0.76 | **0.007** |
| Identification | 1.34 | 1.08 – 1.67 | **0.007** |
| Legit. x GVI | 2.82 | 0.87 – 9.19 | 0.085 |
| Illegit. x GVI | 2.03 | 0.63 – 6.51 | 0.233 |
| GVI x Identification | 1.97 | 1.28 – 3.03 | **0.002** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 1.01 | | |
| $\tau_{11\ \text{Subject.GVI}}$ | 2.48 | | |
| $\rho_{01\ \text{Subject}}$ | -0.03 | | |
| ICC | 0.40 | | |
| $N_{\text{Subject}}$ | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.362 / 0.618 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

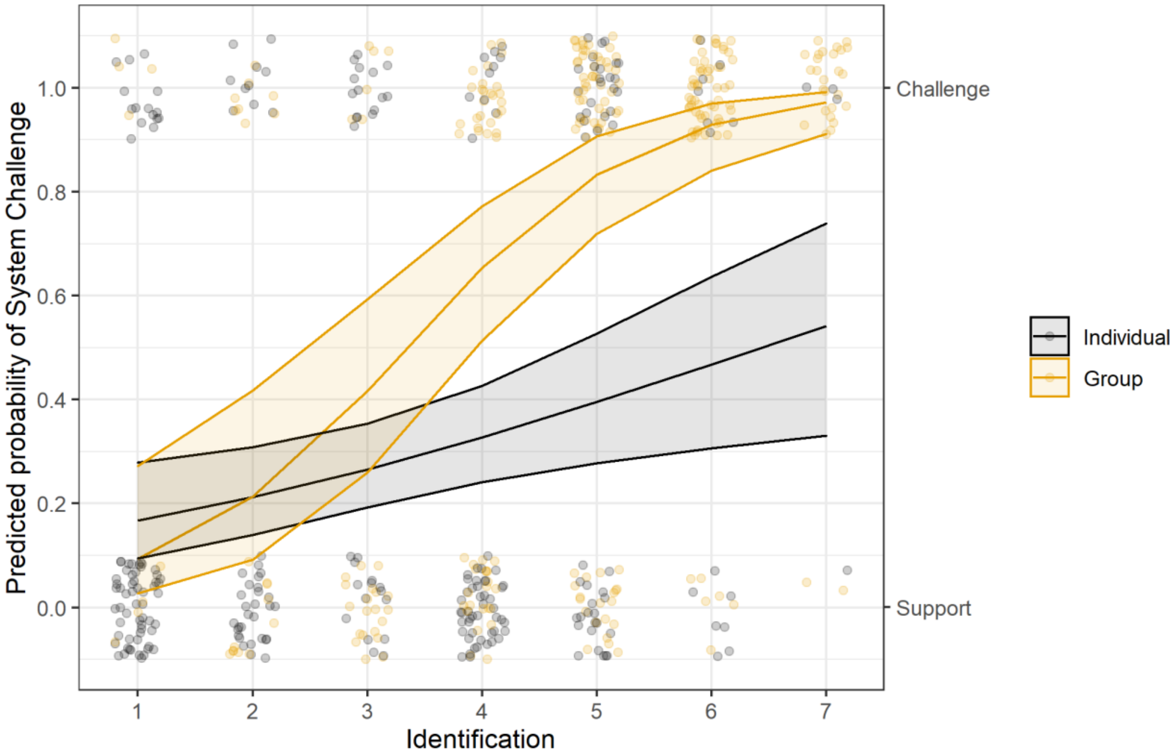


*Figure 1.13.* Predicted probability of System Challenge as a function of Identification in Group and Individual scenarios. Shaded areas represent the 95% confidence interval. Datapoints (jittered) correspond to the right vertical axis and represent raw decisions.

## ***Fairness***

The Fairness model contained a significant main effect of Fairness and a significant Elite x Fairness interaction (***Table 1.30***). Post-hoc tests showed that the Fairness trend was significant in the No Elite condition ($trend_{No\ Elite} = 0.38$, $z = 3.49$, $p < .001$), but non-significant in the Legitimate ($trend_{Legitimate} = .06$, $z = .50$, $p = .62$) and Illegitimate ($trend_{Illegitimate} = .03$, $z = .27$, $p = .79$) conditions. Note however that Tukey contrasts adjusting for multiple comparisons showed non-significant differences between the Legitimate and Illegitimate Elite conditions and the No Elite condition (***Table 1.31***), suggesting that this finding may lack statistical power. ***Figure 1.14*** plots these effects.

**Table 1.30.** *Generalised linear mixed effects model predicting System Challenge from GVI, Elite, and Fairness*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.24 | 0.04 – 1.55 | 0.134 |
| Risk preference | 1.24 | 1.09 – 1.41 | **0.001** |
| Group size | 1.03 | 0.91 – 1.17 | 0.624 |
| GVI† | 4.02 | 1.90 – 8.51 | **< 0.001** |
| Legitimate elite†† | 1.10 | 0.34 – 3.61 | 0.869 |
| Illegitimate elite†† | 1.20 | 0.37 – 3.87 | 0.760 |
| Fairness | 1.46 | 1.18 – 1.81 | **< 0.001** |
| Legit. x GVI | 2.60 | 0.91 – 7.47 | 0.075 |
| Illegit. x GVI | 2.00 | 0.70 – 5.69 | 0.193 |
| Legit. x Fairness | 0.72 | 0.55 – 0.96 | **0.025** |
| Illegit. x Fairness | 0.71 | 0.53 – 0.95 | **0.020** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 1.20 | | |
| ICC | 0.27 | | |
| $N_{\text{Subject}}$ | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.263 / 0.460 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

**Table 1.31.** *Post-hoc contrasts of the influence of Fairness between Elite conditions*

| *Contrast* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|
| No Elite - Legitimate Elite | 0.324 | 0.144 | Inf | 2.249 | 0.063 |
| No Elite - Illegitimate Elite | 0.348 | 0.149 | Inf | 2.330 | 0.052 |
| Legitimate Elite - Illegitimate Elite | 0.024 | 0.150 | Inf | 0.158 | 0.986 |

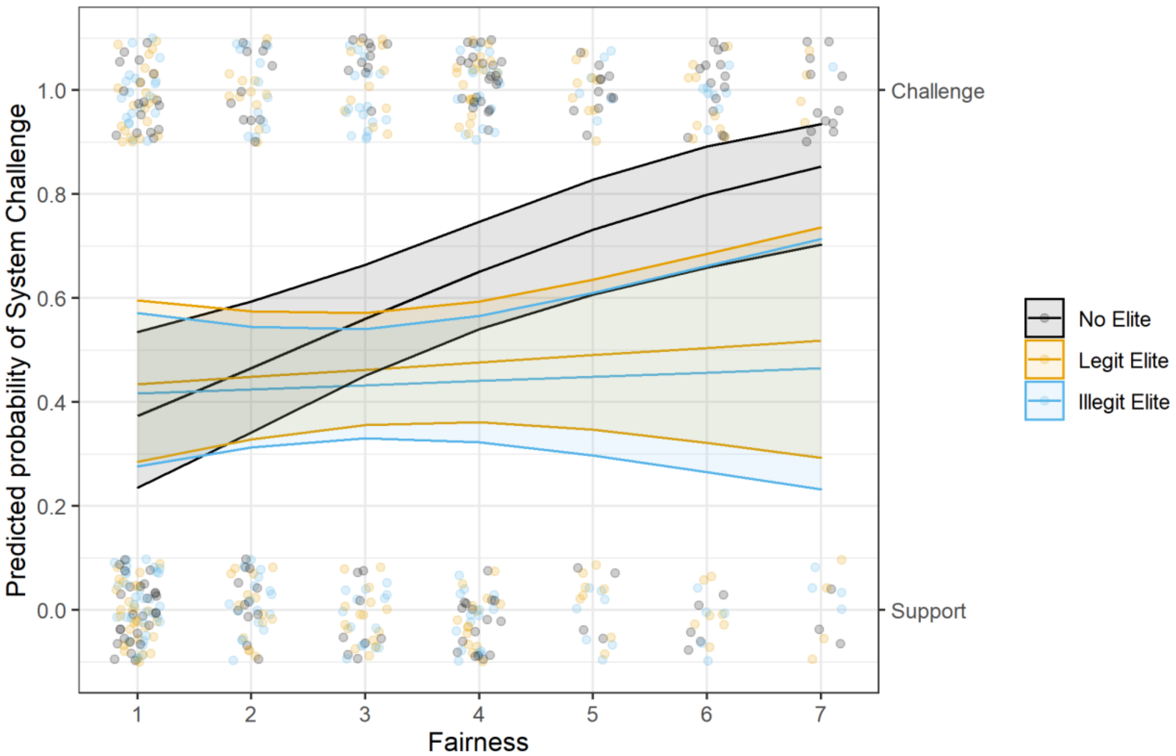


*Figure 1.14.* Predicted probability of System Challenge as a function of Fairness across Elite conditions. Shaded areas represent the 95% confidence interval. Datapoints (jittered) correspond to the right vertical axis and represent raw decisions.

**Information effects**

I next investigated how Communicated information (intentions in group scenarios, number of coins in individual scenarios) translated into beliefs about the likelihood of the system being overturned. To do this, I computed a new variable by dividing Number belief by

Communicated information. This new variable, called "Believed information", described the extent to which the communicated information was reflected in participants' Number belief. For example, a Believed information value of 50% would indicate that a participant believed that half the communicated intentions to challenge/coins tossed would actually challenge/land on heads. A Believed information value of 100% would indicate that the participant believed that all communicated intentions to challenge/coins tossed would challenge/land on heads. Believed information values above 100% would indicate that the participant believed that number of others actually challenging the system was *greater* than the number originally communicated (note that in individual scenarios this could never be the case).

I constructed a model predicting Believed information from GVI and Elite, controlling for Group size. Diagnostics for this model indicated that the assumption of homoscedasticity had been violated. I therefore used the '*robustlmm'* package to create a robust model. The robust model showed the same effects as the original. For the purposes of conventional significance-based interpretation, I report the original model (constructed using the *lmerTest* package). In this model, GVI and Group size were significant predictors of Believed intentions (***Table 1.32***); the number of others believed to have followed up on their intentions in the group scenarios was greater than the number of coins believed to land on heads in the individual scenarios, and intentions were believed less/fewer coins were believed to land on heads in larger compared to smaller groups. ***Figure 1.15*** shows the GVI effect. In the individual scenarios, the translation from number of coins tossed to number of heads believed to have occurred was approximately 52.56%. In the group scenarios, other circles' intentions were treated with greater odds; the translation from number of others intending to challenge to number of others believed to challenge was approximately 85.75%. This indicates that whereas coin tosses were treated as expected (i.e., with 50/50 chance of heads), in groups participants had greater confidence in other group members to follow through with their intentions.

**Table 1.32.** *Standard and Robust Generalised Linear Mixed Model predicting Believed Information from GVI and Elite conditions*

| | **Standard model** | | | **Robust model** | | |
|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI (95%)* | *p* | *Estimates* | *CI (95%)* | *p* |
| (Intercept) | 0.83 | 0.66 – 1.01 | < **0.001** | 0.75 | 0.61 – 0.88 | < **0.001** |
| Group size | -0.03 | -0.04 – -0.01 | < **0.001** | -0.02 | -0.03 – -0.01 | < **0.001** |
| GVI† | 0.35 | 0.26 – 0.43 | < **0.001** | 0.38 | 0.29 – 0.46 | < **0.001** |
| Legitimate Elite†† | 0.03 | -0.05 – 0.11 | 0.503 | 0.01 | -0.07 – 0.09 | 0.807 |
| Illegitimate Elite†† | -0.00 | -0.09 – 0.08 | 0.920 | -0.01 | -0.09 – 0.08 | 0.890 |
| GVI x Legit. Elite | -0.05 | -0.17 – 0.07 | 0.397 | -0.02 | -0.14 – 0.09 | 0.717 |
| GVI x Illegit. Elite | 0.01 | -0.11 – 0.12 | 0.925 | 0.01 | -0.10 – 0.13 | 0.820 |
| **Random Effects** | | | | | | |
| $\sigma^2$ | 0.08 | | | 0.07 | | |
| $\tau_{00}$ | 0.01 Subject | | | 0.00 Subject | | |
| ICC | 0.14 | | | 0.00 | | |
| N | 87 Subject | | | 87 Subject | | |
| Observations | 522 | | | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.260 / 0.361 | | | 0.340 / 0.340 | | |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

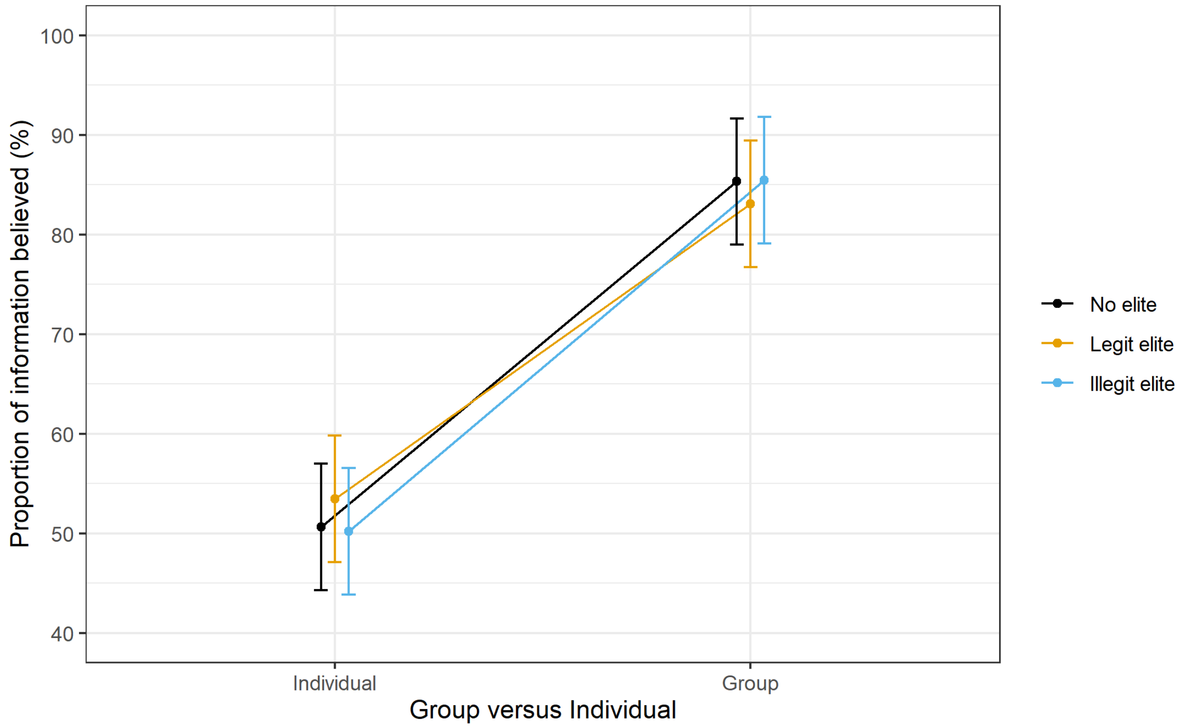


*Figure 1.15.* Believed information (i.e., proportion of information believed) as a function of Group versus Individual scenarios and Elite conditions. Error bars represent the 95% confidence interval.

Finally, I created a model to test the overall effect of the availability of information on the likelihood of challenging the system. To do so I combined the Experiment 1 and Experiment 2 datasets and created a new binary categorical variable describing whether information was available (Experiment 2 dataset) or not (Experiment 1 dataset). I named this variable "Information" (it should be noted that this variable is distinct from the "Communicated information" variable in the preceding analyses). Using this combined dataset, I specified a model predicting System Challenge from the three-way interaction of GVI, Elite, and Information, in addition to the covariates group size and risk preference. The resulting model is shown in ***Table 1.33***. The main effect of Information was significant. Overall, the presence of Intention and coin toss information increased the likelihood of challenging the system by two-and-a-half times compared to when that Information was not available. In addition, the three-way interaction was significant, and post-hocs showed that the presence versus absence of Information had a particularly strong effect in the Individual No Elite scenario (***Table 1.34; Figure 1.16***).

**Table 1.33.** *Generalised linear mixed effects model predicting System Challenge from GVI, Elite, and Information*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.73 | 0.21 – 2.55 | 0.627 |
| Group size | 0.98 | 0.90 – 1.07 | 0.704 |
| Risk preference | 1.29 | 1.17 – 1.43 | **< 0.001** |
| GVI† | 6.32 | 3.02 – 13.25 | **< 0.001** |
| Legitimate Elite†† | 1.81 | 0.88 – 3.72 | 0.104 |
| Illegitimate Elite†† | 1.40 | 0.68 – 2.89 | 0.359 |
| Information | 2.52 | 1.12 – 5.69 | **0.026** |
| GVI x Legit. Elite | 0.45 | 0.17 – 1.24 | 0.124 |
| GVI x Illegit. Elite | 0.52 | 0.19 – 1.47 | 0.219 |
| GVI x Information | 0.62 | 0.22 – 1.72 | 0.357 |
| Legit. Elite x Information | 0.18 | 0.07 – 0.52 | **0.001** |
| Illegit. Elite x Information | 0.24 | 0.08 – 0.67 | **0.007** |
| GVI x Legit. Elite x Information | 6.55 | 1.53 – 28.07 | **0.011** |
| GVI x Illegit. Elite x Information | 4.31 | 1.00 – 18.68 | 0.051 |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 1.57 | | |

| | |
|---|---|
| ICC | 0.32 |
| N Subject | 178 |
| Observations | 1055 |
| Marginal $R^2$ / Conditional $R^2$ | 0.204 / 0.461 |

*Note.* † Reference group = Individual scenario; Treatment group = Group scenario

†† Reference group = No Elite

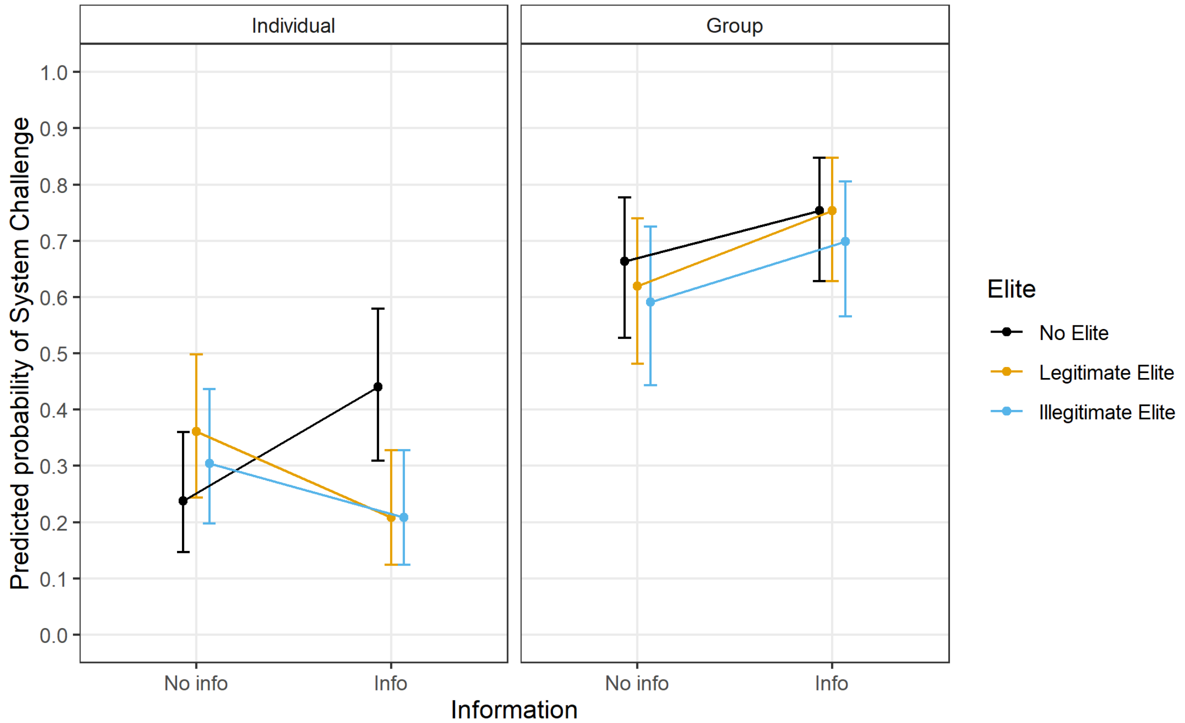


*Figure 1.16.* Predicted probability of a participant challenging the system as a function of GVI, Elite, and Information conditions. The only significant effect of Information was observed between the Individual No Elite conditions. Error bars represent the 95% confidence interval.

**Table 1.34.** *Post-hoc contrasts of the effect of Information on System Challenge in GVI and Elite conditions*

| *Contrast* | *GVI* | *Elite* | *Estimate (log odds)* | *SE* | *df* | *z* | *p* |
|---|---|---|---|---|---|---|---|
| No info - Info | Individual | No Elite | -0.924 | 0.416 | Inf | -2.223 | **0.026** |
| No info - Info | Group | No Elite | -0.441 | 0.418 | Inf | -1.056 | 0.291 |
| No info - Info | Individual | Legitimate Elite | 0.763 | 0.423 | Inf | 1.803 | 0.071 |
| No info - Info | Group | Legitimate Elite | -0.634 | 0.416 | Inf | -1.523 | 0.128 |
| No info - Info | Individual | Illegitimate Elite | 0.506 | 0.426 | Inf | 1.188 | 0.235 |
| No info - Info | Group | Illegitimate Elite | -0.473 | 0.425 | Inf | -1.113 | 0.266 |

## Interim Summary

In Experiment 2 I sought to investigate how the contextual factors of group versus individual decision-making and the presence and type of Elite influenced decision-making when coordinating information was available. Overall, I found that the addition of Information had a positive influence on the likelihood of challenging the system, and this effect was particularly strong when decisions were made individually with No Elite in the system.

The similarities and differences of the results of Experiment 2 compared to Experiment 1 offer insight into how the availability of extra coordinating Information influenced the process by which decisions were made, and provided an opportunity to identify effects that did and did not replicate across samples. Firstly, I found, consistent with Experiment 1, that collective decision-making markedly increased the likelihood of challenging the system over individual decision-making. With extra information, participants were also less likely to challenge the system in the individual scenarios in the presence of an Elite compared to when

No Elite was present, which was not the case in the no information conditions. Secondly, in contrast to Experiment 1, feelings of efficacy were equally influential in group and individual scenarios. The fact that the effect in Experiment 1 was moderated by identification suggests that when coordinating information was lacking, participants drew on feelings of identification to guide decision-making but did not need to do so in Experiment 2 because they had more concrete (albeit ambiguous) information available. Thirdly, I found evidence supporting the notion that feelings of Identification were particularly relevant for guiding decisions when group boundaries were more salient. Consistent with Experiment 1, Identification was more influential for collective decisions than individual ones. In contrast, however, this effect was not sensitive to the presence or type of Elite.

An important consideration was the extent to which the extra information was believed. I found that the information was considerably more believed in the group scenarios compared to the individual scenarios. This highlights that participants were relatively confident that others would follow up on their intentions, whereas in individual scenarios the coin toss information was believed approximately at the 50% rate that would be expected from a coin toss.

## Discussion

Across two experiments I investigated how the framing of an unfair social system influenced individuals' subjective experience and their propensity to challenge the system at personal cost. I also examined how extra coordinating information altered experiences and decisions. Overall, unfair systems were more likely to be challenged when decisions were made collectively, with each individual's decision influencing the outcome for the entire group, compared to when decisions were made individually and outcomes were determined by chance. Efficacy beliefs (i.e., expectations relating to the probability of success) and social identification were sensitive to changes in context and were important in guiding decisions. The availability of extra information (intentions in the group scenarios and coin tosses in the individual scenarios) increased the likelihood of challenging, with its greatest influence evident in the Individual No Elite condition. The availability of information also had an effect on the predictive power of efficacy. When no information was available, efficacy beliefs were more influential for decisions in group scenarios when identification was high. When information was available, efficacy beliefs were not more influential in group scenarios.

The findings provide empirical insight to the growing literature on the core motivations underlying collective action. In particular, the findings shed light on the intricacies of the relations between feelings of social identification and efficacy and how these influence decisions to support or challenge social arrangements. Below, I first consider the direct influence of identification on decisions to challenge. I then go on to consider the moderating role of identification on the influence of efficacy. Finally, I consider how the availability of coordinating information affected these processes.

### The direct influence of identification

Consistent with prior research (e.g., Thomas et al., 2012; van Zomeren et al., 2008), I found evidence that feelings of identification were relevant to decisions both directly and indirectly. Critically, however, the influence of identification was not universal but rather context-dependent. Under conditions of limited information, in situations with no class difference identification had no influence on decisions, either alone or in groups. In contrast, when there was a “Legitimate” class-difference, greater identification was associated with a greater likelihood of challenging the system both alone and in groups, and this effect was stronger for group decision-making compared to individual decision-making. Interestingly, in Experiment 2 – where information was available (intentions in the group scenarios and number of coins to be tossed in the individual scenarios) – the qualifying effect of the status differential disappeared. That is, when information was available the presence and Legitimacy of the Elite was inconsequential to the influence that feelings of shared identity had on decisions. Instead, the influence of identification was sensitive to whether decisions were made as a group or individually: both individually and in groups, feelings of shared identity increased the likelihood of challenging, but the effect was stronger when making decisions as part of a group.

The finding that the relevance of identification varied with context challenges the notion that it is the central causal centrepiece of collective action (van Zomeren et al., 2008). Instead, the present evidence highlights the dynamic nature of identity processes and corroborates the importance of emergent identities in dealing with novel social issues (Thomas et al., 2009, 2012, 2016). For example, Thomas et al. (2016) directly compared the ability of the Encapsulation Model of Social Identity in Collective Action (EMSICA) – which posits a reflexive and dynamic role of identity in collective action – to van Zomeren et al.’s (2008) Social Identity Model of Collective Action (SIMCA) – which argues that social identity precedes other subjective experiences and action – to explain action-relevant beliefs, emotions, and intentions after group interaction and found that in the context of newly-formed groups, the encapsulating role of social identity was a more appropriate explanation. They therefore suggested that in cases of newly forming groups the formation of social identities may be influenced by group interaction and subjective feelings arising from them, such as feelings of efficacy and anger. Akfirat et al. (2020) further showed in a meta-analysis of 40 studies that the influence of identification on collective action participation is stronger in emergent groups than pre-existing groups, supporting the notion that identity processes are particularly influential in coordinating behaviour in newly formed groups. My findings add to this literature by demonstrating that the guiding influence of identification in novel social dilemmas is sensitive to changes in context.

Specifically, I have presented evidence to suggest that feelings of identity are particularly important when there is a clear distinction between one's own group and an outgroup. An explanation for why this was the case can be derived from research on the emergence of "politicised" collective identity. A politicised collective identity "can be understood as a form of collective identity that underlies group members' explicit motivations to engage" in a power struggle (Simon & Klandermans, 2001, p. 323); simply put, a politicised collective identity is a collective identity that involves intentions to participate in a contest of power. As alluded to in the Introduction, collective identity politicises when (i) there is a shared feeling of group-based deprivation which is (ii) attributable to an adversary/adversarial group, and (iii) which is acted upon so that members of broader society are forced to take a stance on the issue. These processes also align with the primary and secondary appraisals suggested to guide coping (Lazarus, 1991; Lazarus & Folkman, 1984). According to Lazarus (1991), primary appraisal involves identifying the relevance of disadvantage for the group, and secondary appraisal involves attributing blame for disadvantage. Though the design of the present study does not perfectly align with the societal configuration of Simon and

Klandermans’ (2001) point (iii) above, there is consistency with both Simon and Klandermans (2001) and Lazarus (1991) with respect to the shared feelings of group-deprivation and its causal attribution to a problematic system or group; across both experiments reported here, identification was a stronger predictor of decisions to challenge in groups compared to individually (i.e., under conditions of group-based deprivation), and in Experiment 1 the presence of an Elite (especially the Legitimate Elite) also rendered the influence of identification stronger (i.e., the deprivation was attributed to someone). Furthermore, as Klandermans (2014) outlines, politicisation of identity is not all or nothing but instead involves degrees of identity transformation as a power struggle unfolds. The endpoint of this process is “full politicisation”, but the stages of politicisation prior to this are nonetheless consequential. The influence of identity in the present study might thus be considered a demonstration of emergent politicised collective identity.

This line of reasoning must by necessity address why – in Experiment 1 – the presence of the “Legitimate” Elite was so galvanising. It was anticipated that systems with a Legitimate Elite – who was meritocratically-designated – would be perceived as fairer than systems whose advantaged participant was “Illegitimate” and designated by chance, and that decisions to challenge the system would reflect this. This expectation was based on the assumption that a merit-worthy Elite would be perceived as more legitimate than a non-merit-worthy one, and thus the system over which they presided would be interpreted as relatively more stable (Tajfel & Turner, 1979). Contrary to expectations, the systems involving the so-called Legitimate Elite were rated equally as unfair as those involving the Illegitimate Elite and inspired greater identification-derived action. The legitimacy manipulation therefore did not elicit the intended responses. It is possible that disadvantaged participants perceived themselves to be more socially distant from the Legitimate Elite relative to the Illegitimate Elite because, whereas the Illegitimate Elite could have been any participant in the study, the Legitimate Elite was distinct in being more mathematically adept than the rest. As a result, it is possible that the Legitimate Elite position was paradoxically perceived as illegitimate because it was defined by excellence in addition tasks, whilst the Illegitimate position was perceived as more democratic insofar as all participants stood an equal chance of adopting it. The finding that the contravention of equality rules is likely to be more important in mobilising left-leaning individuals than the contravention of merit-based rules (Mikołajczak & Becker, 2019), combined with the likelihood that the majority of the present study's sample was left-leaning, lends weight to the idea that the Legitimate Elite was perceived as illegitimate. If this was indeed the case, the social identity explanation holds true; the less legitimately perceived system was subject to the greatest rates of challenge because its lower legitimacy rendered it less stable.

An alternative explanation that does not assume the non-legitimacy of the Legitimate Elite invokes self-categorisation theory (Turner et al., 1987). According to this approach, identities are usually nested within superordinate identity categories and become relevant at different levels of abstraction. It is possible that the participant in the Illegitimate Elite position, since they were chosen randomly and could have been any one of the participants, was considered at the superordinate level of "participant", which subsumed the categories of "Elite" and "non-Elite". As a result, when making decisions in systems with the Illegitimate Elite, non-Elites may simply have perceived the Illegitimate Elite as "one of us". In contrast, since the Legitimate Elite was explicitly contextualised in the experiment as the highest scorer in the maths task (and not "any other participant"), non-Elites may have felt more alienated from them. As a result, in scenarios involving the Legitimate Elite the lower-level identity categories of "Elite" and "Non-Elite" may have been more salient, and comparisons of these identities could have encouraged greater identity-driven action. By this interpretation, the greater identity-driven mobilisation against the Legitimate Elite is explained by the level of abstraction of group comparisons.

**The moderating role of identification on efficacy beliefs**

In addition to the direct influence of identification, in Experiment 1 I also found that identification moderated the extent to which efficacy beliefs encouraged decisions to challenge the system. Specifically, I found that the greater relevance of efficacy in groups compared to individual decision settings could be attributed to social identity; when identification was low, feelings of efficacy were no different in predicting decisions to challenge in group compared to individual scenarios. However, when identification was high, feelings of efficacy were more influential when decisions were made in groups compared to when they were made alone. The significance of this finding lies in the fact that feelings of efficacy directly mapped onto expected value; greater feelings of efficacy represented beliefs in higher probabilities of success. Thus, in both individual and group scenarios, a number belief (i.e., efficacy) rating of 50 represented the same believed probability in success and therefore should have been valued equally. The fact that efficacy beliefs were acted upon as if they had greater value in the groups compared to individually when identification was high suggests that strong feelings of identification added extra utility to the expected value of challenging the system in groups. At higher levels of identification, a lower threshold of efficacy was required to elicit an action response; in a sense, feelings of efficacy were bolstered by greater identification, with the result that at higher levels of identification a lower level of efficacy elicited the same response. Put simply, higher identification appeared to enhance subjective feelings of group efficacy.

This finding is at odds with the predictions of the dynamic dual pathway model (van Zomeren et al., 2012), which predicts that efficacy beliefs should be more predictive of collective action when identification is lower. It is suggested that efficacy beliefs are a means by which low identifiers can determine whether engaging in action is worthwhile (Doosje et al., 2002; Kelly & Breinlinger, 1995). This makes intuitive sense insofar as lower identifiers may be less motivated by collective goals and therefore action will be guided – if at all – by other factors such as instrumental expectations of success. There is indeed empirical data to suggest that efficacy beliefs are more predictive of collective action when group identity is less relevant (van Zomeren et al., 2008). This line of reasoning suggests that the predictive influences of identity and efficacy are inversely related to each other.

It is notable then that the findings reported in the present study follow an exactly opposite pattern. The evidence I present suggests that group identification has an enhancing effect on the relation between efficacy beliefs and mobilisation, and there is also research to support this notion. In a conceptual replication of the Stanford Prison Experiment (Haney et al., 1973), Haslam and Reicher (2006) found that increased identification amongst participants in the prisoner role led to greater social support and more effective coping with the stressors of the situation. Outten et al. (2009) showed that greater identification as a Black American was associated with greater problem-focused (i.e., efficacy-related) group-based coping strategies, which in turn improved well-being outcomes. Constantine et al. (2002) found that stronger identification amongst African Americans was associated with greater use of collective coping strategies. Tausch et al. (2011) investigated the dual pathway model in three distinct politicised identities (students protesting against tuition fees in Germany, Indian Muslims' support for action addressing ingroup disadvantage, and British Muslims' appraisals of British foreign policy in Muslim countries) and found that efficacy was positively related to normative action (e.g., peaceful protest, petitions) but negatively related to non-normative action (e.g., violent protest). In addition, both SIMCA and EMSICA suggest an enhancing role of identification on efficacy (Thomas et al., 2009, 2012; van Zomeren et al., 2008). In general, these findings, including those of the present study, are consistent with the social identity approach to coping, which emphasises that relevant identities in a given situation influence how that situation is appraised, including expectations of efficacy (e.g., Haslam et al., 2004, 2005). It therefore appears that there is evidence in support of the notion that greater identification can render feelings of efficacy *more* influential in eliciting action.

The opposing lines of evidence may be explained by the degree to which the relevant identities are politicised. For example, it is significant to observe that the action of identification-enhanced efficacy-related coping in the Outten et al. (2009) study was specifically in the *intergroup* domain (as opposed to *intra*group coping). Though the authors did not provide any strong theoretical explanation for this based on the evidence, the finding resembles the above-discussed result showing that the direct influence of identification on decisions to challenge was strongest in the scenario where the ingroup/outgroup distinction was most salient (i.e., when participants decided in groups in the presence of the Legitimate Elite). In Study 1 of van Zomeren et al. (2008), though participants were students taking part in a real-life demonstration against financial cuts to education, the measure of identity obtained was *chronic* identification as a student, rather than *emergent* politicised identity as a protestor. It is possible that this created a disconnect between the established measured identity and the emergent politicised identity that was relevant to the immediate situation. As a result, identification as a student may have confounded identification as an activist. For example, highly identified students may have felt more dependent on the higher education system and so their efficacy beliefs about being able to challenge the financial cuts may have been less powerful in guiding action. In contrast, those who identified less as a student may have felt less constrained as a member of the system, making their perceived efficacy more consequential. Indeed, it is even conceivable that those *least* identified as students may have been the *most* identified as activists in that particular context. Similarly, the moderating role of gender identification on how much feelings of efficacy predicted women's participation in collective action was based on identification as a woman and not identification as an activist (Kelly & Breinlinger, 1995). This notwithstanding, efficacy was equally if not more influential in predicting *intentions* to participate amongst strong identifiers compared to weak identifiers, and since intentions are found to be strongly predictive of behaviour (Ajzen, 1991) this finding casts doubt on the notion that stronger effects of efficacy are always associated with weaker identification. There are thus some grounds to doubt the assertion that low identification leads to a greater influence of efficacy. Previous findings, as well as my own, may be understood to indicate that *politicised* identity enhances the power of efficacy beliefs to mobilise individuals to action.

**Decisions with extra coordinating information**

An extremely consistent finding in the social dilemma literature is that communication increases cooperative behaviour. Meta-analytic research has estimated that communication can increase cooperation in social dilemmas by approximately 40 percent on average compared to when communication is not possible (Sally, 1995), though several studies report even larger effects (e.g., Bornstein, 1992; Frohlich & Oppenheimer, 1998; Hackett et al., 1994). Consistent with prior research, my results demonstrate that making communication possible (allowing participants to communicate intentions in group conditions and receiving analogous coin-tossing information in the individual conditions) increased the likelihood of challenging the system. To recap, comparing challenge rates between Experiment 1 (no information) and Experiment 2 (extra coordinating information) showed that the availability of extra information led to an increase in the likelihood of challenging the system. This was especially the case when participants made decisions individually with No Elite present. In addition to the direct effect of information, there were also some notable differences between Experiments 1 and 2 in how subjective experiences predicted decisions. Specifically, with extra coordinating information, the influence of efficacy beliefs was not different between individual and group scenarios. As has been considered above, in Experiment 1, where no information was available, the influence of efficacy beliefs was greatest when decisions were made in groups and identification was high. This suggests that the additional information provided in Experiment 2 allowed participants to feel efficacious regardless of the extent to which they identified with one another. This, by extension, implies that feelings of identification acted as a means of social coordination in the absence of explicit communication.

Explanations for the communication effect observed in the literature include the notion that communication allows group members to identify with others (i.e., reduce social distance), appraise the character of others to estimate the likelihood that they will cooperate, and commit to a promise of cooperation (Bicchieri & Lev-On, 2007; Dawes, 1980; He et al., 2017). He et al. (2017) presented evidence to suggest that the most potent of these in increasing cooperation is what they termed “type detection”; the appraisal of others’ cooperativeness. They also found that the commitment aspect was also important, whilst social distance was apparently not influential. The notion that social distance does not play a primary role in the communication effect has been echoed by others (Bicchieri & Lev-On, 2007; Bouas & Komorita, 1996). Though Bouas and Komorita (1996) did observe that communication increased feelings of group identification, they concluded that this identification alone did not elicit greater cooperation. Instead, they suggested from their evidence that arriving at group consensus (i.e., the commitment factor) was of primary importance. This research converges on the idea that the reason why communication improves cooperation is related to the greater ability to coordinate as a group.

The findings I have presented are consistent with this notion that communication serves as a coordination tool. In Experiment 1, where information was limited, participants appeared to used group membership and feelings of identification to gauge the relation between feelings of efficacy and decisions. This was not necessary for participants in Experiment 2, where extra information provided the opportunity to coordinate without the need to make identity-derived inferences. Thus, though identification was more influential in the group scenarios in Experiment 2, it had no bearing on how feelings of efficacy translated into action.

The finding that the effect of information was particularly strong when making decisions individually in the No Elite condition might suggest that the availability of information was complicated by the structure of the decision scenarios. In effect, the Individual No Elite scenarios were the least complex of the experiment; decisions only affected oneself and there was no intergroup structure. In terms of appraisal processes (e.g., Lazarus, 1991), there was therefore minimum identification as a member of a disadvantaged group (primary appraisal) and minimum attribution of blame (secondary appraisal). Thus, in this scenario, it is possible that the addition of information did nothing but provide a tool with which to make a more informed probability-based decision. That is, in the absence of more complex social information, the decision amounted solely to determining the expected value of the decision to challenge (i.e., the value of the reward multiplied by the probability of receiving it), comparing it to the expected value of not challenging, and deciding which option was preferred.

The role of information in supporting social movements, and the finding that this can happen without the need for a salient sense of group identity, may have important implications for, and indeed help to explain, modern social media activism. In the 21st Century a large amount of social interaction occurs virtually, and the virtual social world provides a unique platform to connect with unknown others who share the same outlook on an issue. Social media activism has been conceptualised as a “digital form of consensus mobilisation” (Foster, Tassone, & Matheson, 2021, pp. 742), alluding to a necessary precursory process of raising public knowledge and appreciation of an issue in need of being addressed (Klandermans & Oegema, 1987). In support of this notion, in an experimental setting, tweeting on Twitter following exposure to sexism was found to strengthen social identity and collective action intentions amongst women (Foster et al., 2021). In this context, identification as woman featured as a consequence of activism rather than as a necessary prerequisite. This is consistent with the findings of the present study, where the effect of communication served as a coordinating tool to encourage action, independent of feelings of identification. The implication of this finding, more generally, is that, in the context of modern social media activism, a strong sense of social identity may not be a necessary prerequisite to action. Instead, the act of communicating beliefs and intentions may serve as a necessary spark to ignite feelings of group identity and subsequent collective action. A key to success that follows from this argument (which social media activists have no doubt already discerned) is to communicate ideological messages widely and frequently in the virtual social space.

An important limitation of the present (and the following) study is the extent to which the decisions made by participants can be considered to represent real-world activist decision-making. In particular, it must be kept in mind that activism can involve heavy costs that may not be reflected by the decision environment of the experiment. Activists are subject to economic risks such as losing their jobs and being arrested, as well as risks to their physical and mental wellbeing. Even where risk for activists is exclusively financial in nature, the stakes are likely higher for real-world decisions compared to those in the present experiment. For activists in non-Democratic societies, the potential costs may be even higher, with an elevated likelihood of incarceration, physical harm and death, which often occur in direct contravention of international Human Rights laws. For example, during and following the mass protests in Belarus concerning the 2020 Presidential elections, protesters were detained and sentenced at high rates, likely without due course of law, and an estimated 1000 cases of torture of detainees was reported (Human Rights Center “Viasna”, 2021). The overwhelming mistreatment of protesters prompted the United Nations Human Rights Office to release a statement condoning the situation and calling for the Belarusian authorities to release those protestors arrested for exercising their right to peaceful protest (*Office of the High Commissioner Human Rights*, 2020). Thus, in non-Democratic contexts, the potential costs of activism can be expected to be higher than in Democratic contexts because of the harsh responses to public dissent exercised by these regimes, which may be more likely to ignore basic human rights concerns. The exact nature of real-world activism is therefore not guaranteed to exactly follow the patterns observed in the present experiment, and future research could aim to address this limitation by investigating similar scenarios of social justice in a more naturalistic setting through field studies, perhaps complemented by qualitative methods.

A second limitation of this study concerns sample size and the observed power of the reported effects. Because no formal power analysis was conducted, and because there is little utility in calculating observed power post-hoc (Hoenig & Heisey, 2001), the power with which effects were detected in the present study is unknown. However, Brysbaert (2019) suggests that for a repeated measures design including an interaction, 110 participants are required for 80% power. This suggests that the power achieved in this study (and Study 2) was lower than 80%. It is therefore possible that some effects were not detected because of a lack of sensitivity due to insufficient sample size, and that the present analysis only revealed those effects that were sufficiently large to be detected with low power. The coarseness of this analysis may therefore have missed important subtle effects that are important for social decision-making and collective action. To remedy this, a replication study could be conducted with sample size determined by power analysis of the effects observed in the present study. This replication would provide more insight by demonstrating whether the results observed here can be reproduced, and in addition would reveal potentially overlooked effects that were too small to be detected.

In conclusion, the two experiments reported here provide empirical insight into the context-dependent influences of identity and efficacy in social dilemmas. Through the use of artificial unfair social systems and incentivised decision-making, I demonstrated and quantified the moderating influence of collective identification on the extent to which feelings of efficacy encouraged decisions to challenge the unfair system. In addition, I found that feelings of identification were more influential in eliciting system challenge responses when they were relatively more politicised. Though the availability of coordinating information affected decisions only in scenarios where politicisation of identity was at a minimum, I found that the pattern of results was slightly different under conditions of information. Specifically, it appears that extra information reduced the need for participants to infer efficacy from feelings of identification. The present study used novel methods to explore collective decision-making by framing it in a way that allowed for the influence of efficacy beliefs – and their moderation by feelings of identity and the ability to communicate – to be quantified numerically. In doing so, the results extend previous findings by demonstrating that group identity may act primarily as a coordinating tool to manage expectations and efficacy beliefs, and that when coordination can be achieved by other means, group identity may be less influential in eliciting action.

## References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Akfirat, S., Uysal, M. S., Bayrak, F., Ergiyen, T., Üzümçeker, E., Yurtbakan, T., & Özkan, Ö. S. (2020). *Social Identification and Collective Action Participation in the Internet Age: A Meta-Analysis*. PsyArXiv. https://doi.org/10.31234/osf.io/uqb7a

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, *9*(3), 75–78. https://doi.org/10.1111/1467-8721.00064

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bicchieri, C., & Lev-On, A. (2007). Computer-mediated communication and cooperation in social dilemmas: An experimental analysis. *Politics, Philosophy & Economics*, *6*(2), 139–168. https://doi.org/10.1177/1470594X07077267

Bornstein, G. (1992). The free-rider problem in intergroup conflicts over step-level and continuous public goods. *Journal of Personality and Social Psychology*, *62*(4), 597–606. https://doi.org/10.1037/0022-3514.62.4.597

Bouas, K. S., & Komorita, S. S. (1996). Group discussion and cooperation in social dilemmas. *Personality and Social Psychology Bulletin*, *22*(11), 1144–1150. https://doi.org/10.1177/01461672962211005

Brewer, M. B., & Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology*, *50*(3), 543–549. https://doi.org/10.1037/0022-3514.50.3.543

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1), 16. https://doi.org/10.5334/joc.72

Cohen-Chen, S., & Van Zomeren, M. (2018). Yes we can? Group efficacy beliefs predict collective action, but only when hope is high. *Journal of Experimental Social Psychology*, *77*, 50–59. https://doi.org/10.1016/j.jesp.2018.03.016

Constantine, M. G., Donnelly, P. C., & Myers, L. J. (2002). Collective self-esteem and Africultural coping styles in African American adolescents. *Journal of Black Studies*, *32*(6), 698–710. https://doi.org/10.1177/00234702032006004

Dawes, R. M. (1980). Social Dilemmas. *Annual Review of Psychology*, *31*(1), 169–193. https://doi.org/10.1146/annurev.ps.31.020180.001125

De Cremer, D., & Van Vugt, M. (2002). Intergroup and intragroup aspects of leadership in social dilemmas: A relational model of cooperation. *Journal of Experimental Social Psychology*, *38*(2), 126–136. https://doi.org/10.1006/jesp.2001.1499

Doosje, B., Spears, R., & Ellemers, N. (2002). Social identity as both cause and effect: The development of group identification in response to anticipated and actual changes in the intergroup status hierarchy. *The British Journal of Social Psychology*, *41*, 57–76. https://doi.org/10.1348/014466602165054

Drury, J., & Reicher, S. (2000). Collective action and psychological change: The emergence of new social identities. *British Journal of Social Psychology*, *39*(4), 579–604. https://doi.org/10.1348/014466600164642

Drury, J., & Reicher, S. (2005). Explaining enduring empowerment: A comparative study of collective action and psychological outcomes. *European Journal of Social Psychology*, *35*(1), 35–58. https://doi.org/10.1002/ejsp.231

Drury, J., & Reicher, S. (2009). Collective psychological empowerment as a model of social change: researching crowds and power. *Journal of Social Issues*, *65*(4), 707–725. https://doi.org/10.1111/j.1540-4560.2009.01622.x

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. https://doi.org/10.1007/s10683-006-9159-4

Foster, Mindi D., Adrianna Tassone, and Kimberly Matheson. "Tweeting about sexism motivates further activism: A social identity perspective." *British Journal of Social Psychology* 60, no. 3 (2021): 741–64. https://doi.org/10.1111/bjso.12431.r

Frohlich, N., & Oppenheimer, J. (1998). Some consequences of e-mail vs. face-to-face communication in experiment. *Journal of Economic Behavior & Organization*, *35*(3), 389–403. https://doi.org/10.1016/S0167-2681(98)00044-4

Hackett, S., Schlager, E., & Walker, J. (1994). The role of communication in resolving commons dilemmas: experimental evidence with heterogeneous appropriators. *Journal of Environmental Economics and Management*, *27*(2), 99–126. https://doi.org/10.1006/jeem.1994.1029

Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *The Sociology of Corrections International Journal of Criminology & Penology, 1(1), 69–97.* Haslam, S. A., Jetten, J., O'Brien, A., & Jacobs, E. (2004). Social identity, social influence and reactions to potentially stressful tasks: Support for the self-categorization model of stress. *Stress and Health*, *20*(1), 3–9. https://doi.org/10.1002/smi.995

Haslam, S. A., O'Brien, A., Jetten, J., Vormedal, K., & Penna, S. (2005). Taking the strain: Social identity, social support, and the experience of stress. *The British Journal of Social Psychology*, *44*, 355–370. https://doi.org/10.1348/014466605X37468

Haslam, S. A., & Reicher, S. (2006). Stressing the group: Social identity and the unfolding dynamics of responses to stress. *Journal of Applied Psychology*, *91*(5), 1037–1052. https://doi.org/10.1037/0021-9010.91.5.1037

He, S., Offerman, T., & van de Ven, J. (2017). The sources of the communication gap. *Management Science*, *63*(9), 2832–2846. https://doi.org/10.1287/mnsc.2016.2518

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician*, *55*(1), 19–24. https://doi.org/10.1198/000313001300339897

Hornsey, M. J., Blackwood, L., Louis, W., Fielding, K., Mavor, K., Morton, T., O’Brien, A., Paasonen, K.-E., Smith, J., & White, K. M. (2006). Why do people engage in collective action? revisiting the role of perceived effectiveness. *Journal of Applied Social Psychology*, *36*(7), 1701–1722. https://doi.org/10.1111/j.0021-9029.2006.00077.x

Human Rights Center "Viasna" (2021). *Human Rights Situation in Belarus in 2020. Analytical Review by Viasna*. Human Rights Center “Viasna.” http://spring96.org/en/news/101223

Jost, J. T. (2019). A quarter century of system justification theory: Questions, answers, criticisms, and societal applications. *British Journal of Social Psychology*, *58*(2), 263–314. https://doi.org/10.1111/bjso.12297

Kelly, C., & Breinlinger, S. (1995). Identity and injustice: Exploring women’s participation in collective action. *Journal of Community & Applied Social Psychology*, *5*(1), 41–57. https://doi.org/10.1002/casp.2450050104

Klandermans, B. (1984). Mobilization and participation: Social-psychological expansions of resource mobilization theory. *American Sociological Review*, *49*(5), 583–600. https://doi.org/10.2307/2095417

Klandermans, P. G. (2014). Identity politics and politicized identities: identity processes and the dynamics of protest. *Political Psychology*, *35*(1), 1–22. https://doi.org/10.1111/pops.12167

Klandermans, Bert, and Dirk Oegema. Potentials, networks, motivations, and barriers: steps towards participation in social movements. *American Sociological Review* 52, no. 4 (1987): 519–31. https://doi.org/10.2307/2095297.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer publishing company.

Lenth, R. V., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.6.0) [Computer software]. https://CRAN.R-project.org/package=emmeans

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Masson, T., Jugert, P., & Fritsche, I. (2016). Collective self-fulfilling prophecies: Group identification biases perceptions of environmental group norms among high identifiers. *Social Influence*, *11*(3), 185–198. https://doi.org/10.1080/15534510.2016.1216890

McCarthy, J. D., & Zald, M. N. (1977). Resource mobilization and social movements: A partial theory. *American Journal of Sociology*, *82*(6), 1212–1241. https://doi.org/10.1086/226464

Mikołajczak, G., & Becker, J. C. (2019). What is (un)fair? political ideology and collective action. *Journal of Social and Political Psychology*, *7*(2), 810–829. https://doi.org/10.5964/jspp.v7i2.1230

Mummendey, A., Kessler, T., Klink, A., & Mielke, R. (1999). Strategies to cope with negative social identity: Predictions by social identity theory and relative deprivation theory. *Journal of Personality and Social Psychology*, *76*(2), 229–245. http://dx.doi.org/10.1037/0022-3514.76.2.229

Office of the High Commissioner Human Rights (2020). *UN human rights experts: Belarus must stop torturing protesters and prevent enforced disappearances.* Office of the High Commissioner Human Rights. *https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=26199&LangID=E*

Olson, J. M., Roese, N. J., Meen, J., & Robertson, D. J. (1995). The preconditions and consequences of relative deprivation: Two field studies. *Journal of Applied Social Psychology*, *25*(11), 944–964. https://doi.org/10.1111/j.1559-1816.1995.tb02384.x

Osborne, D., & Sibley, C. G. (2013). Through rose-colored glasses: System-justifying beliefs dampen the effects of relative deprivation on well-being and political mobilization. *Personality and Social Psychology Bulletin*, *39*(8), 991–1004. https://doi.org/10.1177/0146167213487997

Osborne, D., Sibley, C. G., & Sengupta, N. K. (2015). Income and neighbourhood-level inequality predict self-esteem and ethnic identity centrality through individual- and group-based relative deprivation: A multilevel path analysis. *European Journal of Social Psychology*, *45*(3), 368–377. https://doi.org/10.1002/ejsp.2087

Outten, H. R., Schmitt, M. T., Garcia, D. M., & Branscombe, N. R. (2009). Coping options: Missing links between minority group identification and psychological well-being. *Applied Psychology: An International Review*, *58*(1), 146–170. https://doi.org/10.1111/j.1464-0597.2008.00386.x

Pettigrew, T. F., Christ, O., Wagner, U., Meertens, R. W., Dick, R. V., & Zick, A. (2008). Relative deprivation and intergroup prejudice. *Journal of Social Issues*, *64*(2), 385–401. https://doi.org/10.1111/j.1540-4560.2008.00567.x

Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology*, *52*(4), 597–617. https://doi.org/10.1111/bjso.12006

Runciman, W. G. (1966). *Relative deprivation and social justice: A study of attitudes to social inequality in twentieth-century England*. Routledge & Kegan Paul.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*(1), 58–92. https://doi.org/10.1177/1043463195007001004

Simon, B., & Klandermans, B. (2001). Politicized collective identity: A social psychological analysis. *American Psychologist*, *56*, 319–331. https://doi.org/10.1037/0003-066X.56.4.319

Snow, D. A., & Corrigall-Brown, C. (2015). Collective identity. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 174–180). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.10403-9

Sturmer, S., & Simon, B. (2004). Collective action: Towards a dual-pathway model. *European Review of Social Psychology*, *15*(1), 59–99. https://doi.org/10.1080/10463280340000117

Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. Academic Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of inter-group relations* (pp. 33–47). Brooks/Cole.

Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of intergroup behavior. In J. T. Jost & J. Sidanius (Eds.), *Political Psychology* (pp. 276–293). Psychology Press. https://doi.org/10.4324/9780203505984-16

Tausch, N., Becker, J. C., Spears, R., Christ, O., Saab, R., Singh, P., & Siddiqui, R. N. (2011). Explaining radical group behavior: Developing emotion and efficacy routes to normative and nonnormative collective action. *Journal of Personality and Social Psychology*, *101*(1), 129–148. https://doi.org/10.1037/a0022728

Thomas, E. F., Mavor, K. I., & McGarty, C. (2012). Social identities facilitate and encapsulate action-relevant constructs: A test of the social identity model of collective action. *Group Processes & Intergroup Relations*, *15*(1), 75–88. https://doi.org/10.1177/1368430211413619

Thomas, E. F., McGarty, C., & Mavor, K. (2016). Group interaction as the crucible of social identity formation: A glimpse at the foundations of social identities for collective action. *Group Processes & Intergroup Relations*, *19*(2), 137–151. https://doi.org/10.1177/1368430215612217

Thomas, E. F., McGarty, C., & Mavor, K. I. (2009). Aligning identities, emotions, and beliefs to create commitment to sustainable social and political action. *Personality and Social Psychology Review*, *13*(3), 194–218. https://doi.org/10.1177/1088868309341563

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

van Zomeren, M. (2013). Four core social-psychological motivations to undertake collective action. *Social and Personality Psychology Compass*, *7*(6), 378–388. https://doi.org/10.1111/spc3.12031

van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin*, *134*(4), 504–535. https://doi.org/10.1037/0033-2909.134.4.504

Vestergren, S., Drury, J., & Chiriac, E. H. (2018). How collective action produces psychological change and how that change endures over time: A case study of an environmental campaign. *British Journal of Social Psychology*, *57*(4), 855–877. https://doi.org/10.1111/bjso.12270

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Wright, S. C., Taylor, D. M., & Moghaddam, F. M. (1990). Responding to membership in a disadvantaged group: from acceptance to collective protest. *Journal of Personality and Social Psychology*, *58*(6), 994–1003. https://doi.org/10.1037/0022-3514.58.6.994

# Study 2. The Weakness of Groups: How the opportunity to freeride affects collective decisions to challenge unfair social systems

## Abstract

In real-world settings, the incentives of group-based decisions to change unfair systems may vary. Specifically, it is often the case that conflict of motives exists between what is optimal for the individual and what is the best outcome for the group. In this study, I investigated whether and how the incentive structure of group decision-making amongst disadvantaged individuals influences the likelihood of challenging unfair systems by comparing decisions in two types of social dilemma; the Assurance Game – where freeriding was not possible and cooperation is the best for both the individual and the disadvantaged group – and the Prisoner's Dilemma – where freeriding was possible and the best strategy for the individual conflicted with the best strategy for the group. I explored this both under conditions where participants were able and unable to communicate intentions before deciding. One-hundred-and-seventy-eight participants made decisions using interactively linked computers in an artificial social system that was rendered unfair by allocating two participants to "Elite" roles based either on merit ("Legitimate Elite") or by "stealing" an Elite position from a top scorer ("Illegitimate Elite"). The Elites in the experiment could receive considerably higher payoffs than the non-Elites. I used a combination of generalised linear and linear mixed model analysis to identify how participants' decisions were affected by the type of Elite present and whether freeriding was possible. When communication was not possible (Experiment 1, *N* = 91, ~70% female, ~91% undergraduate, of which 65% were psychology students), I found that freeriding decreased the likelihood of challenging the system when the system had No Elite beneficiary but not when there was either type of Elite present. When communication was possible (Experiment 2, *N* = 87, ~62% female, ~93% undergraduate, of which 90% were psychology students), under conditions of freeriding participants were more likely to challenge the system

when the meritocratically-assigned Elite benefited from the systemic unfairness, compared to when No Elite benefited. Overall, the results suggest that participants were especially mobilised against the Legitimate Elite in the higher-stakes Prisoner's Dilemma scenario where freeriding was possible.

## Introduction

In Study 1, I found that individuals feel more efficacious and are more likely to challenge an unfair social arrangement if they are part of a group rather than a lone agent, and that feelings of identification were important for this. Due to the payoff structure of the group contexts in Study 1, there was a strong incentive to make choices that were believed to be consistent with the choices of others (see below for explanation as to why). As a result, personal and group incentives were aligned.

In many real-world collective decision scenarios, however, this alignment of individual and collective interest often cannot be taken for granted. For example, recycling behaviour benefits the collective by helping to sustain the planet, but involves personal costs in terms of time and effort, and the benefits of such behaviour are less tangible on a personal level than the immediate costs, because they are not immediately evident (Lee et al., 2019; Smith et al., 1994). Likewise, individuals can avoid contributing to public media services (e.g., the British Broadcasting Company), but still enjoy its productions. In terms of activism, the personal costs of engaging in a social movement can be non-negligible (e.g., burnout and stress; Chen & Gorski, 2015) and this can make it even more appealing to avoid pursuing a collective goal and instead let others bear the burden. In these cases, incentives to avoid collaboration can dominate intentions to engage in collective efforts. That is, the best option for individual decision-makers is not to act, since doing so means they avoid paying a personal cost and yet still benefit from any positive outcomes achieved by others. Clearly, though, if everyone followed this strategy, no collective gains could be achieved, making the outcome of mutual non-cooperation worse than mutual cooperation. For example, if nobody engages with sustainability behaviours, everybody will suffer the effects of a less healthy planet; if nobody contributes to public services those services will operate less well or even disappear; if nobody votes then the Democratic process of government will not function as it should; and if nobody engages in a collective action to challenge a problematic social issue, nobody will enjoy positive social change.

The above incentive structure characterises a social dilemma, known in contemporary literature specifically as the Prisoner's Dilemma (Dawes, 1980; Hardin, 1968; Olson, 1965). In behavioural economic terms, the situation is a Prisoner's Dilemma (i) if there is a dominant strategy for an individual to defect, and (ii) the benefits of unanimous cooperation are greater than the benefits of unanimous defection. That is, (i) it is always better for an individual not to cooperate and (ii) if nobody cooperates, the result is worse for everybody than if everyone had cooperated. Formal definition of the Prisoner's Dilemma criteria can be described with the following three equations:

$$\mathrm{P_n} > \mathrm{Q_0} \tag{1}$$

$$\mathrm{P_i} < \mathrm{Q_{i\text{-}1}}, \text{ for } i = 1, \ldots, n \tag{2}$$

$$\mathrm{P_i} < \mathrm{Q_i}, \text{ for } i = 1, \ldots, n - 1 \tag{3}$$

where

$P$ = payoff for cooperators,

Q = payoff for defectors,

$n$ = the number of decision-makers,

$i$ = the number of cooperators and $i \neq n$ and $i \neq 0$.

Equation 1 specifies that unanimous cooperation yields a greater payoff than unanimous defection (the value of the difference between $P_n$ and $Q_0$ – if it is positive – is known as *cooperative gain*). Equation 2 specifies that a decision-maker can increase their payoff if they choose to defect instead of cooperate (known as the *temptation*). Equation 3 specifies that in situations where there is at least one cooperator and one defector, defection will always be more profitable than cooperation. It follows from Equations 2 and 3 that in the Prisoner's Dilemma the only Nash equilibrium is defection; there is no choice other than defection that any one decision-maker could make that would improve their result.

Variations on this incentive structure model the interdependencies in social dilemmas in different ways (Van Lange et al., 2013). One such variation is known as the Assurance Game (also known as the Stag Hunt game). In the Assurance Game, Equation 1 holds and there is a non-zero cooperative gain; unanimous cooperation is more profitable that unanimous defection. In contrast to the Prisoner's Dilemma, there is no dominant strategy in the Assurance Game (i.e., Equations 2 and 3 do not apply). Instead, there are two equilibria: unanimous defection *or* unanimous cooperation. That is, cooperation is the best strategy if everyone else cooperates, whereas defection is best strategy if everyone else defects. This means that the best strategy depends on what others (are believed to) do, placing importance on social coordination (McAdams, 2009; Skyrms, 2004), and that there is no temptation to defect since the interests of the collective are always aligned with those of the individual.

For the purposes of this study, these formal definitions of the Prisoner's Dilemma and Assurance Game can be summarised by one critical difference; in the Assurance Game it is not possible to "freeride", whereas in the Prisoner's Dilemma it is. Freeriding refers to the act of not contributing to a collective endeavour and yet still benefiting from any positive outcomes that it confers (Olson, 1965). As alluded to above, the opportunity to freeride has the potential to undermine collective goals if individuals value their own interests over group interests.

Comparison of the Assurance Game and the Prisoner's Dilemma highlights the impact the different incentive structures have on cooperation rates. Boone et al. (2010) showed that cooperation is higher in the Assurance Game (around 59% cooperation rate) compared to the Prisoner's Dilemma (34-39% cooperation rate). However, despite the strong incentives to freeride if the opportunity presents, individuals are consistently found to deviate from the pure defection strategy in social dilemmas that profit-maximising decision-makers would adopt. For example, a meta-analysis of 17 studies investigating the public goods game identified *conditional* cooperation as the prevailing decision strategy (Thöni & Volk, 2018). A conditional co-operator is classified as a decision-maker whose likelihood of cooperating increases with the extent to which others are known or believed to also cooperate (Fischbacher et al., 2001a; Thöni & Volk, 2018). In contrast, approximately 30% of participants are classified as freeriders (e.g., Fischbacher et al., 2001; Thöni & Volk, 2018), who always defect regardless of the (believed or known) decisions of others. Likewise, in the Prisoner's Dilemma decision-makers are more cooperative than rational self-interest would predict (Colman, 2003; Kollock, 1998). Thus, most individuals tend not to be motivated by pure self-interest but instead show concern for the social context and collective outcomes.

A possible explanation for the non-zero cooperation observed in the Prisoner's Dilemma is that decision-makers transform the objective payoffs of the scenario into subjective utilities based on their preferences. In this way, researchers have argued that cooperation in the Prisoner's Dilemma may be achieved by individuals making decisions as if they were in an Assurance Game (Boone et al., 2010; Kollock, 1998). That is, they transform a problem involving mixed motives (i.e., self-interest versus group-interest) into a problem of coordination (Simpson, 2004). It appears that this transformation is most common amongst individuals who have relatively strong concern for others' outcomes (i.e., those with a prosocial Social Value Orientation [SVO]; Bogaert et al., 2008; Boone et al., 2010; Van Lange, 1999), supporting the notion that personal preferences become integrated into the subjective value associated with decision options.

The ability for decision-makers to communicate with one another is understood to enhance cooperation in social dilemmas. On average, being able to coordinate responses by communicating is suggested to increase cooperation by approximately 40% (Sally, 1995). In a more recent meta-analysis, Balliet (2010) reiterated the finding of greater cooperation rates under communication conditions, with the added insights that face-to-face – versus written – communication, and communication in larger – versus smaller – groups leads to greater cooperation rates. Communication can therefore act as an opposing force to the deleterious effects of freeriding on cooperation.

Beyond the incentive structures of these games, the context in which decisions are presented is likely to be important in guiding decision-making in social dilemmas. In Study 1, I showed that decision-making in groups was more conducive to cooperation than decision-making individually. Surprisingly, the involvement of an out-group class of participant had no direct influence on cooperation rates in Study 1. It did, however, play a role in the extent to which feelings of identification induced cooperative decisions, suggesting that intergroup comparisons were (indirectly) relevant for guiding behaviour, a conclusion consistent with social psychological theory and research (e.g., Simon & Klandermans, 2001; Turner et al., 1987; van Zomeren et al., 2008). An interesting question, therefore, concerns whether and how intergroup comparisons are relevant in the context of freeriding. Based on the insight from this literature and Study 1, it is possible that the galvanisation afforded by intergroup comparisons may sustain cooperation when freeriding is possible.

These comparisons are not just a theoretical exercise. In the context of real-world social issues, an understanding of the conditions under which individuals are more and less likely to cooperate collectively can have important societal implications. Especially in the context of social inequality – the ill effects of which manifest insidiously in many aspects of society (e.g., Clark et al., 2014; Patel et al., 2020; Perlin et al., 2001) – understanding the most effective ways in which (disadvantaged) individuals are able to cooperate towards the collective goal of overcoming inequality can inform, for example, best practice for social movement campaigns. More effective grassroots initiatives, in turn, will have greater influence on policymakers to create the necessary infrastructure for a fairer society. As alluded to above, however, a considerable challenge concerns the power of incentives at the individual level to undermine group cooperation.

In this study, I investigated whether experimentally disadvantaged individuals' decisions to challenge the system would differ depending on whether it was possible to freeride or not. I did this by creating social systems that differed in terms of their incentive structure. One type of system was an Assurance Game – where freeriding was not possible and the best outcome for the individual was aligned with best outcome for the collective – whilst the other was a Prisoner's Dilemma – where freeriding was possible and the best outcome for the individual always resulted from not cooperating. In all other respects these systems were identical. The incentive structures were manipulated in this way to explore whether the group-based elevation of cooperation observed in Study 1 would still hold when it was possible for group members to take advantage of one another. This is important because many real-world instances of collective action involve opportunities to shirk on collective responsibility and yet still benefit from any positive outcomes of action. As in Study 1, the systems also differed in the extent to which social inequality was made salient. This was done by creating an "Elite" class of participant who could receive considerably more compensation for the experiment and a non-Elite class whose compensation was lower. The Elite either earned their position by scoring the highest at an addition task ("Legitimate Elite") or by successfully "stealing" the role from the second-highest scorer ("Illegitimate Elite"). Thus, this study followed a 2 x 3 factorial design. I report the results of two experiments in this framework; one where communication of intentions prior to deciding was not possible and one where it was. I hypothesised that, overall, freeriding would have a negative influence on the likelihood of challenging the system. However, the deleterious effects of freeriding were expected to be mitigated by two important factors: whether participants could communicate intentions and thereby coordinate their responses, and whether group membership was made salient by the comparisons with an "Elite" outgroup. As in Study 1, I aimed to build on the existing literature by measuring participants' experiences of fairness, group identification, and group efficacy to explore how these perceptions were affected by the opportunity to freeride and how they in turn influenced participants' decision-making. I hypothesised that all three experiences would be lower when freeriding was possible and that each would be negatively related to the likelihood of deciding the challenge the system.

## Experiment 1

## Method

### Overview of the experiment

The data for Study 2 were acquired from the same experimental sessions as those described in Study 1. In addition to the paradigms detailed in Study 1, participants also completed a further three paradigms, which I detail more fully below. The focus of Study 2's analysis was the comparison of no-freeriding group scenarios to freeriding group scenarios. The no-freeriding scenarios for the present purposes were termed the "group" scenarios in Study 1. The additional three paradigms not yet detailed were very similar to the group scenarios detailed in Study 1, but with the difference that freeriding was possible in these paradigms (whereas it was not in the previously described paradigms). Full details of this difference are provided below. Hereafter I refer only to the methodology relevant for the present experiment, but the reader is reminded that all data considered herein were collected at the same time as the data considered in Study 1, and indeed the data corresponding to the no-freeriding paradigms here are exactly the same data as the "group" paradigms of Study 1. Studies 1 and 2 were conducted concurrently in this way for efficiency. Because the experiments were fairly complex, separating the two studies would have involved instructing multiple participant pools, requiring considerably more time and resources. Instead, by combining the two research questions into one data collection event I took advantage of the fact that participants were already familiar with the basic principles of the experiment, making it possible to collect more data in less time and using fewer resources.

Participants took part in six games, reflecting the 2 (No Freeriding vs Freeriding) x 3 (No Elite vs Legitimate Elite vs Illegitimate Elite) within-subjects design. Recall that, in each game, participants were endowed with 100 coins (equivalent to £8.00). Participants had to make a choice between supporting the system by paying 50 coins (i.e., half their endowment) and keeping their remaining 50 coins for certain, or challenging the system by not paying for a chance of keeping all 100 coins (with a reciprocal chance of receiving 0, making this a risky choice). Note that "challenging" here is the cooperative choice; a decision to challenge increases the chance of overturning the system for everybody but comes at the personal risk of being caught out and receiving 0, whereas a decision to support has no impact on the chances of overturning the system but ensures that the individual receives 50. In the conditions involving an Elite, the payments of participants would go directly to the Elite, whereas payments in the No Elite condition would not go to anyone. Decisions were incentivised; participants were aware that the result from one game would be randomly selected at the end of the experiment and they would be paid according to their monetary outcome in that game. The key difference of the Freeriding manipulation was that, under conditions of No Freeriding, a choice not to challenge (i.e., support) the system excluded participants from benefiting in the event that the system was overturned, whereas in the Freeriding condition, if the system was overturned even those individuals who did not challenge the system would benefit. Thus, participants had the opportunity to freeride by avoiding taking a personal risk but still potentially benefiting from the risks taken by others (see ***Figure 2.1*** below for a graphical illustration of how these differing incentives compare).

## Materials

### *The Games*

Six games – comprising a 2 x 3 within-subjects design – were described and presented as social systems in which participants made decisions. Two factors determined the structure of each game: whether participants had to challenge the system to benefit from its overturn or not (No Freeriding vs Freeriding, hereafter referred to as 'Freeriding' or 'FR') and whether there was No Elite, the Legitimate Elite, or the Illegitimate Elite to whom payments were made (hereafter referred to as 'Elite'). Each game consisted of two phases. Phase 1 presented the System (i.e., specifying the Freeriding and Elite condition) and participants were asked to decide either to pay 50 of their 100-coin endowment to the system, or to challenge the system by not paying for a chance of keeping all 100 coins. The probability of successfully challenging the system and receiving 100 coins was determined by the number of disadvantaged participants who chose to challenge (see below for details). In Phase 2, participants were presented with a series of questions about the scenario (see Study 1 Methods for full details of Phase 2). These were presented one after another on the participants' computers, in the following order. All games were one-shot decisions (i.e., only one decision per paradigm) to avoid the possibility of 'bet-hedging' within each game (for example, by making alternating decisions across a series of trials). No feedback about the outcome of decisions was provided so that knowledge of past outcomes could not influence decisions. To incentivise participants, the outcome of one game in the experimental session was randomly chosen at the end and converted to Great British Pounds and paid to participants in addition to their showup fee.

**No Freeriding.** As explained above, these games were the ‘group’ games presented in Study 1. A separate system was presented for each level of the Elite condition. Participants decided between paying 50 coins of their endowment and keeping 50 for certain or to challenge the system by attempting to pay 0. Each circle who challenged the system by paying 0 would increase the likelihood of the System being overturned by approximately [85 / group size] %. If the system was overturned, those who decided to challenge would successfully keep all 100 coins of their endowment, whilst those who paid 50 to the system would still only keep the 50 remaining to them after paying. If the system was not successfully overturned, those who tried to avoid paying would be punished and receive 0, whilst those who paid 50 would keep their 50. As described in Study 1, participants completed a series of comprehension questions to check their understanding of these contingencies. Thus, it was clear to participants that they could only benefit from system overturn by taking a risk and challenging. This payoff structure is reflected in ***Table 2.1*** and ***Figure 2.1***.

**Table 2.1.** *Expected payoffs for player i based on number of other (j) players challenging the system when freeriding is not possible*

| Number of other (*j*) challengers | Probability of overturn if player *i* challenges | System challenge expected value | System support expected value |
|---|---|---|---|
| 0 | 0.05 | 5 | 50 |
| 1 | 0.1 | 10 | 50 |
| 2 | 0.15 | 15 | 50 |
| 3 | 0.2 | 20 | 50 |
| 4 | 0.25 | 25 | 50 |
| 5 | 0.3 | 30 | 50 |
| 6 | 0.35 | 35 | 50 |
| 7 | 0.4 | 40 | 50 |
| 8 | 0.45 | 45 | 50 |
| 9 | 0.5 | 50 | 50 |
| 10 | 0.55 | 55 | 50 |
| 11 | 0.6 | 60 | 50 |
| 12 | 0.65 | 65 | 50 |
| 13 | 0.7 | 70 | 50 |
| 14 | 0.75 | 75 | 50 |
| 15 | 0.8 | 80 | 50 |
| 16 | 0.85 | 85 | 50 |

In formal terms, the expected value, $\pi_i$, of each decision is provided by the following equations. A choice to challenge and pay 0 is expressed with $\mathbb{C} = 0$. A choice to support and pay 50 is expressed with $\mathbb{C} = 1$.

If $\mathbb{C}_i = 0$,

$$\pi_i = 100\alpha + 100(\alpha \sum_{\substack{j=1 \\ j \neq i}}^{n-1} (1 - \mathbb{C}_j\ ))$$

If $\mathbb{C}_i = 1$,

$$\pi_i = 50$$

where $\alpha = 0.85 \times \frac{1}{n}$,

$\mathbb{C}_j =$ choice of player j,

and $n =$ number of participants in interaction group.

These equations specify that the expected value of a choice to challenge is derived from one's own decision and the number of individuals who challenge. In contrast, the expected value of a choice to support is always 50.

**Freeriding.** The Freeriding games presented a near-identical scenario to the No Freeriding games, except that in Freeriding games it was not necessary for a participant to challenge the system to benefit from its overturn. In these games, if the system was overturned, *all participants* would successfully keep all 100 coins of their endowment. That is, even those who decided to pay 50 would receive their payment back if the system was overturned. If the system was not successfully overturned, those who challenged would be punished and receive 0, whilst those who paid 50 would keep their remaining 50. Thus, in these games there was a dominant defection strategy of paying 50. Regardless of the decisions of others, it was always more profitable (in terms of expected value) to decide to pay 50; in doing so a participant secures a certain 50-coin payoff but also benefits from a possible additional 50 if the system is overturned. This payoff structure is reflected in ***Table 2.2*** and ***Figure 2.1***.

**Table 2.2.** *Expected payoffs for player i based on number of other (j) players challenging the system when freeriding is possible*

| Number of other (*j*) challengers | Probability of overturn if player *i* challenges | System challenge expected value | System support expected value |
|---:|---:|---:|---:|
| 0 | 0.05 | 5 | 50 |
| 1 | 0.1 | 10 | 52.5 |
| 2 | 0.15 | 15 | 55 |
| 3 | 0.2 | 20 | 57.5 |
| 4 | 0.25 | 25 | 60 |
| 5 | 0.3 | 30 | 62.5 |
| 6 | 0.35 | 35 | 65 |
| 7 | 0.4 | 40 | 67.5 |
| 8 | 0.45 | 45 | 70 |
| 9 | 0.5 | 50 | 72.5 |
| 10 | 0.55 | 55 | 75 |
| 11 | 0.6 | 60 | 77.5 |
| 12 | 0.65 | 65 | 80 |
| 13 | 0.7 | 70 | 82.5 |
| 14 | 0.75 | 75 | 85 |
| 15 | 0.8 | 80 | 87.5 |
| 16 | 0.85 | 85 | 90 |

Again, expressing this payoff structure formally, the expected value, $\pi_i$, of each decision is provided by the following equations. A choice to challenge and pay 0 is expressed with $\mathbb{C} = 0$. A choice to support and pay 50 is expressed with $\mathbb{C} = 1$.

If $\mathbb{C}_i = 0$,

$$\pi_i = 100\alpha + 100(\alpha \sum_{\substack{j=1 \\ j \neq i}}^{n-1} \left(1 - \mathbb{C}_j\right))$$

If $\mathbb{C}_i = 1$,

$$\pi_i = 50 + 50(\alpha \sum_{\substack{j=1 \\ j \neq i}}^{n-1} \left(1 - \mathbb{C}_j\right))$$

where $\alpha = 0.85 \times \frac{1}{n}$,

$\mathbb{C}_j =$ choice of player j,

and $n =$ number of participants in interaction group.

The first equation is identical to that in the No Freeriding condition and specifies that the expected value of a choice to challenge is derived from one's own decision and the number of individuals who challenge. The second equation is distinct in specifying that a choice to support the system provides a certain value of 50 and an additional expected value contingent on the number of others who challenge.

As in Study 1, following their decision in each paradigm, participants completed the ratings phase where they provided their experience of the system in terms of feelings of identification, number belief, and fairness (detailed in the 'Phase 2' subsection of the Methods section of Study 1, Experiment 1).

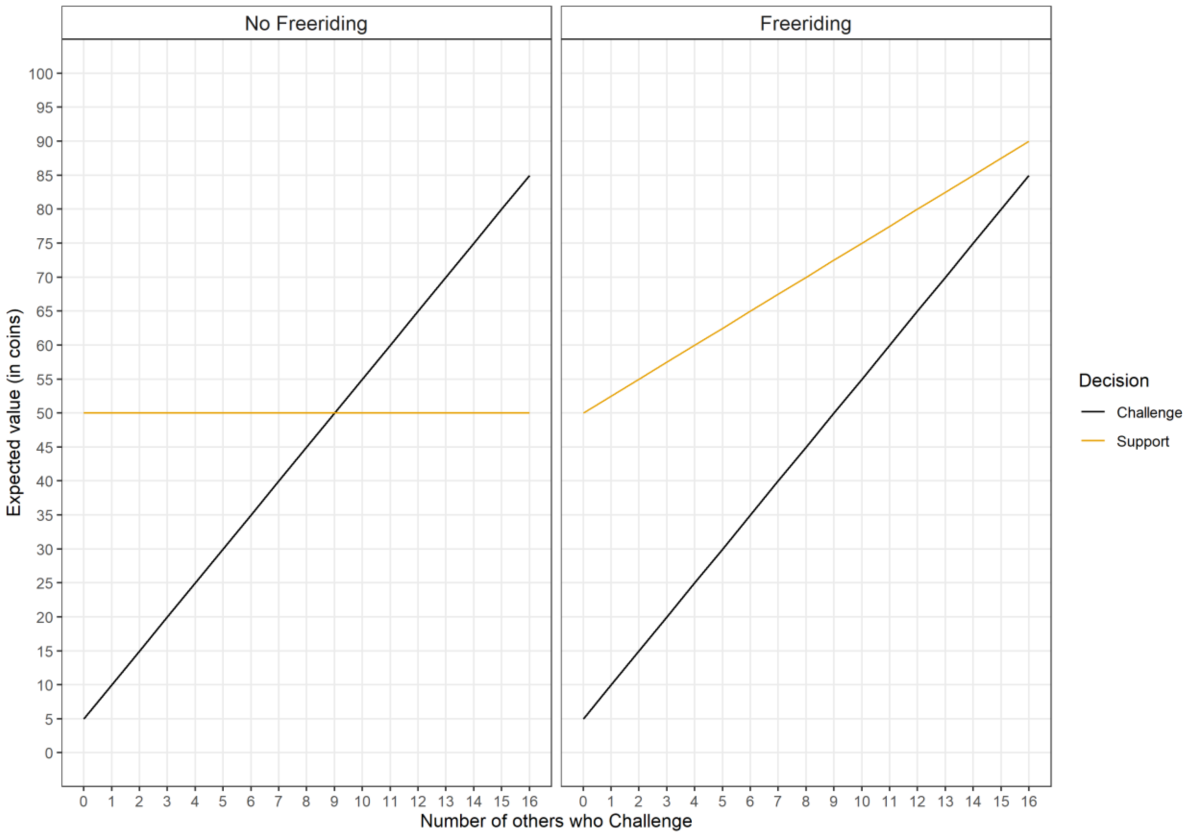


*Figure 2.1.* Expected value of each decision for player i based on the number of others who challenge and whether freeriding is possible. When freeriding is possible, the expected value of supporting the system will always be greater than the expected value of challenging the system.

# Results

## Computed variables

### *Number belief*

As in Study 1, Number belief was expressed as a proportion of group size minus one. For example, for a given decision made as part of a group with nine other circles, if a participant believed that 5 other circles would pay 0, their number belief would be transformed into a proportion like so: 5 / 9 = 0.56. This value was then multiplied by 100, so that an increment of 1 corresponded to a proportion change of 1% (this was done so that odds ratios were more readily interpretable; without doing so would mean that an increment of 1 would correspond to a proportion change of 100%).

***Risk preference***

A measure of Risk preference was constructed in the same way as in Study 1. Participants' responses in the Risk preference task were coded according to the degree of risk seeking or risk avoidance they represented. First, decisions to choose the risky 100 option were coded as 'risky' and decisions to choose the certain 50 option were coded as 'non-risky'. Risk-seeking was quantified as

$$risk\ seeking = 1 - P(chosen\ risky\ option) - 0.1$$

The extra subtraction of 0.1 was included so that the maximum and minimum possible values for risk-seeking were the same as those for risk avoidance. Risk avoidance was quantified as

$$risk\ avoidance = 1 - (1 - P(non - chosen\ risky\ option))$$

To clarify, consider a choice between the certain option and a more rewarding but risky option with probability 0.7. If the risky option is chosen, risk-seeking for this choice would be quantified as 1 – 0.7 – 0.1 = 0.2, reflecting the relatively low risk associated with the choice (recall the maximum probability of a risky option is always 0.85). If the non-risky option is chosen, risk-avoidance would instead be quantified as 1 – (1 – 0.7) = 0.7, reflecting the relatively risk avoidant nature of the choice. To create the final measure of RP, the values for risk-seeking and risk-avoidance for each choice were summed and the total risk avoidance score was subtracted from the total risk-seeking score:

$$risk\ preference = (\sum risk\ seeking) - (\sum risk\ avoidance)$$

In this way, individuals scoring above 0 are relatively risk-seeking and individuals scoring below zero are relatively risk-avoidant.

Analyses were conducted using RStudio. The majority of analyses were achieved using the ‘lme4’ package (version 1.1.26; Bates et al., 2015). I also made extensive use of ‘ggplot2’ (version 3.3.3; Wickham, 2016) and ‘ggeffects’ (version 1.0.1; Lüdecke, 2018) for visualisation and ‘emmeans’ (version 1.5.4; Lenth et al., 2021) for post-hoc analysis. For each model presented I ran a series of diagnostic tests to check that underlying model assumptions were met. In the interests of space, I only report where diagnostic tests flagged potential issues. In all analyses, I implemented the ‘bobyqa’ optimizer and allowed a maximum of 200,000 iterations for convergence.

**System Challenge decisions**

I first investigated the effect of whether it was possible to freeride and the type of Elite present on decisions to challenge the System. This was achieved by creating generalised linear mixed models (GLMMs) predicting the binomial variable “System Challenge” from the Freeriding and Elite manipulations. I controlled for Risk preference and Group size. Freeriding was a significant negative predictor of System Challenge (*log odds* = -1.43, *odds ratio (OR)* = 0.24, $z$ = -3.19, $p$ = 0.001); the odds of a circle deciding to challenge the system were 0.24 times lower when Freeriding was possible compared to when it was not. Whilst the main effects of the Elite conditions were non-significant, there was a significant Elite x Freeriding interaction (***Table 2.3***).

**Table 2.3.** *Generalised linear mixed model predicting System Challenge from the experimental manipulations - Freeriding and Elite - and their interaction*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 1.25 | 0.10 – 16.21 | 0.864 |
| Group size | 1.17 | 0.95 – 1.44 | 0.133 |
| Risk preference | 1.09 | 0.89 – 1.33 | 0.396 |
| Freeriding† | 0.24 | 0.10 – 0.58 | **0.001** |
| Legitimate elite†† | 0.78 | 0.35 – 1.73 | 0.545 |
| Illegitimate elite†† | 0.70 | 0.30 – 1.61 | 0.399 |
| Legit. x Freeriding | 3.51 | 1.10 – 11.26 | **0.034** |
| Illegit. x Freeriding | 4.61 | 1.39 – 15.32 | **0.013** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00}$ Subject | 4.51 | | |
| $\tau_{11}$ Subject.Freeriding | 0.59 | | |
| $\rho_{01}$ Subject | -0.27 | | |
| ICC | 0.57 | | |
| N Subject | 91 | | |
| Observations | 519 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.055 / 0.594 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

Post-hoc tests showed that the interaction was driven by a significant influence of Freeriding in the No Elite condition ($b$ = 1.43, $z$ *ratio* = 3.19, $p$ = 0.001; ***Table 2.4***); the opportunity to freeride was found to reduce system challenge decisions when No Elite was present but to have to no effect when either Legitimate or Illegitimate Elites were present (***Figure 2.2***).

***Table 2.4.*** *Post-hoc tests comparing the differences between Freeriding scenarios for each Elite condition*

| *Contrast* | *Elite* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| No FR - FR | No Elite | 1.430 | 0.448 | Inf | 3.190 | **0.001** |
| No FR - FR | Legit. Elite | 0.174 | 0.422 | Inf | 0.412 | 0.680 |
| No FR - FR | Illegit. Elite | -0.099 | 0.437 | Inf | -0.225 | 0.822 |

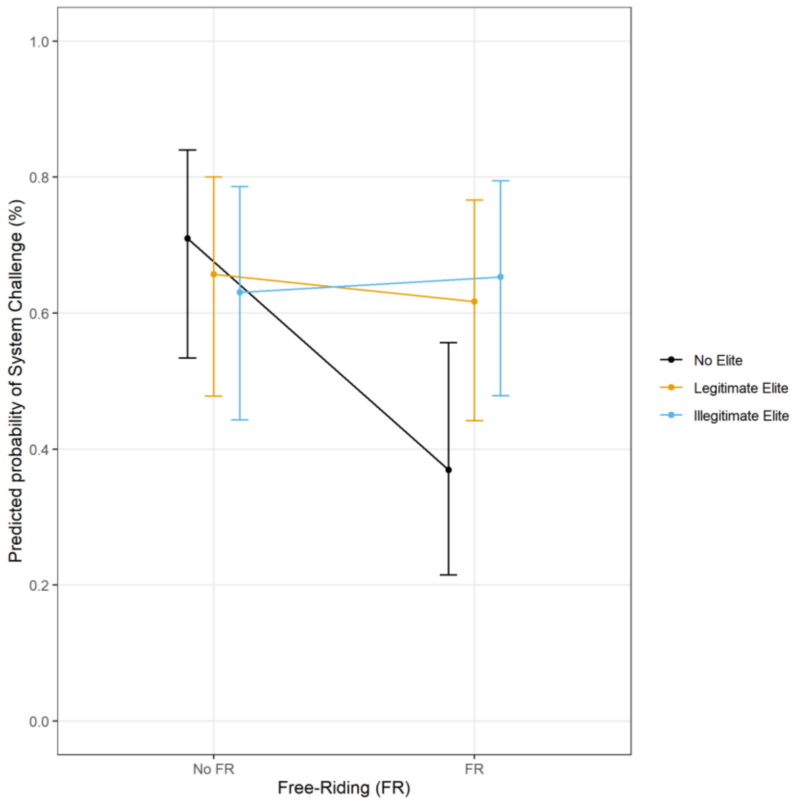


*Figure 2.2.* Predicted probability of System Challenge as a function of whether it was possible to Freeride and Elite condition.

## Subjective experience of the System

Recall from the Methods section of Study 1 that in each game, after making a decision, participants were asked to rate their experience of each system by indicating how fair they felt it to be (“Fairness”), how much they felt associated with the other group members (“Identification”), and how many other group members they believed had decided to challenge the system (“Number belief”). I explored the influence of Freeriding and Elite on these ratings. I did this by specifying a model for each variable predicted by Freeriding, Elite, and their interaction, controlling for group size. For all models in these analyses, I first attempted modelling the most complex random effects structure including random slopes for both Freeriding and Elite. If this most complex model did not converge (which was most often the case), I remodelled by removing one random slope at a time until the model converged.

### *Fairness*

System Fairness was not predicted by Freeriding ($b$ = -0.32, $t(429.66) = -1.8$, $p$ = 0.072). Both Legitimate ($b$ = -0.81, $t(428.13) = -4.85$, $p < .001$) and Illegitimate ($b$ = -1.03, $t(429.72) = -5.87$, $p < .001$) Elite conditions negatively predicted Fairness, compared to the Elite absent condition; when in a system with either the Legitimate or Illegitimate Elite, circles rated the system as less fair than when they were in a system with no Elite present. All combinations of the Freeriding x Elite interaction effect on fairness ratings were non-significant (see ***Table 2.5***).

**Table 2.5.** *Linear mixed effects model predicting Fairness*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 3.99 | 3.63 – 4.34 | < **0.001** |
| Group size | -0.02 | -0.12 – 0.08 | 0.687 |
| Freeriding (FR)† | -0.32 | -0.66 – 0.03 | 0.071 |
| Legitimate Elite†† | -0.81 | -1.14 – -0.48 | < **0.001** |
| Illegitimate Elite†† | -1.03 | -1.37 – -0.69 | < **0.001** |
| Legit. x FR | 0.20 | -0.28 – 0.67 | 0.418 |
| Illegit. x FR | 0.38 | -0.11 – 0.87 | 0.125 |
| **Random Effects** | | | |
| $\sigma^2$ | 1.28 | | |
| $\tau_{00}$ Subject | 1.70 | | |
| ICC | 0.57 | | |
| N Subject | 91 | | |
| Observations | 519 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.049 / 0.592 | | |

*Note.* † Reference group = No freeriding; Treatment group = Freeriding

†† Reference group = No Elite

Post-hoc contrasts showed in both the No Freeriding and the Freeriding conditions there were significant differences in Fairness between the Legitimate Elite and No Elite conditions and the Illegitimate and No Elite conditions, but no difference between the Illegitimate and Legitimate Elite conditions (***Table 2.6, Figure 2.3***).

**Table 2.6.** *Post-hoc contrasts of Fairness ratings between Elite conditions for No Freeriding and Freeriding scenarios*

| *Contrast* | *Freeriding* | *Estimate* | *SE* | *df* | *t ratio* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legitimate Elite | No FR | 0.813 | 0.169 | 433.054 | 4.826 | **< 0.001** |
| No Elite - Illegitimate Elite | No FR | 1.030 | 0.177 | 434.667 | 5.834 | **< 0.001** |
| Legitimate Elite - Illegitimate Elite | No FR | 0.217 | 0.177 | 434.667 | 1.229 | 0.437 |
| No Elite - Legitimate Elite | FR | 0.616 | 0.177 | 434.610 | 3.474 | **0.002** |
| No Elite - Illegitimate Elite | FR | 0.649 | 0.177 | 434.610 | 3.660 | **0.001** |
| Legitimate Elite - Illegitimate Elite | FR | 0.033 | 0.169 | 433.054 | 0.196 | 0.979 |

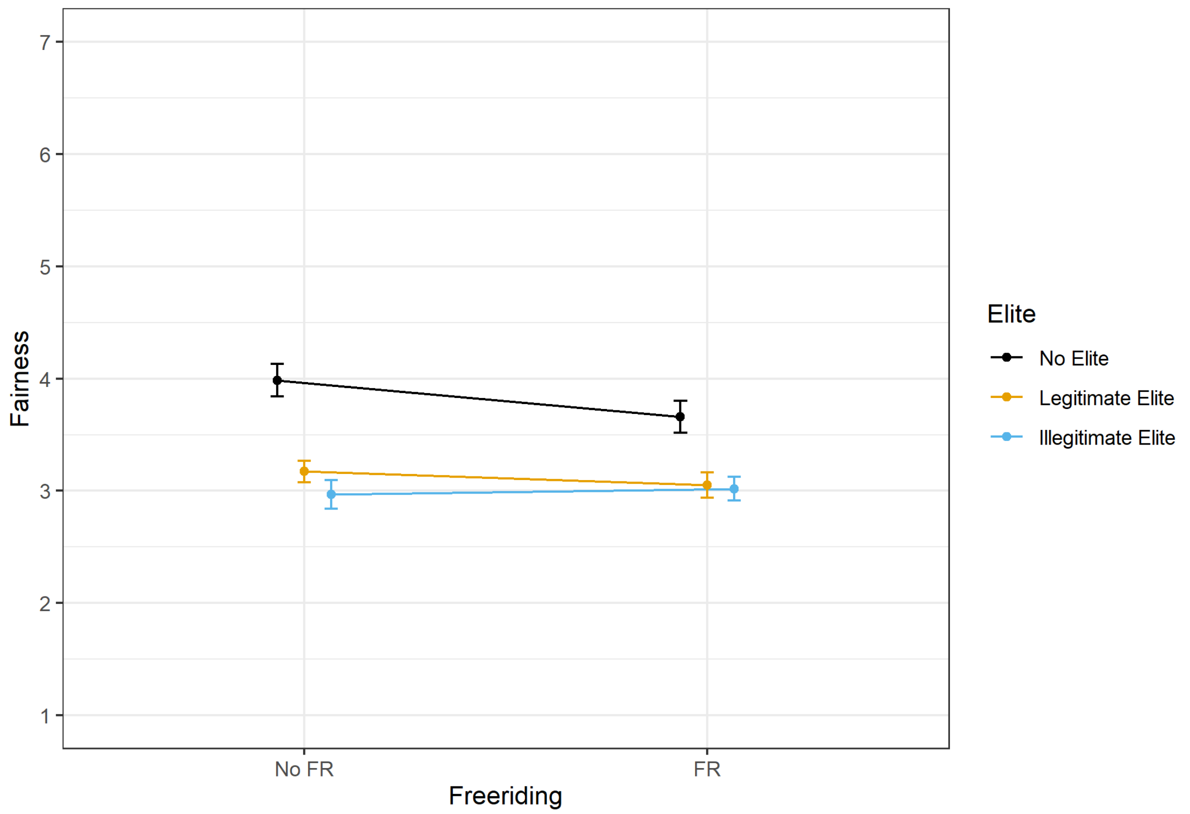


*Figure 2.3.* Mean ratings of System Fairness in each Freeriding and Elite condition. Error bars represent within-subject standard error of the mean.

### *Identification*

Identification was not predicted by any factor in the model (***Table 2.7***).

**Table 2.7.** *Linear mixed effects model predicting Identification*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 4.18 | 3.79 – 4.56 | **< 0.001** |
| Group size | 0.06 | -0.06 – 0.17 | 0.328 |
| Freeriding (FR)† | -0.21 | -0.57 – 0.16 | 0.267 |
| Legitimate Elite†† | 0.13 | -0.19 – 0.45 | 0.419 |
| Illegitimate Elite†† | 0.25 | -0.09 – 0.58 | 0.148 |
| Legit. X FR | 0.07 | -0.39 – 0.54 | 0.754 |
| Illegit. X FR | -0.12 | -0.59 – 0.36 | 0.625 |
| **Random Effects** | | | |
| $\sigma^2$ | 1.21 | | |
| $\tau_{00\ \mathrm{Subject}}$ | 2.31 | | |
| $\tau_{11\ \mathrm{Subject.Freeriding}}$ | 0.44 | | |
| $\rho_{01\ \mathrm{Subject}}$ | -0.24 | | |
| ICC | 0.65 | | |
| $N_{\mathrm{Subject}}$ | 91 | | |
| Observations | 519 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.013 / 0.659 | | |

*Note.* † Reference group = No freeriding; Treatment group = Freeriding

†† Reference group = No Elite

***Number belief***

An initial model indicated potentially influential cases (Cooks Distance > 0.4). Removing these cases and retesting the model addressed the outlier issue and yielded the same model effects. I therefore report the original model. This model showed a significant effect of

Freeriding; Number belief was significantly lower when Freeriding was possible compared to when it was not (***Table 2.8***). Post-hoc tests showed that this overall effect was driven primarily by the difference between Freeriding conditions in the No Elite condition (***Table 2.9, Figure 2.4***).

**Table 2.8.** *Linear mixed effects model predicting Number belief*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 56.69 | 51.25 – 62.13 | **< 0.001** |
| Group size | 0.60 | -0.72 – 1.93 | 0.372 |
| Freeriding (FR)† | -6.97 | -13.57 – -0.38 | **0.038** |
| Legitimate Elite†† | -4.04 | -9.12 – 1.04 | 0.119 |
| Illegitimate Elite†† | 2.89 | -2.47 – 8.24 | 0.291 |
| Legit. x FR | 4.84 | -2.56 – 12.23 | 0.200 |
| Illegit. x FR | 1.66 | -5.92 – 9.25 | 0.667 |
| **Random Effects** | | | |
| $\sigma^2$ | 305.39 | | |
| $\tau_{00\ \text{Subject}}$ | 395.66 | | |
| $\tau_{11\ \text{Subject.Freeriding}}$ | 345.32 | | |
| $\rho_{01\ \text{Subject}}$ | -0.51 | | |
| ICC | 0.55 | | |
| $N_{\text{Subject}}$ | 91 | | |
| Observations | 519 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.020 / 0.562 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.9.** *Post-hoc contrasts of Number belief between Freeriding conditions for each Elite condition*

| *Contrast* | *Framing* | *Estimate* | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| No FR - FR | No Elite | 6.974 | 3.385 | 256.775 | 2.060 | **0.040** |
| No FR - FR | Legit. Elite | 2.135 | 3.260 | 234.926 | 0.655 | 0.513 |
| No FR - FR | Illegit. Elite | 5.310 | 3.375 | 255.045 | 1.573 | 0.117 |

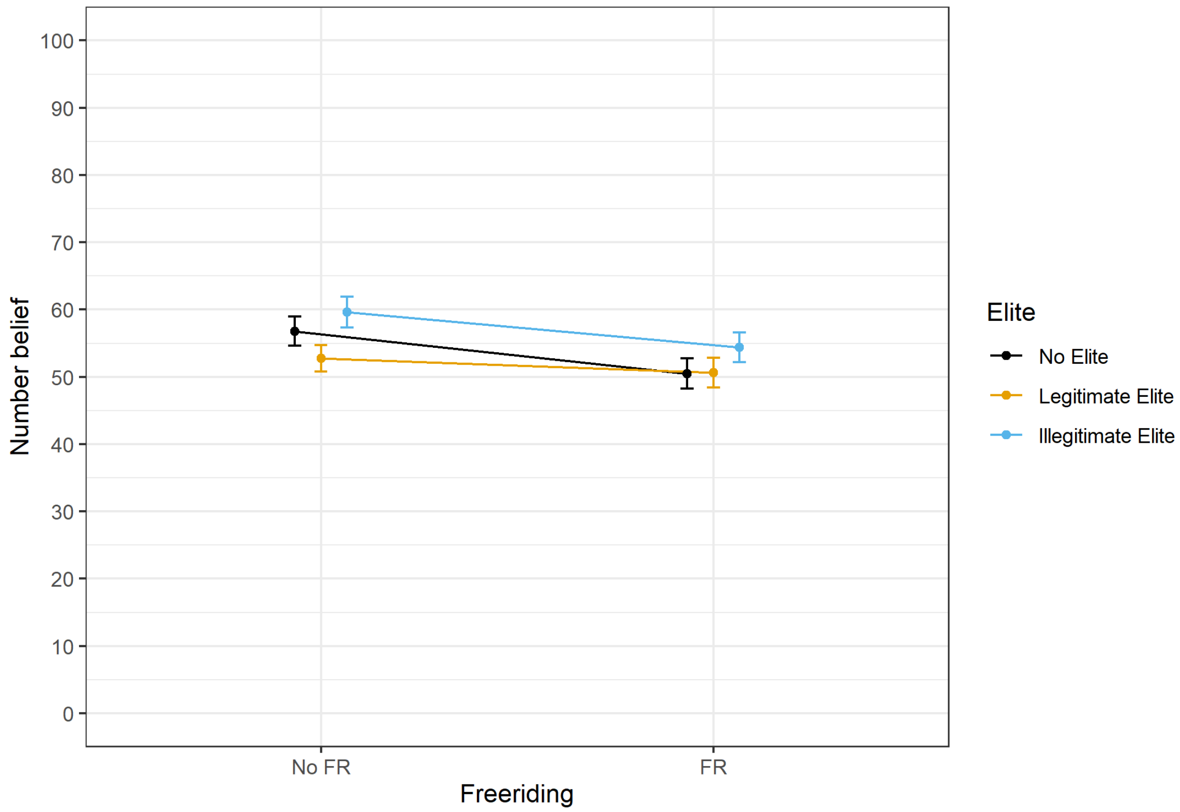


*Figure 2.4.* Mean ratings of Number belief in each Freeriding and Elite condition. Error bars represent within-subject standard error of the mean.

## Influence of subjective experiences on System Challenge decisions

As in Study 1, I was interested to explore how the subjective feelings of Identification, Fairness, and Number belief influenced decisions. To this end, in separate models I added each predictor to the above base model, and also tested the interaction of each with Freeriding and Elite.

### *Identification*

The modelling procedure revealed a significant main effect of Identification and significant interactions of Identification with each of Freeriding and Elite. The Freeriding x Identification model was a better fit than the Identification main effect model, $\chi^2$ (1, 91) = 7.01, $p$ = .008, as was the Elite x Identification model, $\chi^2$ (2, 91) = 6.20, $p$ = .045, and the main effect model was itself a better fit than the base model, $\chi^2$ (1, 91) = 64.45, $p$ < .001. ***Table 2.10*** shows the main effect and Freeriding interaction models. ***Table 2.11*** shows the main effect and Elite interaction models. I next tested whether a three-way Freeriding x Elite x Identification interaction existed by combining the appropriate terms into one model. The resultant model (***Table 2.12***) was a better fit than either interaction model alone (versus Freeriding interaction model: $\chi^2$ (4, 91) = 15.64, $p$ = .004; versus Elite interaction model: $\chi^2$ (3, 91) = 16.44, $p$ = .001). ***Figure 2.5*** shows the final model with the three-way interaction.

Post-hoc tests revealed that Identification was a positive predictor of System Challenge in all scenarios expect the No freeriding, No Elite condition (***Table 2.13***). Put differently, Freeriding increased the influence of Identification only in No Elite conditions (***Table 2.14***). This pattern of results suggests that when Freeriding was not possible, Identification influenced decisions only in situations with salient class differences. In contrast, when Freeriding was possible, Identification was equally influential regardless of the presence or type of Elite.

**Table 2.10.** *GLMMs predicting System Challenge from Freeriding, Elite, and Identification and its interaction with Freeriding*

| | Main effect model | | | Interaction model | | |
|---|---|---|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* | *Odds Ratios* | *CI (95%)* | *p* |
| Intercept | 0.08 | 0.01 – 0.70 | **0.023** | 0.33 | 0.03 – 3.35 | 0.346 |
| Group size | 1.13 | 0.95 – 1.35 | 0.167 | 1.17 | 0.98 – 1.39 | 0.090 |
| Risk preference | 1.06 | 0.91 – 1.24 | 0.451 | 1.01 | 0.86 – 1.19 | 0.908 |
| Freeriding† | 0.26 | 0.10 – 0.65 | **0.004** | 0.03 | 0.00 – 0.23 | **0.001** |
| Legitimate elite†† | 0.70 | 0.30 – 1.66 | 0.424 | 0.74 | 0.34 – 1.64 | 0.460 |
| Illegitimate elite†† | 0.54 | 0.22 – 1.33 | 0.179 | 0.62 | 0.27 – 1.42 | 0.257 |
| Legit. x Freeriding | 2.12 | 1.72 – 2.62 | **< 0.001** | 1.71 | 1.35 – 2.15 | **< 0.001** |
| Illegit. x Freeriding | 3.25 | 0.97 – 10.89 | 0.056 | 3.65 | 1.04 – 12.74 | **0.043** |
| Identification | 5.49 | 1.55 – 19.40 | **0.008** | 6.20 | 1.69 – 22.75 | **0.006** |
| Freeriding x Identification | | | | 1.71 | 1.12 – 2.61 | **0.014** |
| **Random Effects** | | | | | | |
| $\sigma^2$ | 3.29 | | | 3.29 | | |
| $\tau_{00}$ | 3.77 $_{\text{Subject}}$ | | | 2.45 $_{\text{Subject}}$ | | |
| $\tau_{11}$ | 1.11 $_{\text{Subject.Freeriding}}$ | | | 1.10 $_{\text{Subject. Freeriding}}$ | | |
| $\rho_{01}$ | -0.65 $_{\text{Subject}}$ | | | 0.12 $_{\text{Subject}}$ | | |
| ICC | 0.48 | | | 0.49 | | |
| N | 91 $_{\text{Subject}}$ | | | 91 $_{\text{Subject}}$ | | |
| Observations | 519 | | | 519 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.279 / 0.622 | | | 0.309 / 0.650 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.11**. *GLMMs predicting System Challenge from Freeriding, Elite, and Identification and its interaction with Elite*

| | Main effect model | | | Interaction model | | |
|---|---|---|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* | *Odds Ratios* | *CI (95%)* | *p* |
| Intercept | 0.08 | 0.01 – 0.70 | **0.023** | 0.21 | 0.02 – 2.24 | 0.196 |
| Group size | 1.13 | 0.95 – 1.35 | 0.167 | 1.14 | 0.95 – 1.37 | 0.147 |
| Risk preference | 1.06 | 0.91 – 1.24 | 0.451 | 1.06 | 0.90 – 1.24 | 0.509 |
| Freeriding† | 0.26 | 0.10 – 0.65 | **0.004** | 0.27 | 0.11 – 0.67 | **0.004** |
| Legitimate elite†† | 0.70 | 0.30 – 1.66 | 0.424 | 0.13 | 0.02 – 0.66 | **0.015** |
| Illegitimate elite†† | 0.54 | 0.22 – 1.33 | 0.179 | 0.17 | 0.03 – 0.94 | **0.043** |
| Legit. x Freeriding | 2.12 | 1.72 – 2.62 | < **0.001** | 1.68 | 1.27 – 2.21 | < **0.001** |
| Illegit. x Freeriding | 3.25 | 0.97 – 10.89 | 0.056 | 3.10 | 0.92 – 10.45 | 0.068 |
| Identification | 5.49 | 1.55 – 19.40 | **0.008** | 5.35 | 1.54 – 18.64 | **0.008** |
| Legit. x Identification | | | | 1.56 | 1.08 – 2.24 | **0.017** |
| Illegit. x Identification | | | | 1.35 | 0.94 – 1.93 | 0.107 |
| **Random Effects** | | | | | | |
| $\sigma^2$ | 3.29 | | | 3.29 | | |
| $\tau_{00}$ | 3.77 Subject | | | 3.60 Subject | | |
| $\tau_{11}$ | 1.11 Subject.Freeriding | | | 1.00 Subject.Freeriding | | |
| $\rho_{01}$ | -0.65 Subject | | | -0.58 Subject | | |
| ICC | 0.48 | | | 0.48 | | |
| N | 91 Subject | | | 91 Subject | | |
| Observations | 519 | | | 519 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.279 / 0.622 | | | 0.295 / 0.632 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.12.** *Final Identification model predicting System Challenge from GVI, Elite, Identification, and their three-way interaction*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 2.34 | 0.17 – 32.85 | 0.529 |
| Group size | 1.20 | 1.00 – 1.45 | 0.053 |
| Risk preference | 1.00 | 0.84 – 1.18 | 0.955 |
| Freeriding† | 0.00 | 0.00 – 0.04 | **< 0.001** |
| Legitimate elite†† | 0.02 | 0.00 – 0.20 | **0.001** |
| Illegitimate elite†† | 0.05 | 0.01 – 0.47 | **0.009** |
| Identification | 1.11 | 0.81 – 1.53 | 0.504 |
| Legit. x Freeriding | 477.23 | 11.20 – 20341.45 | **0.001** |
| Illegit. x Freeriding | 301.27 | 7.32 – 12400.05 | **0.003** |
| Freeriding x Identification | 3.45 | 1.69 – 7.08 | **0.001** |
| Legit. x Identification | 2.42 | 1.44 – 4.06 | **0.001** |
| Illegit. x Identification | 1.86 | 1.13 – 3.06 | **0.014** |
| Legit. x Freeriding x Identification | 0.30 | 0.13 – 0.71 | **0.006** |
| Illegit. x Freeriding x Identification | 0.39 | 0.17 – 0.89 | **0.026** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \mathrm{Subject}}$ | 2.64 | | |
| $\tau_{11\ \mathrm{Subject.Freeriding}}$ | 1.34 | | |
| $\rho_{01\ \mathrm{Subject}}$ | 0.12 | | |
| ICC | 0.52 | | |
| $N_{\mathrm{Subject}}$ | 91 | | |
| Observations | 519 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.344 / 0.684 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.13.** *Post-hoc tests of the trend of Identification in each Elite condition and each Freeriding scenario*

| *Freeriding* | *Elite* | *Identification trend* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| No FR | No Elite | 0.108 | 0.161 | Inf | 0.669 | .504 |
| FR | No Elite | 1.347 | 0.334 | Inf | 4.030 | **<0.001** |
| No FR | Legit. Elite | 0.991 | 0.226 | Inf | 4.381 | < **0.001** |
| FR | Legit. Elite | 1.036 | 0.249 | Inf | 4.163 | < **0.001** |
| No FR | Illegit. Elite | 0.728 | 0.211 | Inf | 3.447 | **0.001** |
| FR | Illegit. Elite | 1.028 | 0.259 | Inf | 3.973 | < **0.001** |

**Table 2.14.** *Post-hoc tests comparing the influence of Identification between Freeriding scenarios in each Elite condition*

| *Contrast* | *Elite* | *Estimate* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| No FR - FR | No Elite | -1.240 | 0.366 | Inf | -3.386 | **0.001** |
| No FR - FR | Legitimate Elite | -0.046 | 0.326 | Inf | -0.140 | 0.889 |
| No FR - FR | Illegitimate Elite | -0.300 | 0.321 | Inf | -0.934 | 0.350 |

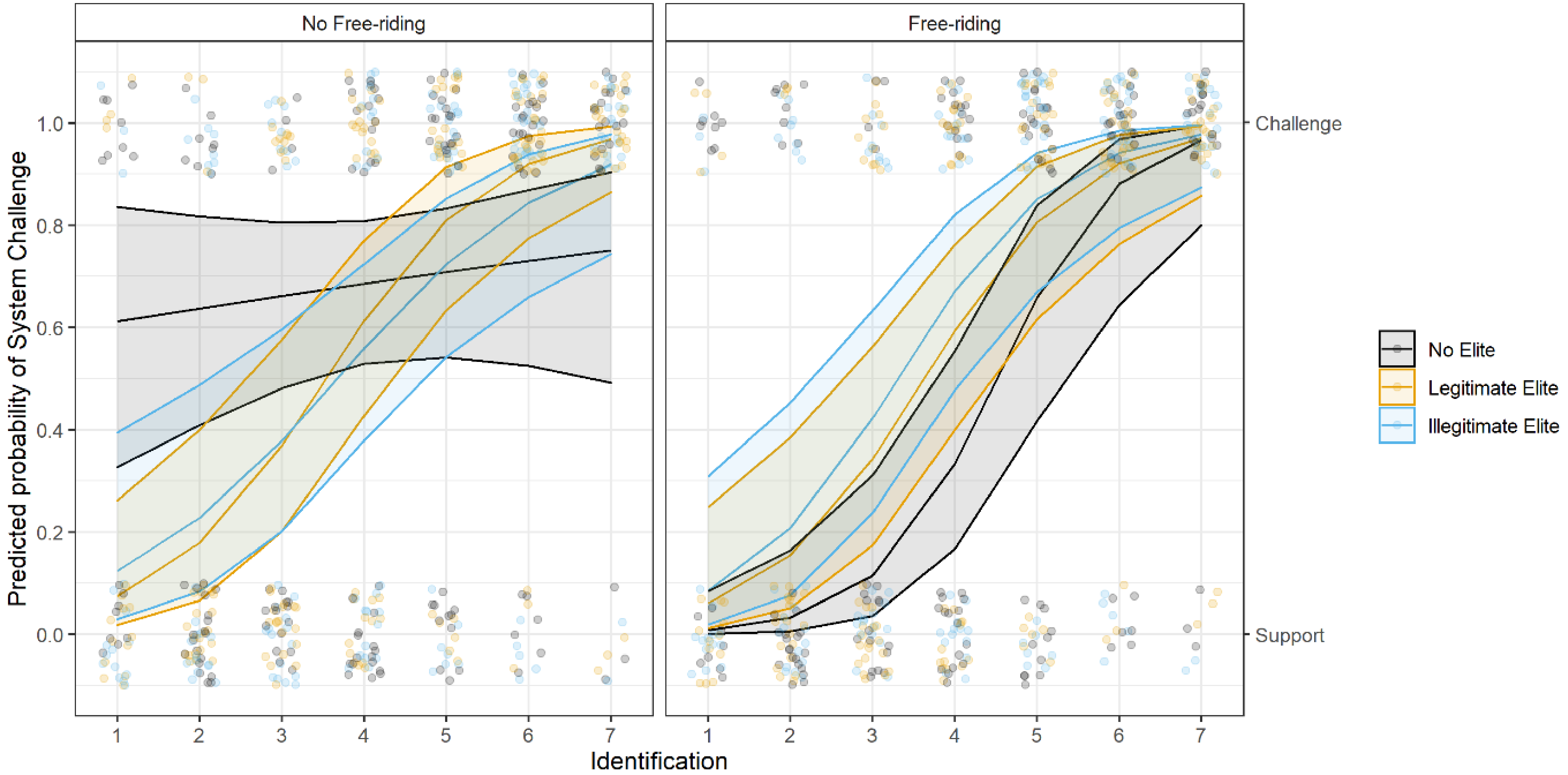


*Figure 2.5.* Model-derived predicted probabilities depicting the three-way interaction of Identification, Freeriding, and Elite predicting System Challenge decisions. Shaded areas represent the 95% confidence interval. Data points (jittered) correspond to the right vertical axis and represent raw decisions.

### *Number belief*

The best-fitting Number belief model showed a significant main effect of Number belief in addition to the observed significant effects of Freeriding and the Elite x Freeriding interactions (***Table 2.15***). This model was a better fit than the base model, $\chi^2$ (1, 91) = 126.37, $p$ < .001. Adding a Freeriding x Number belief interaction term did not improve the model fit, $\chi^2$ (1, 91) = 0.30, $p$ = .59, and neither did adding an Elite x Number belief interaction, $\chi^2$ (2, 91) = 0.51, $p$ = .78.

**Table 2.15.** *Final Number belief model predicting System Challenge from GVI, Elite, and Number belief*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.06 | 0.00 – 0.73 | **0.028** |
| Group size | 1.20 | 0.97 – 1.48 | 0.098 |
| Risk preference | 1.00 | 0.83 – 1.22 | 0.965 |
| Freeriding† | 0.27 | 0.10 – 0.73 | **0.010** |
| Legitimate elite†† | 1.00 | 0.41 – 2.45 | 0.999 |
| Illegitimate elite†† | 0.47 | 0.18 – 1.21 | 0.119 |
| Number belief | 1.08 | 1.06 – 1.10 | **< 0.001** |
| Legit. x Freeriding | 4.24 | 1.08 – 16.58 | **0.038** |
| Illegit. x Freeriding | 7.56 | 1.82 – 31.43 | **0.005** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 2.73 | | |
| $\tau_{11\ \text{Subject.Freeriding}}$ | 0.99 | | |
| $\rho_{01\ \text{Subject}}$ | 0.68 | | |
| ICC | 0.57 | | |
| $N_{\text{Subject}}$ | 91 | | |
| Observations | 519 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.380 / 0.733 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

### *Fairness*

A model including Fairness was no better fit than the base model, $\chi^2$ (1, 91) = 0.017, $p$ = .90.

## Interim Summary

In Experiment 1 I investigated how the possibility to freeride influenced subjective experiences of fairness, social identity, and group efficacy, and the likelihood of challenging the system amongst individuals disadvantaged by a social system. I did this by comparing measurements across two types of artificial unfair social system that were identical in all ways expect for the payoff structure; in No Freeriding scenarios participants could only improve their payoff by taking a personal risk and challenging the system, whereas in Freeriding scenarios it was possible for a participants' payoff to improve even if they didn't take a personal risk. I found that the possibility of freeriding decreased the likelihood that participants would challenge the system, but only in the absence of social inequality. In the presence of social inequality, decisions to challenge were no more or less likely when freeriding was possible compared to when freeriding was not possible. These results suggest that when social inequality is apparent, the deleterious impact of freeriding on cooperation is neutralised. Feelings of identification amongst disadvantaged participants positively predicted decisions to challenge in all scenarios except when there was no social inequality and it was not possible to freeride, suggesting that social identity was less relevant to decisions when there were no social boundaries and interpersonal trust was less important for collective and individual outcomes. Feelings of group efficacy (i.e., Number belief) were lower when freeriding was possible, compared to when it was not possible.

# Experiment 2

Experiment 2 aimed to elaborate on the findings of Experiment 1 by investigating how the opportunity to communicate intentions modified the decision process. Since communication is commonly found to improve cooperation in social dilemmas (e.g., Sally, 1995), it was expected that overall the opportunity to coordinate responses by communicating intentions would improve cooperation, even when there was an incentive to freeride. Beyond cooperation rates, in Experiment 2 I also sought to explore how the addition of communication influenced the subjective ratings provided by participants, and how these in turn affected decisions.

## Method

### Participants

Participants were the same sample that took part in Study 1, Experiment 2. Excluding the Elite and displaced Elites this amounted to 87 participants. Of these, 54 (62%) were female, 11 (13%) were male, 2 (2%) preferred not to say, one (1%) was 'other' and 19 (22%) were not defined (because the experimenter erroneously omitted the gender questionnaire for some participants). Of the 76 participants who provided the information, 71 (93%) were Undergraduates and of these 64 (90%) were Psychology students.

### Materials and Procedure

As was the case in Study 1, Experiment 2 of the present study was nearly identical to Experiment 1 but differed with the addition of an 'intention phase' before a final decision was made. As in the group games in Study 1, the intention phase presented the scenario in the same way as in Experiment 1 but, instead of making a decision immediately, participants were asked to virtually communicate an intention by selecting 'I intend to Pay 50' or 'I intend to Pay 0'. It was made clear to participants that their expressed intentions were non-binding, and they could change their mind in the decision phase if they wanted to. Participants formed these intentions for each of the six games before continuing to the decision phase where, for each game, the scenario was presented again along with the intentions communicated. Intentions were presented in two sentences – one describing how many participants expressed an intention to pay 50, and one describing how many participants expressed an intention to pay 0. The participants' own intention was made clear in these sentences; depending on their intention, one sentence would read "You and [x] others expressed an intention to [pay 50/pay0]", whilst the other sentence read "[x] others expressed an intention to [pay 50/pay 0]".

## Results

### Computed variables

In addition to the variables computed in Experiment 1, I also created the variable "Others' intentions". This variable represented the number of other group members expressing an intention to challenge the system as a proportion of the total number of Group members present. This provided a value ranging from 0 (indicating that *no* others intended to challenge) to 1 (indicating that *all* others intended to challenge). As was the case for Number belief, this value was multiplied by 100 so that an increment of 1 corresponded to a proportion change of 1% (this was done so that odds ratios were more readily interpretable and so that the scale for Other's intentions was identical to that of Number belief).

### System challenge decisions

As in Experiment 1, I first examined how the Freeriding and Elite manipulations influenced the likelihood of challenging the system by creating generalised linear mixed models (GLMMs) predicting the binomial variable "System Challenge", controlling for Risk preference and Group size. The main effects of Freeriding and Elite were both non-significant.

However, there was a significant Freeriding x Elite interaction whereby System Challenge was significantly more likely when the Legitimate Elite was present compared to the No Elite condition in scenarios where Freeriding was possible (*OR* = 3.88, *z* = 2.17, *p* = 0.030, ***Table 2.16***). Post-hoc tests indicated that there was as significant difference in the likelihood of challenging between the No Elite and Legitimate Elite conditions when Freeriding was possible but not when Freeriding was not possible (***Table 2.17***). Inspection of ***Figure 2.6*** shows that this result was due to a relative decrease in odds between No Freeriding and Freeriding conditions in the No Elite condition combined with a relative increase in odds between No Freeriding and Freeriding conditions in the Legitimate Elite condition, though neither of these effects were themselves significant (***Table 2.18***).

**Table 2.16.** *Generalised linear mixed effects model predicting System Challenge from Freeriding scenarios and Elite conditions*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 7.46 | 2.72 – 20.44 | **< 0.001** |
| Risk preference | 1.32 | 1.06 – 1.63 | **0.012** |
| Group size | 0.93 | 0.76 – 1.14 | 0.460 |
| Freeriding (FR)† | 0.60 | 0.24 – 1.47 | 0.265 |
| Legitimate Elite†† | 1.00 | 0.45 – 2.23 | 1.000 |
| Illegitimate Elite†† | 0.72 | 0.33 – 1.60 | 0.422 |
| Legit. x FR | 3.88 | 1.14 – 13.25 | **0.030** |
| Illegit. x FR | 2.89 | 0.90 – 9.32 | 0.076 |

**Random Effects**

| | |
|---|---|
| $\sigma^2$ | 3.29 |
| $\tau_{00\ \text{Subject}}$ | 2.84 |

| | |
|---|---|
| $\tau_{11}$ Subject.Freeriding | 0.54 |
| $\rho_{01}$ Subject | 0.65 |
| ICC | 0.54 |
| N Subject | 87 |
| Observations | 522 |
| Marginal $R^2$ / Conditional $R^2$ | 0.090 / 0.584 |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.17.** *Post-hoc contrasts of the likelihood of challenging the System between Elite conditions in each Freeriding condition*

| *Contrast* | *Freeriding* | *Estimate (log odds)* | *SE* | *df* | *z* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legit. Elite | No FR | -0.000 | 0.409 | Inf | -0.000 | 1.000 |
| No Elite - Illegit. Elite | No FR | 0.325 | 0.405 | Inf | 0.803 | 0.701 |
| Legit. Elite - Illegit. Elite | No FR | 0.325 | 0.405 | Inf | 0.803 | 0.701 |
| No Elite - Legit. Elite | FR | -1.356 | 0.474 | Inf | -2.860 | **0.012** |
| No Elite - Illegit. Elite | FR | -0.736 | 0.439 | Inf | -1.677 | 0.214 |
| Legit. Elite - Illegit. Elite | FR | 0.620 | 0.462 | Inf | 1.341 | 0.372 |

**Table 2.18.** *Post-hoc contrasts of the likelihood of challenging the System between Freeriding conditions in each Elite condition*

| *Contrast* | *Elite* | *Estimate (log odds)* | *SE* | *df* | *z* | *p* |
|---|---|---|---|---|---|---|
| No FR - FR | No Elite | 0.512 | 0.459 | Inf | 1.114 | 0.265 |

| | | | | | | |
|---|---|---|---|---|---|---|
| No FR - FR | Legit. Elite | -0.845 | 0.540 | Inf | -1.564 | 0.118 |
| No FR - FR | Illegit. Elite | -0.550 | 0.482 | Inf | -1.140 | 0.254 |

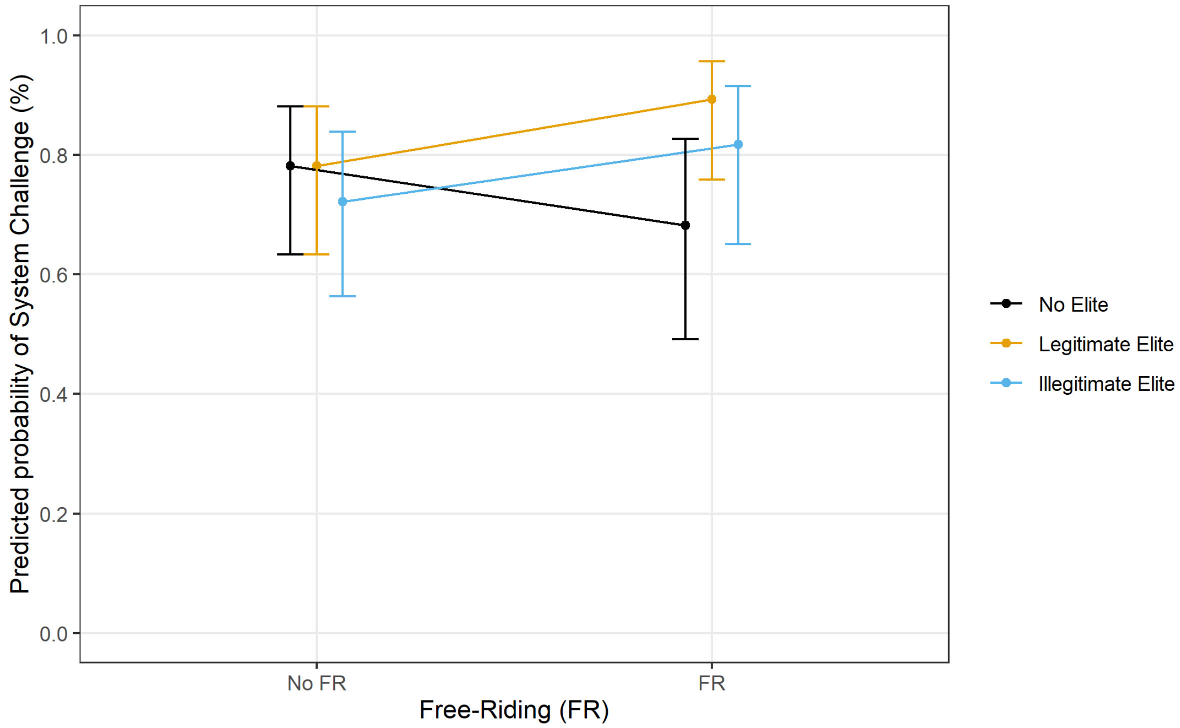


*Figure 2.6.* Predicted probability of System Challenge based on Freeriding scenario and Elite condition. Error bars represent standard error of the mean.

## Subjective experience of the System

To investigate the possible motivations behind decisions, I next examined how the Freeriding and Elite conditions influenced the ratings of how fair participants felt the systems were, how much they identified with one another within the systems, and how many other group members they believed would challenge the system.

### *Fairness*

System Fairness was not predicted by Freeriding ($b$ = 0.01, $t$(259.76) = 0.06, $p$ = .949). The Illegitimate Elite condition negatively predicted Fairness, compared to the No Elite condition ($b$ = -0.38, $t$(348) = -2.69, $p$ = 0.007; ***Table 2.19***). There was also a significant

Legitimate Elite x Freeriding interaction; systems with the Legitimate Elite were rated as fairer when Freeriding was possible compared to when it was not (***Table 2.20; Figure 2.7***). Further post-hoc tests showed that under conditions of No Freeriding systems with No Elite were rated as fairer than systems with an Illegitimate Elite, whilst when Freeriding was possible Systems with a Legitimate Elite were rated fairer than systems with an Illegitimate Elite (***Table 2.21***).

**Table 2.19.** *Linear mixed effects model predicting Fairness from Freeriding and Elite conditions*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 3.17 | 1.43 – 4.92 | **< 0.001** |
| Group size | 0.02 | -0.11 – 0.16 | 0.760 |
| Freeriding (FR)† | 0.00 | -0.31 – 0.31 | 1.000 |
| Legitimate Elite†† | -0.17 | -0.45 – 0.10 | 0.221 |
| Illegitimate Elite†† | -0.38 | -0.66 – -0.10 | **0.007** |
| Legit. x FR | 0.40 | 0.01 – 0.79 | **0.044** |
| Illegit. x FR | 0.23 | -0.16 – 0.62 | 0.249 |
| **Random Effects** | | | |
| $\sigma^2$ | 0.86 | | |
| $\tau_{00\ \text{Subject}}$ | 2.67 | | |
| $\tau_{11\ \text{Subject.Freeriding}}$ | 0.48 | | |
| $\rho_{01\ \text{Subject}}$ | -0.27 | | |
| ICC | 0.75 | | |
| $N_{\text{Subject}}$ | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.011 / 0.754 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.20.** *Post-hoc contrasts of Fairness rating between Freeriding and No Freeriding conditions in each Elite condition*

| *Contrast* | *Elite* | *Estimate* | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| No FR - FR | No Elite | -0.000 | 0.160 | 296.698 | -0.000 | 1.000 |
| No FR - FR | Legit. Elite | -0.402 | 0.160 | 296.698 | -2.511 | **0.013** |
| No FR - FR | Illegit. Elite | -0.230 | 0.160 | 296.698 | -1.435 | 0.152 |

**Table 2.21.** *Post-hoc contrasts of Fairness ratings between Elite conditions in each Freeriding condition*

| *Contrast* | *Freeriding* | *Estimate* | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legit. Elite | No FR | 0.172 | 0.142 | 352.046 | 1.216 | 0.444 |
| No Elite - Illegit. Elite | No FR | 0.379 | 0.142 | 352.046 | 2.676 | **0.021** |
| Legit. Elite - Illegit. Elite | No FR | 0.207 | 0.142 | 352.046 | 1.460 | 0.312 |
| No Elite - Legit. Elite | FR | -0.230 | 0.142 | 352.046 | -1.622 | 0.238 |
| No Elite - Illegit. Elite | FR | 0.149 | 0.142 | 352.046 | 1.054 | 0.543 |
| Legit. Elite - Illegit. Elite | FR | 0.379 | 0.142 | 352.046 | 2.676 | **0.021** |

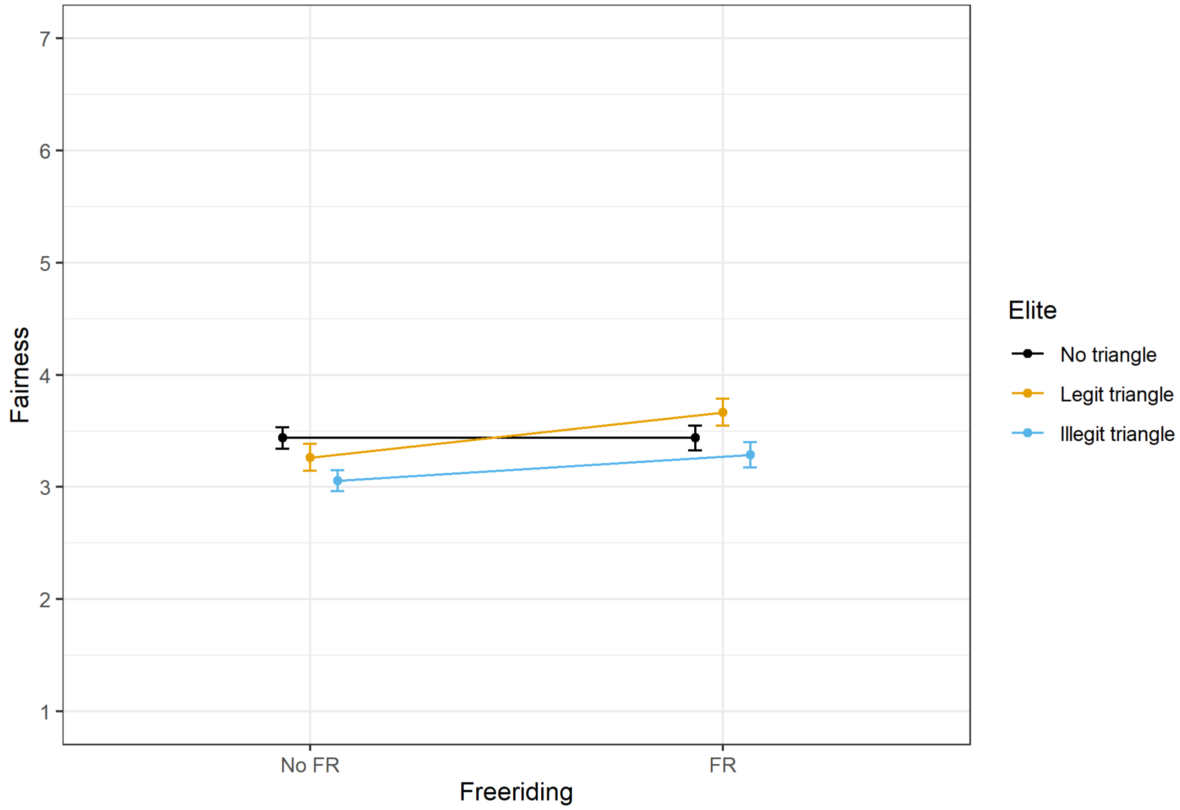


*Figure 2.7.* Ratings of system fairness as a function of Freeriding and Elite conditions. The system with a Legitimate Elite was perceived as fairer when Freeriding was possible compared to when it was not. Error bars represent within-subject standard error of the mean.

### *Identification*

Identification was not predicted by the main effects of Freeriding or Elite. However, there was a significant Freeriding x Elite interaction (***Table 2.22***). Post-hoc contrasts showed that in systems with the Legitimate Elite, Identification was higher when Freeriding was possible compared to when it was not possible (***Table 2.23, Figure 2.8***). This rendered Identification significantly higher in the Legitimate Elite system compared to the No Elite system when Freeriding was possible (***Table 2.24***).

**Table 2.22.** *Linear mixed effects model predicting Identification from Freeriding and Elite conditions*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 4.89 | 3.73 – 6.04 | **< 0.001** |
| Group size | -0.01 | -0.10 – 0.08 | 0.805 |
| Freeriding (FR)† | -0.17 | -0.56 – 0.22 | 0.388 |
| Legitimate Elite†† | 0.05 | -0.32 – 0.42 | 0.808 |
| Illegitimate Elite†† | -0.11 | -0.49 – 0.26 | 0.544 |
| Legit. x FR | 0.59 | 0.06 – 1.11 | **0.028** |
| Illegit. x FR | 0.31 | -0.21 – 0.83 | 0.246 |
| **Random Effects** | | | |
| $\sigma^2$ | 1.56 | | |
| $\tau_{00}$ Subject | 0.86 | | |
| $\tau_{11}$ Subject.Freeriding | 0.35 | | |
| $\rho_{01}$ Subject | -0.16 | | |
| ICC | 0.38 | | |
| N Subject | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.017 / 0.388 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.23.** *Post-hoc contrasts of Identification between Freeriding conditions in each Elite condition*

| *Contrast* | *Elite* | *Estimate* | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| No FR - FR | No Elite | 0.172 | 0.201 | 351.629 | 0.859 | 0.391 |
| No FR - FR | Legit. Elite | -0.414 | 0.201 | 351.629 | -2.062 | **0.040** |
| No FR - FR | Illegit. Elite | -0.138 | 0.201 | 351.629 | -0.687 | 0.492 |

**Table 2.24.** *Post-hoc contrasts of Identification between Elite conditions in each Freeriding condition*

| *Contrast* | *Freeriding* | *Estimate* | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legit. Elite | No FR | -0.046 | 0.190 | 352.049 | -0.242 | 0.968 |
| No Elite - Illegit. Elite | No FR | 0.115 | 0.190 | 352.049 | 0.604 | 0.818 |
| Legit. Elite - Illegit. Elite | No FR | 0.161 | 0.190 | 352.049 | 0.846 | 0.675 |
| No Elite - Legit. Elite | FR | -0.632 | 0.190 | 352.049 | -3.322 | **0.003** |
| No Elite - Illegit. Elite | FR | -0.195 | 0.190 | 352.049 | -1.027 | 0.560 |
| Legit. Elite - Illegit. Elite | FR | 0.437 | 0.190 | 352.049 | 2.295 | 0.058 |

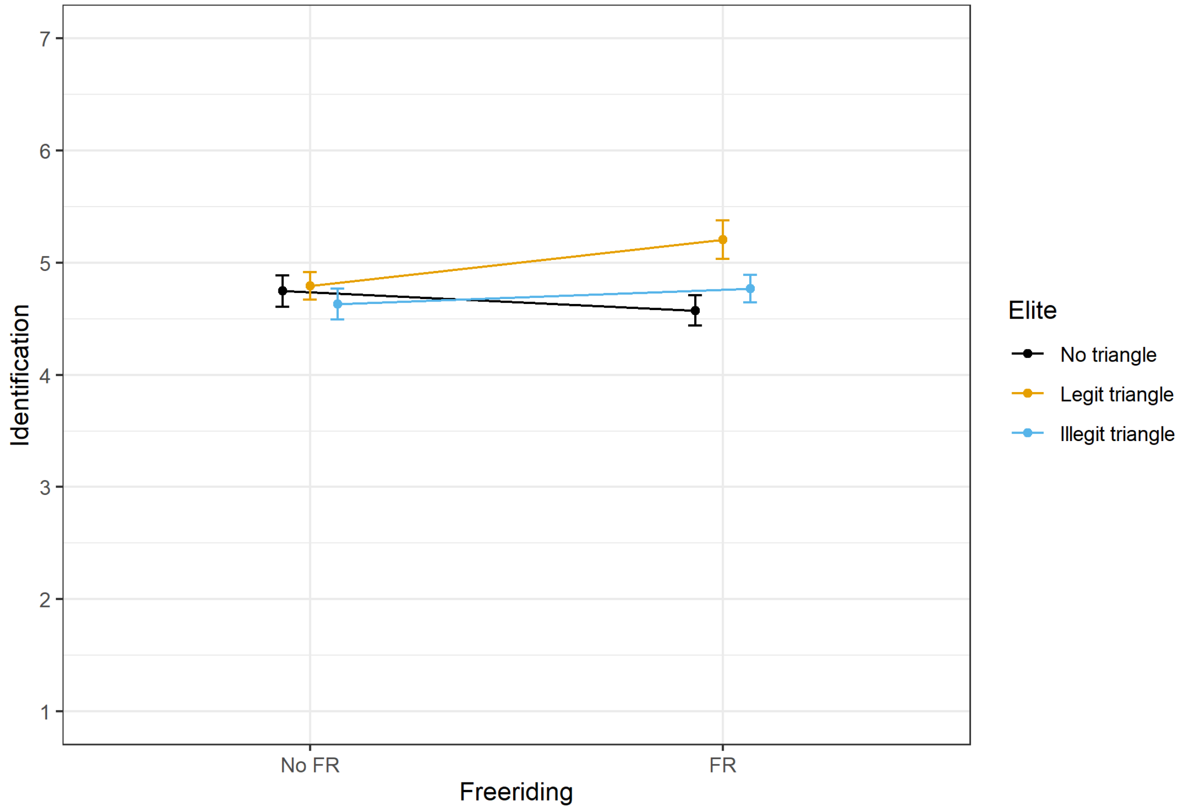


*Figure 2.8.* Ratings of identification with other non-Elites as a function of Freeriding and Elite conditions. In the Legitimate Elite system, Identification was significantly higher when Freeriding was possible compared to when it was not. Error bars represent standard error of the mean.

***Number belief***

A model predicting Number belief showed a significant Freeriding x Elite interaction (***Table 2.25***). Number belief was predicted only from the Legitimate Elite x Freeriding interaction ($b$ = 15.53, $t(348)$ = 3.21, $p$ = 0.001); in the presence of a Legitimate Elite, Number belief was higher when Freeriding was possible compared to when it was not. This was not the case for the other Elite conditions (***Table 2.26***). In the Freeriding scenario, Number belief was significantly higher in the Legitimate Elite condition compared to the No Elite ($b$ = -12.68, $t(352)$ = -3.69, $p$ = 0.001) and Illegitimate Elite ($b$ = 10.33, $t(352)$ = 3.00, $p$ = 0.008) conditions (***Table 2.27, Figure 2.9***).

**Table 2.25.** *Linear mixed effects model predicting Number belief from Freeriding and Elite conditions*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 50.16 | 32.99 – 67.33 | **< 0.001** |
| Group size | 0.86 | -0.44 – 2.17 | 0.194 |
| Freeriding (FR)† | -6.26 | -13.30 – 0.78 | 0.081 |
| Legitimate Elite†† | -2.86 | -9.56 – 3.85 | 0.404 |
| Illegitimate Elite†† | -0.95 | -7.66 – 5.75 | 0.780 |
| Legit. x FR | 15.53 | 6.05 – 25.01 | **0.001** |
| Illegit. x FR | 3.30 | -6.18 – 12.78 | 0.495 |
| **Random Effects** | | | |
| $\sigma^2$ | 508.83 | | |
| $\tau_{00}$ Subject | 108.51 | | |
| $\tau_{11}$ Subject.Freeriding | 103.88 | | |
| $\rho_{01}$ Subject | 0.42 | | |
| ICC | 0.29 | | |
| N Subject | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.028 / 0.307 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.26.** *Post-hoc contrasts of Number belief between Freeriding conditions in each Elite condition*

| *Contrast* | *Elite* | *Estimate* | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| No FR - FR | No Elite | 6.262 | 3.611 | 355.501 | 1.734 | 0.084 |
| No FR - FR | Legit. Elite | -9.273 | 3.611 | 355.501 | -2.568 | **0.011** |
| No FR - FR | Illegit. Elite | 2.962 | 3.611 | 355.501 | 0.820 | 0.413 |

**Table 2.27.** *Post-hoc contrasts of Number belief between Elite conditions in each Freeriding condition*

| *Contrast* | *Freeriding* | *Estimate* | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legit. Elite | No FR | 2.857 | 3.440 | 352.047 | 0.830 | 0.684 |
| No Elite - Illegit. Elite | No FR | 0.953 | 3.440 | 352.047 | 0.277 | 0.959 |
| Legit. Elite - Illegit. Elite | No FR | -1.903 | 3.440 | 352.047 | -0.553 | 0.845 |
| No Elite - Legit. Elite | FR | -12.678 | 3.440 | 352.047 | -3.686 | **0.001** |
| No Elite - Illegit. Elite | FR | -2.346 | 3.440 | 352.047 | -0.682 | 0.774 |
| Legit. Elite - Illegit. Elite | FR | 10.332 | 3.440 | 352.047 | 3.004 | **0.008** |

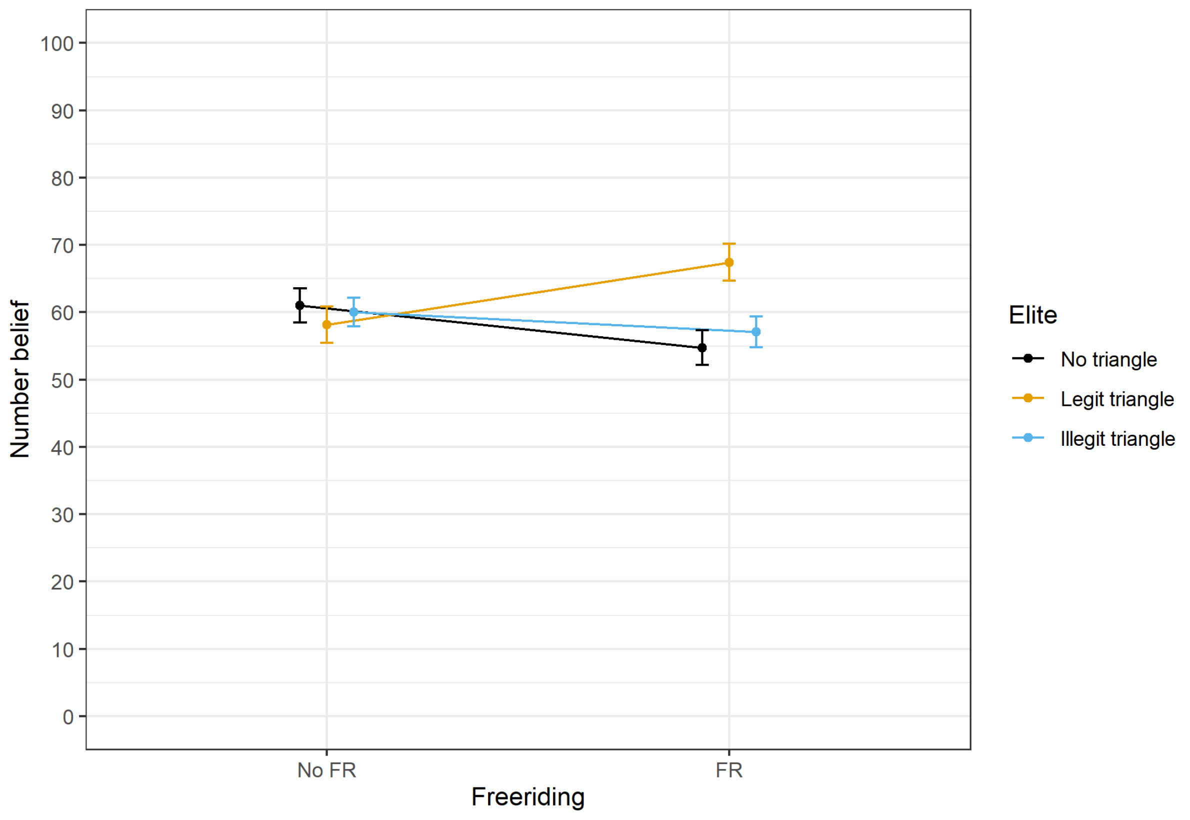


*Figure 2.9.* Belief in the number of others deciding to challenge (expressed as a proportion of total number of others) as a function of Freeriding and Elite conditions. In the Legitimate Elite system, Number belief was higher when Freeriding was possible compared to when it was not. Error bars represent standard error of the mean.

## Influence of subjective experiences on System Challenge decisions

As in previous analyses, I next explored how the subjective experiences of Number belief, Identification, and Fairness contributed to decisions to challenge the system by testing the effects of Number belief, Identification, and Fairness and their interactions with each of Freeriding and Elite in separate models.

### *Number belief*

Adding a main effect of Number belief improved model fit over the base model, $\chi^2$ (1, 87) = 68.64, $p$ < .001, and showed that Number belief had a significant positive influence on System Challenge. Adding a Number belief x Freeriding interaction further improved model fit, $\chi^2$ (1, 87) = 4.16, $p$ = .04 (***Table 2.28***). Post-hoc analysis showed that Number belief had a significant influence on decisions in both No-Freeriding and Freeriding conditions (***Table 2.29***) but the influence of Number belief was stronger when Freeriding was not possible compared to when Freeriding was possible (***Table 2.30, Figure 2.10***).

**Table 2.28.** *Generalised linear mixed effects model predicting System Challenge from Freeriding and Elite conditions and Number belief*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.83 | 0.04 – 18.75 | 0.908 |
| Risk preference | 1.37 | 1.09 – 1.72 | **0.007** |
| Group size | 0.91 | 0.73 – 1.14 | 0.415 |
| Freeriding (FR)† | 2.37 | 0.52 – 10.88 | 0.268 |
| Legitimate Elite†† | 1.36 | 0.52 – 3.59 | 0.535 |
| Illegitimate Elite†† | 0.63 | 0.25 – 1.58 | 0.321 |
| Number belief | 1.07 | 1.04 – 1.09 | **< 0.001** |
| Legit. x FR | 1.65 | 0.43 – 6.38 | 0.467 |
| Illegit. x FR | 2.86 | 0.79 – 10.36 | 0.110 |
| FR x Number belief | 0.98 | 0.95 – 1.00 | **0.046** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 4.50 | | |
| ICC | 0.58 | | |
| $N_{\text{Subject}}$ | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.271 / 0.692 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.29.** *Post-hoc analysis estimating the trend of Number belief in each Freeriding condition*

| *Freeriding* | *Trend* | *SE* | *df* | *z* | *p* |
|---|---|---|---|---|---|
| No Freeriding | 0.063 | 0.011 | Inf | 6.037 | < **.001** |
| Freeriding | 0.040 | 0.008 | Inf | 4.954 | < **.001** |

**Table 2.30.** *Post-hoc contrast of the influence of Number belief between Freeriding conditions*

| *Contrast* | *Estimate (log odds)* | *SE* | *df* | *z* | *p* |
|---|---|---|---|---|---|
| No FR - FR | 0.024 | 0.012 | Inf | 1.999 | **0.046** |

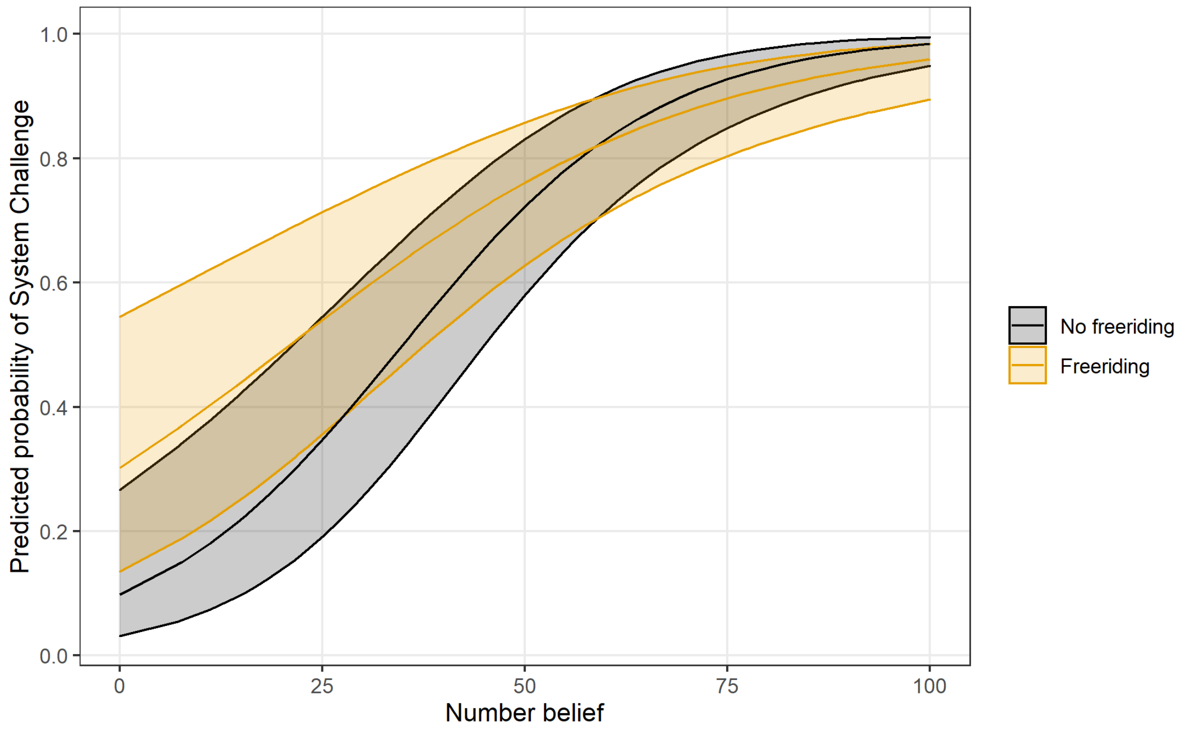


*Figure 2.10.* Predicted probability of a participant deciding to challenge the system as a function of Number belief and Freeriding condition. The influence of Number belief was stronger when Freeriding was possible. Shaded areas represent the 95% confidence interval.

### *Identification*

Adding a main effect of Identification to the base model improved the model fit, $\chi^2$ (1, 87) = 118.88, $p$ < .001, and showed that greater Identification was associated with greater likelihood of challenging the system (***Table 2.31***). Neither of the interaction models improved the fit of this model (Freeriding interaction model vs main effect model: $\chi^2$ (1, 87) = 0.14, $p$ = .70; Elite interaction model vs main effect model: $\chi^2$ (2, 87) = 1.54, $p$ = .46).

**Table 2.31.** *Generalised linear mixed effects model predicting System Challenge from Freeriding and Elite conditions and Identification*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 0.05 | 0.00 – 1.70 | 0.095 |
| Risk preference | 1.43 | 1.10 – 1.87 | **0.008** |
| Group size | 0.97 | 0.76 – 1.25 | 0.829 |
| Freeriding (FR)† | 0.53 | 0.19 – 1.46 | 0.220 |
| Legitimate elite†† | 0.92 | 0.33 – 2.56 | 0.877 |
| Illegitimate elite†† | 0.70 | 0.26 – 1.90 | 0.483 |
| Identification | 3.68 | 2.65 – 5.11 | **< 0.001** |
| Legit. x FR | 2.36 | 0.53 – 10.57 | 0.260 |
| Illegit. x FR | 2.78 | 0.68 – 11.43 | 0.155 |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 6.04 | | |
| ICC | 0.65 | | |
| $N_{\text{Subject}}$ | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.375 / 0.780 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

### *Fairness*

No models including Fairness were better fits to the data than the base model.

## Overall effect of Information

To test the overall effect of the ability to communicate intentions, I combined the datasets of Experiments 1 and 2 and created the new variable "Information", which described whether it was possible to communicate intentions (Experiment 2) or not (Experiment 1). I constructed a model predicting System Challenge from Freeriding, Elite, and this new Information variable using the combined dataset. The best-fitting model included Freeriding, Elite, Information, and a Freeriding x Elite interaction (***Table 2.32***); model fit was not improved by including Freeriding x Information, $\chi^2$ (1, 178) = 2.63, $p$ = .105, or Elite x Information, $\chi^2$ (2, 178) = 1.30, $p$ = .522, interaction terms. The final model indicated that, overall, the ability to communicate intentions increased the odds of challenging the system by 2.48 times.

**Table 2.32.** *Generalised linear mixed effects model predicting System Challenge from Freeriding, Elite, and Information*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 2.45 | 0.46 – 13.05 | 0.295 |
| Group size | 1.03 | 0.91 – 1.16 | 0.681 |
| Risk preference | 1.25 | 1.09 – 1.44 | **0.002** |
| Freeriding† | 0.36 | 0.20 – 0.64 | **0.001** |
| Legitimate elite†† | 0.88 | 0.50 – 1.55 | 0.666 |
| Illegitimate elite†† | 0.70 | 0.40 – 1.25 | 0.228 |
| Information⁎ | 2.48 | 1.24 – 4.97 | **0.010** |
| Legit. x FR | 3.27 | 1.46 – 7.35 | **0.004** |
| Illegit. x FR | 3.48 | 1.54 – 7.86 | **0.003** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00}$ Subject | 3.72 | | |
| ICC | 0.53 | | |
| N Subject | 178 | | |
| Observations | 1041 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.083 / 0.570 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding.
†† Reference group = No Elite.
⁎ Reference group = No information.

## Intentions

The intention information allowed me to explore how information was used to inform decision-making. Specifically, comparing what a participant said they would choose with their actual decision allowed some limited inferences to be made about participants' strategy. ***Table 2.33* – *Table 2.38*** provide cross-tabulated frequency data mapping intentions and decisions for each condition.

**Table 2.33.** Frequency data of intentions and decisions in the No Freeriding, No Elite condition

| | | Decision | | |
|---|---|---|---|---|
| | | Support | Challenge | Total |
| Intention | Support | 14 (16%) | 9 (10%) | 23 (26%) |
| | Challenge | 13 (15%) | 51 (59%) | 64 (74%) |
| | Total | 27 (31%) | 60 (69%) | 87 |

**Table 2.34.** Frequency data of intentions and decisions in the No Freeriding, Legitimate Elite condition

| | | Decision | | |
|---|---|---|---|---|
| | | Support | Challenge | Total |
| Intention | Support | 15 (17%) | 12 (14%) | 27 (31%) |
| | Challenge | 12 (14%) | 48 (55%) | 60 (69%) |
| | Total | 27 (31%) | 60 (69%) | 87 |

**Table 2.35.** Frequency data of intentions and decisions in the No Freeriding, Illegitimate Elite condition

| | | Decision | | |
|---|---|---|---|---|
| | | Support | Challenge | Total |
| Intention | Support | 18 (21%) | 9 (10%) | 27 (31%) |
| | Challenge | 13 (15%) | 47 (54%) | 60 (69%) |
| | Total | 31 (36%) | 56 (64%) | 87 |

**Table 2.36.** *Frequency data of intentions and decisions in the Freeriding, No Elite condition*

| | | Decision | | |
|---|---|---|---|---|
| | | Support | Challenge | Total |
| Intention | Support | 21 (24%) | 10 (12%) | 31 (36%) |
| | Challenge | 14 (16%) | 42 (48%) | 56 (64%) |
| | Total | 35 (40%) | 52 (60%) | 87 |

**Table 2.37.** *Frequency data of intentions and decisions in the Freeriding, Legitimate Elite condition*

| | | Decision | | |
|---|---|---|---|---|
| | | Support | Challenge | Total |
| Intention | Support | 5 (6%) | 4 (5%) | 9 (11%) |
| | Challenge | 16 (18%) | 62 (71%) | 78 (89%) |
| | Total | 21 (24%) | 66 (76%) | 87 |

**Table 2.38.** *Frequency data of intentions and decisions in the Freeriding, Illegitimate Elite condition*

| | | Decision | | |
|---|---|---|---|---|
| | | Support | Challenge | Total |
| Intention | Support | 15 (17%) | 11 (13%) | 26 (30%) |
| | Challenge | 12 (14%) | 49 (56%) | 61 (70%) |
| | Total | 27 (31%) | 60 (69%) | 87 |

***Figure 2.11*** visualises the proportion of decisions that resulted from communicated intentions in each Freeriding and Elite condition. For example, the top left cell shows that when Freeriding was not possible and there was No Elite, approximately 60% of participants who decided to support the system had previously expressed an intention to support (black), whilst the remaining 40% expressed an intention to challenge (orange). Likewise, approximately 80%

of participants who challenged the system had expressed an intention to challenge (orange), whilst the remaining 20% had expressed an intention to support (black). The figure shows that, in all conditions, the majority of participants made decisions that were consistent with their communicated intentions.

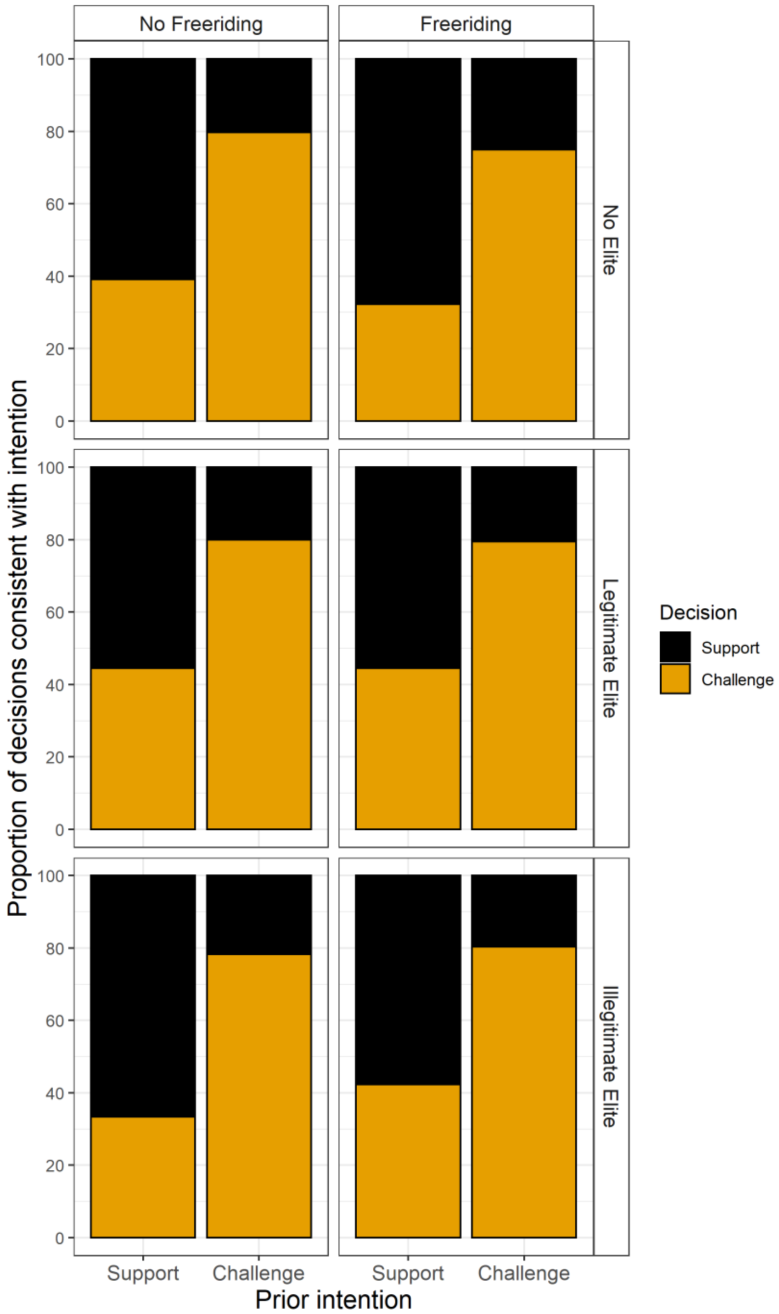


*Figure 2.11.* Proportion of decisions associated as a function of intention and Freeriding and Elite condition. Each bar represents the proportion of decisions to support (black) or challenge (orange) the system that followed a communicated intention to support or challenge (x axis).

### ***Manipulation effects on Intentions***

I tested how the possibility to freeride and the type of Elite present influenced what intention participants would communicate. The binary categorial variable "Intention" ("Don't find out" vs "Find out") describes what decision a participant communicated they intended to make before making their final decision. A model predicting participants' Intention in the communication phase showed a significant Freeriding x Elite interaction effect (***Table 2.39***). Post-hoc analysis showed, consistent with the decision results above, that intentions to challenge were highest when Freeriding was possible and the Legitimate Elite was in charge of the system (***Table 2.40*** and ***Table 2.41***). The interaction effect was driven by a significant difference in means between the Legitimate Elite and No Elite conditions (*estimate* = -3.72, *z ratio* = -3.84, $p < .001$) and the Legitimate and Illegitimate Elite conditions (*estimate* = 3.09, *z ratio* = 3.45, $p = 0.002$) when Freeriding was possible (***Table 2.40; Figure 2.12***). As can be seen in ***Figure 2.12***, there was considerably more variability in the predicted probability of communicated *intentions* to Challenge compared to decisions to Challenge (***Figure 2.6*** above).

**Table 2.39.** *Generalised linear mixed effects model predicting Intention to challenge from Freeriding and Elite conditions*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 36.04 | 0.57 – 2270.26 | 0.090 |
| Group size | 0.95 | 0.70 – 1.28 | 0.713 |
| Risk preference | 1.29 | 0.96 – 1.73 | 0.089 |
| Freeriding (FR)† | 0.71 | 0.11 – 4.43 | 0.716 |
| Legitimate elite†† | 0.64 | 0.25 – 1.63 | 0.345 |
| Illegitimate elite†† | 0.64 | 0.25 – 1.63 | 0.345 |
| Legit. x FR | 64.83 | 7.80 – 538.98 | **< 0.001** |
| Illegit. x FR | 2.96 | 0.74 – 11.82 | 0.124 |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 7.37 | | |
| $\tau_{11\ \text{Subject.Freeriding}}$ | 6.81 | | |
| $\rho_{01\ \text{Subject}}$ | -0.03 | | |
| ICC | 0.76 | | |
| $N_{\text{Subject}}$ | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.141 / 0.796 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.40.** *Post hoc tests of the model predicting Intention to Challenge contrasting the effect of the Elite conditions in each Freeriding scenario*

| *Contrast* | *Freeriding* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legitimate Elite | No FR | 0.454 | 0.480 | Inf | 0.944 | 0.612 |
| No Elite - Illegitimate Elite | No FR | 0.454 | 0.480 | Inf | 0.944 | 0.612 |
| Legitimate Elite - Illegitimate Elite | No FR | -0.000 | 0.467 | Inf | -0.000 | 1.000 |
| No Elite - Legitimate Elite | FR | -3.718 | 0.969 | Inf | -3.839 | < **0.001** |
| No Elite - Illegitimate Elite | FR | -0.633 | 0.517 | Inf | -1.225 | 0.439 |
| Legitimate Elite - Illegitimate Elite | FR | 3.085 | 0.895 | Inf | 3.448 | **0.002** |

**Table 2.41.** *Post hoc tests of the model predicting Intention to Challenge contrasting the effect of Freeriding in each Elite condition*

| *Contrast* | *Elite* | *Estimate (log odds)* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| No FR - FR | No Elite | 0.339 | 0.933 | Inf | 0.364 | 0.716 |
| No FR - FR | Legitimate Elite | -3.833 | 1.537 | Inf | -2.494 | **0.013** |
| No FR - FR | Illegitimate Elite | -0.748 | 0.990 | Inf | -0.755 | 0.450 |

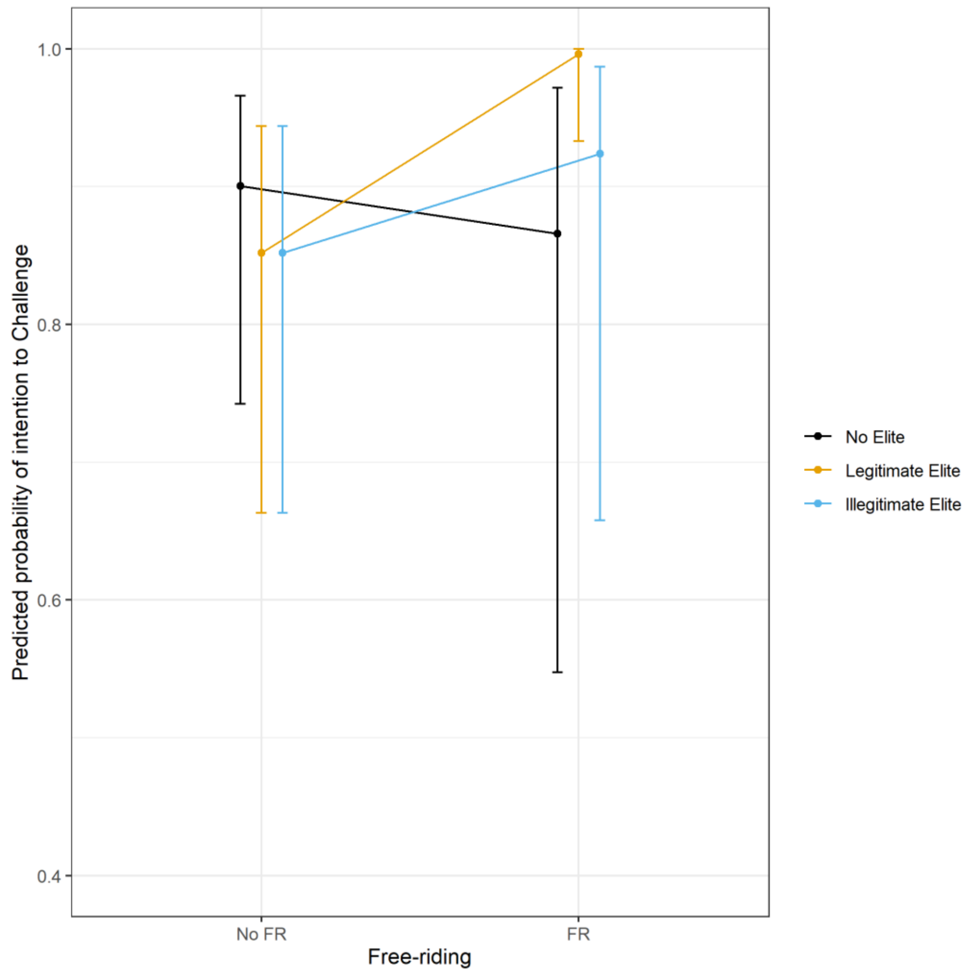


*Figure 2.12.* Predicted probability of expressing an intention to Challenge the System as a function of whether Freeriding was possible and the Elite condition. Error bars represent 95% confidence intervals.

### ***Influence of Intentions on Decisions***

I next estimated the extent to which a participant's intention predicted their decision by specifying a model including the three-way interaction of Intention, Elite, and Freeriding and all their lower-level combinations. In this model, only the main effects of Risk preference and Intention were significant (***Table 2.42***). Participants who had expressed an intention to challenge the system were seven-and-a-half times more likely to challenge the system compared to participants who had expressed an intention to support the system. Having expressed an intention to challenge, the probability of deciding to challenge was approximately 87% (*95% CI* = [79, 92]). In contrast, having expressed an intention to support, the probability of deciding to challenge was approximately 41% (*95% CI* = [25, 58]). The absence of any interaction effects suggests that there were no differences in the extent to which intentions predicted decisions between conditions. That is, it appeared that individuals were no more or less likely to follow up on an intention in any specific condition compared to the others. The finding that participants were not less likely to follow through with their intention when freeriding was possible suggests that they were not pursuing a strategy of deliberately misleading others to challenge and avoiding challenging themselves.

**Table 2.42.** *Generalised linear mixed effects model predicting System Challenge from Freeriding and Elite conditions and Intention*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| (Intercept) | 4.02 | 0.25 – 65.03 | 0.327 |
| Risk preference | 1.32 | 1.08 – 1.61 | **0.006** |
| Group size | 0.94 | 0.78 – 1.14 | 0.544 |
| Freeriding (FR)† | 0.49 | 0.10 – 2.34 | 0.372 |
| Legitimate Elite†† | 1.14 | 0.25 – 5.26 | 0.868 |
| Illegitimate Elite†† | 0.46 | 0.09 – 2.27 | 0.343 |
| Intention* | 7.51 | 1.79 – 31.53 | **0.006** |
| FR x Legit. Elite | 1.22 | 0.09 – 17.00 | 0.885 |
| FR x Illegit. Elite | 4.88 | 0.57 – 42.15 | 0.149 |
| FR x Intention | 1.30 | 0.19 – 8.83 | 0.786 |
| Legit. Elite x Intention | 0.95 | 0.14 – 6.32 | 0.959 |
| Illegit. Elite x Intention | 2.03 | 0.30 – 13.96 | 0.471 |
| FR x Legit. Elite x Intention | 1.49 | 0.07 – 31.59 | 0.797 |
| FR x Illegit. Elite x Intention | 0.37 | 0.03 – 5.11 | 0.455 |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Subject}}$ | 3.16 | | |
| ICC | 0.49 | | |
| $N_{\text{Subject}}$ | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.217 / 0.600 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding.
†† Reference group = No Elite.
* Reference group = Intention to support; Treatment group = Intention to challenge.

***The use of intention information***

I next investigated how communicated intentions in each decision scenario translated into beliefs about how many others actually challenged. I did this by adding the variable "Others' intentions" – which describes, on the same scale as Number belief, how many others expressed an intention to challenge (see 'Computed variables' at beginning of this results section) – to the model predicting Number belief from the manipulations terms and allowing it to interact with both Freeriding and Elite. This new model was a better fit than the base model, $\chi^2$ (6, 87) = 85.85, $p$ < .001. However, model diagnostics indicated that the assumption of normality of residuals was violated and there was possible heteroscedasticity I therefore tested a robust model using the package '*robustlmm*'. This model revealed a significant three-way interaction (***Table 2.43***). Post hoc tests showed that in all conditions Others' intentions was a significant positive predictor of Number belief (***Table 2.44***). However, intentions were significantly less believed when Freeriding was possible in the Legitimate condition compared to when Freeriding was not possible in the Legitimate condition ($b$ = 0.55, $t$ = 2.45, $p$ = .015; ***Table 2.45, Figure 2.13***). Both when Freeriding was and was not possible, there were no differences in the influence of Others' intentions between Elite conditions (***Table 2.46***). ***Figure 2.14*** compares Others' intentions to Number belief in each Freeriding and Elite condition.

**Table 2.43.** *Linear mixed effects model predicting Number belief from Freeriding and Elite conditions and Others' intentions within these conditions*

| | Standard model | | | Robust model | | |
|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI (95%)* | *p* | *Estimates* | *CI (95%)* | *p* |
| Intercept | 0.51 | -28.58 – 29.60 | 0.972 | -5.75 | -31.68 – 20.17 | 0.664 |
| Group size | 0.51 | -0.85 – 1.87 | 0.462 | 0.40 | -0.81 – 1.61 | 0.517 |
| Freeriding (FR)† | -3.89 | -33.16 – 25.39 | 0.795 | -3.13 | -29.20 – 22.95 | 0.814 |
| Legitimate elite†† | -16.24 | -47.38 – 14.89 | 0.307 | -15.31 | -43.04 – 12.42 | 0.279 |
| Illegitimate elite†† | -4.38 | -49.26 – 40.51 | 0.848 | -4.01 | -43.98 – 35.96 | 0.844 |
| Others' intentions | 0.76 | 0.43 – 1.09 | **< 0.001** | 0.89 | 0.59 – 1.18 | **< 0.001** |
| Legit. x FR | 40.86 | -6.76 – 88.49 | 0.093 | 39.40 | -3.02 – 81.81 | 0.069 |
| Illegit. x FR | 23.09 | -33.55 – 79.74 | 0.424 | 23.42 | -27.03 – 73.86 | 0.363 |
| FR x Others' intentions | 0.03 | -0.38 – 0.45 | 0.873 | 0.05 | -0.32 – 0.42 | 0.776 |
| Legit x Others' intentions | 0.23 | -0.21 – 0.66 | 0.303 | 0.24 | -0.15 – 0.63 | 0.226 |
| Illegit. x Others' intentions | 0.07 | -0.57 – 0.71 | 0.830 | 0.07 | -0.50 – 0.64 | 0.809 |
| Legit x FR x Others' intentions | -0.57 | -1.20 – 0.05 | 0.073 | -0.58 | -1.14 – -0.03 | **0.040** |
| Illegit. x FR x Others' intentions | -0.35 | -1.16 – 0.45 | 0.392 | -0.36 | -1.08 – 0.36 | 0.330 |
| **Random Effects** | | | | | | |
| $\sigma^2$ | 447.07 | | | 337.00 | | |
| $\tau_{00}$ | 178.10 $_{\text{Subject}}$ | | | 134.90 $_{\text{Subject}}$ | | |
| ICC | 0.28 | | | 0.29 | | |
| N | 87 $_{\text{Subject}}$ | | | 87 $_{\text{Subject}}$ | | |
| Observations | 522 | | | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.180 / 0.414 | | | 0.287 / 0.491 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding
†† Reference group = No Elite

**Table 2.44.** *Post-hoc tests of the trend of the influence of Others' intentions on Number belief in each Freeriding scenario and Elite condition*

| *Freeriding* | *Elite* | *Others' intentions trend* | *SE* | *df* | *z ratio* | *p* |
|---|---|---|---|---|---|---|
| No FR | No Elite | 0.888 | 0.151 | Inf | 5.883 | < **0.001** |
| FR | No Elite | 0.941 | 0.121 | Inf | 7.753 | < **0.001** |
| No FR | Legitimate Elite | 1.127 | 0.136 | Inf | 8.258 | < **0.001** |
| FR | Legitimate Elite | 0.596 | 0.146 | Inf | 4.076 | < **0.001** |
| No FR | Illegitimate Elite | 0.958 | 0.260 | Inf | 3.681 | < **0.001** |
| FR | Illegitimate Elite | 0.654 | 0.204 | Inf | 3.204 | **0.001** |

**Table 2.45.** *Post-hoc contrasts of the influence of Others' intentions on Number belief between Freeriding conditions in each Elite condition*

| *Contrast* | *Elite* | *Estimate* | *SE* | *df* | *z* | *p* |
|---|---|---|---|---|---|---|
| No FR - FR | No Elite | -0.053 | 0.188 | Inf | -0.284 | 0.776 |
| No FR - FR | Legit. Elite | 0.531 | 0.203 | Inf | 2.622 | **0.009** |
| No FR - FR | Illegit. Elite | 0.303 | 0.328 | Inf | 0.925 | 0.355 |

**Table 2.46.** *Post-hoc contrasts of the influence of Others' intentions on Number belief between Elite conditions in each Freeriding condition*

| *Contrast* | *Freeriding* | *Estimate* | *SE* | *df* | *z* | *p* |
|---|---|---|---|---|---|---|
| No Elite - Legit. Elite | No FR | -0.239 | 0.198 | Inf | -1.210 | 0.447 |
| No Elite - Illegit. Elite | No FR | -0.070 | 0.289 | Inf | -0.241 | 0.968 |
| Legit. Elite - Illegit. Elite | No FR | 0.170 | 0.285 | Inf | 0.595 | 0.823 |
| No Elite - Legit. Elite | FR | 0.345 | 0.193 | Inf | 1.794 | 0.172 |
| No Elite - Illegit. Elite | FR | 0.287 | 0.225 | Inf | 1.278 | 0.408 |
| Legit. Elite - Illegit. Elite | FR | -0.058 | 0.252 | Inf | -0.231 | 0.971 |

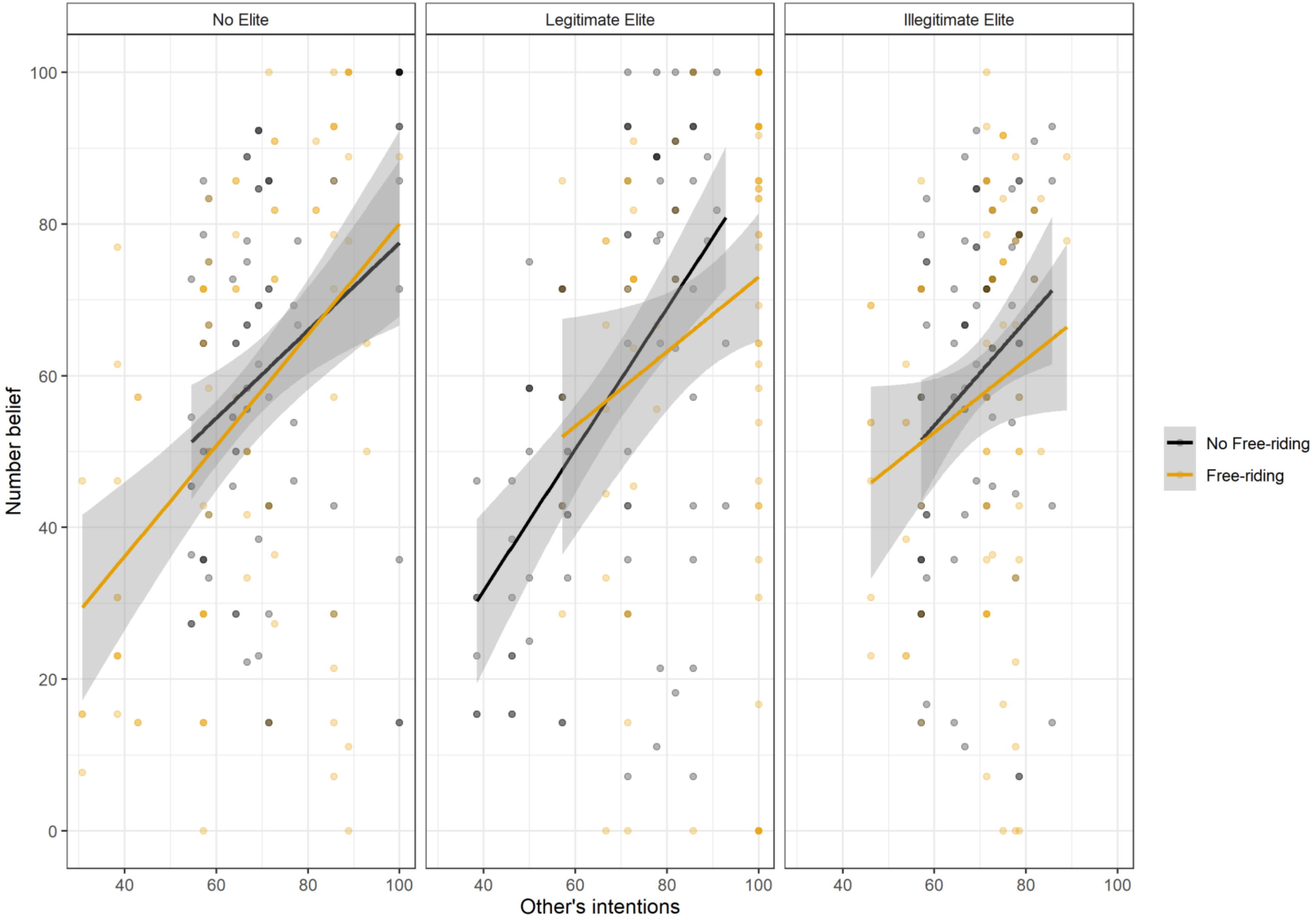


*Figure 2.13.* Number belief predicted by communicated intentions as a function of Freeriding scenario and Elite condition. Shaded areas represent the 95% confidence interval.

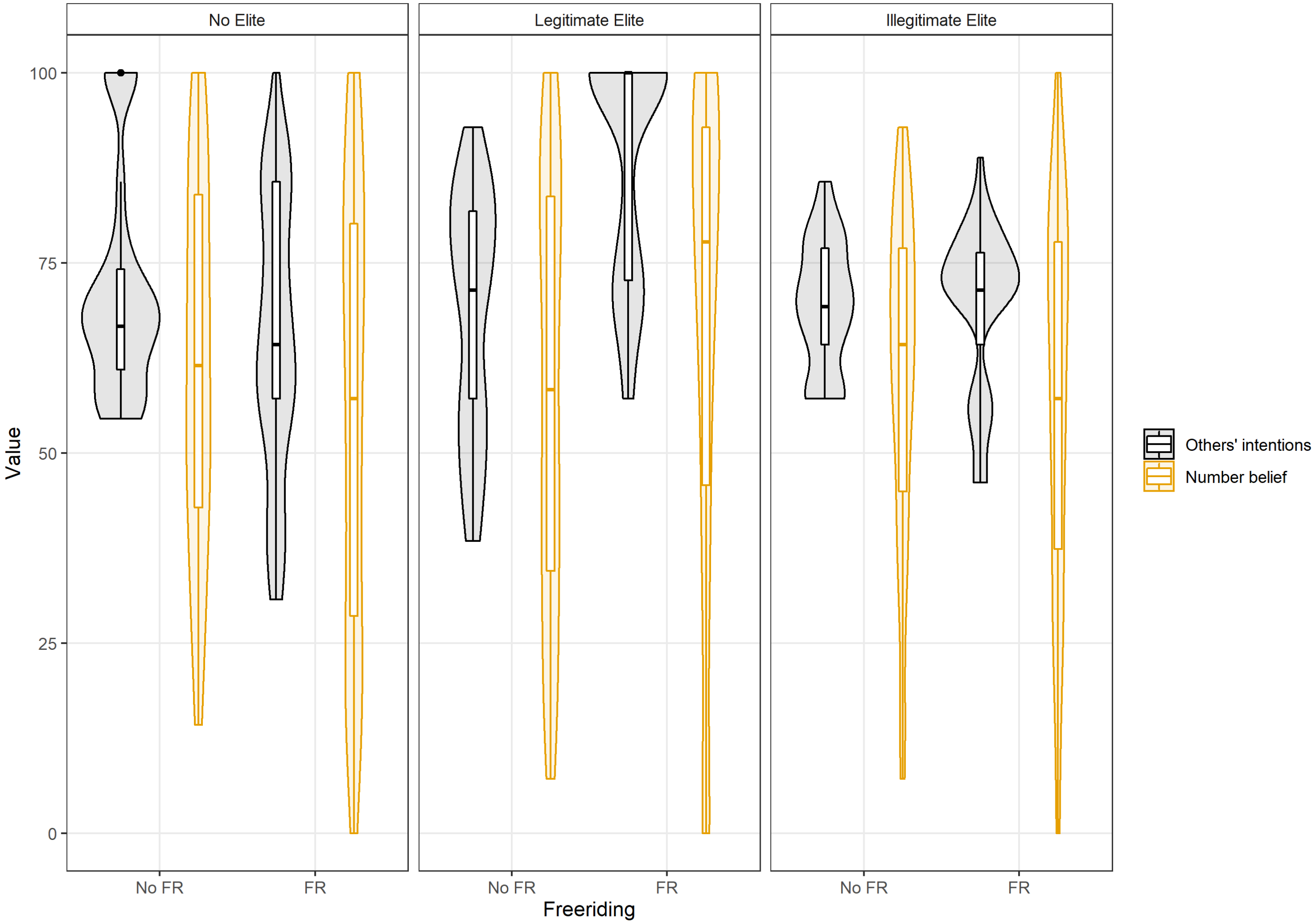


*Figure 2.14.* Violin plot comparing Other's Intentions (black) and beliefs in how many others decided to challenge (orange) in each Freeriding and Elite condition. Violins show the approximate frequency of data at each value. Intentions were highest when Freeriding was possible in the Legitimate Elite system, but in this condition Intentions were also the least believed.

Although most intentions were followed up with consistent decisions (***Figure 2.11***), the finding that intentions were believed to a lesser extent when Freeriding was possible and the Legitimate Elite was present (***Table 2.43.*** Linear mixed effects model predicting Number belief from Freeriding and Elite conditions and Others' intentions within these conditions) warranted further investigation. Specifically, I reasoned that this effect may have been driven by reduced confidence in others following through on their decisions, given the risk of freeriding. I accordingly constructed a model predicting participants' confidence in their Number belief estimation from Freeriding and Elite and their interaction, controlling for group size. Diagnostics flagged potential issues with non-normality of residuals, so I also tested a robust model using 'robustlmm', which showed similar effects. The standard and robust models are presented in ***Table 2.47***. The results show that there were no differences between conditions in how confident participants felt about their estimation of the number of others challenging the system. This result suggests that the lower predictive power of others' intentions on Number belief in the Legitimate Elite Freeriding condition was not due to reduced confidence in others following through on their intentions.

**Table 2.47.** *Standard and robust linear mixed models predicting Confidence in Number belief from Freeriding and Elite conditions.*

| | Standard model | | | Robust model | | |
|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI (95%)* | *p* | *Estimates* | *CI (95%)* | *p* |
| (Intercept) | 0.66 | 0.46 – 0.86 | **< 0.001** | 0.68 | 0.48 – 0.89 | **< 0.001** |
| Group size | -0.01 | -0.03 – 0.00 | 0.155 | -0.01 | -0.03 – 0.00 | 0.137 |
| Freeriding (FR)† | -0.01 | -0.07 – 0.04 | 0.699 | -0.01 | -0.06 – 0.04 | 0.649 |
| Legitimate Elite†† | 0.01 | -0.04 – 0.07 | 0.665 | 0.01 | -0.04 – 0.06 | 0.633 |
| Illegitimate Elite†† | -0.02 | -0.08 – 0.03 | 0.448 | -0.01 | -0.06 – 0.04 | 0.628 |
| FR x Legitimate Elite | 0.06 | -0.02 – 0.14 | 0.123 | 0.06 | -0.01 – 0.13 | 0.095 |
| FR x Illegitimate Elite | 0.03 | -0.05 – 0.11 | 0.447 | 0.02 | -0.05 – 0.09 | 0.620 |
| **Random Effects** | | | | | | |
| $\sigma^2$ | 0.03 | | | 0.03 | | |
| $\tau_{00}$ | 0.03 Subject | | | 0.03 Subject | | |
| ICC | 0.44 | | | 0.52 | | |
| N | 87 Subject | | | 87 Subject | | |
| Observations | 522 | | | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.024 / 0.457 | | | 0.027 / 0.531 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding.
†† Reference group = No Elite

***Thresholds***

Exploring participants' reported Threshold requirements provided further insight. Recall that the Threshold measure asked participants to indicate what number of other participants challenging the system was required for them to decide to also challenge. In essence, this number represents the point at which a participant would *definitely* change their mind from supporting the system (i.e., paying 50) to challenging the system (i.e., paying 0). It was possible that, across conditions, participants had different requirements for how many others they believed would challenge the system before they too would do so. In particular, I reasoned that under conditions of freeriding, participants may have had higher requirements for the number of others claiming an intention to challenge.

I therefore constructed a model predicting this Threshold measure from Group size, Freeriding, Elite, and the Freeriding x Elite interaction. The result of this analysis showed that there were no differences between conditions in this Threshold measure (***Table 2.48***). ***Figure 2.15*** shows that the median Threshold value in all conditions was around 70%. That is, in all conditions, participants reported that they ideally required at least 70% (median value) of other participants to challenge in order to challenge themselves. In the Legitimate Elite Freeriding condition the lower and upper quartiles are slightly higher than in other conditions, suggesting that the Threshold range in this condition was slightly (though not significantly) elevated. The finding that Threshold values were no different across conditions suggests that the changes in context did not overly influence participants' requirements of others. Instead, it appears that participants consistently reported that they would require approximately 70% of others to challenge the system in order to feel confident about challenging the system themselves.

**Table 2.48**. *Linear mixed effects model predicting Threshold from Freeriding and Elite conditions*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| (Intercept) | 57.33 | 34.62 – 80.04 | **< 0.001** |
| Group size | 0.65 | -1.10 – 2.41 | 0.466 |
| Freeriding (FR) † | 0.17 | -4.82 – 5.17 | 0.945 |
| Legitimate Elite†† | 1.53 | -3.47 – 6.52 | 0.549 |
| Illegitimate Elite†† | 0.80 | -4.19 – 5.80 | 0.753 |
| FR x Legit. Elite | -0.00 | -7.06 – 7.06 | 1.000 |
| FR x Illegit. Elite | -1.95 | -9.01 – 5.11 | 0.588 |
| **Random Effects** | | | |
| $\sigma^2$ | 282.51 | | |
| $\tau_{00\ \text{Subject}}$ | 394.04 | | |
| ICC | 0.58 | | |
| $N_{\text{Subject}}$ | 87 | | |
| Observations | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.005 / 0.585 | | |

*Note.* † Reference group = No Freeriding; Treatment group = Freeriding.
†† Reference group = No Elite.

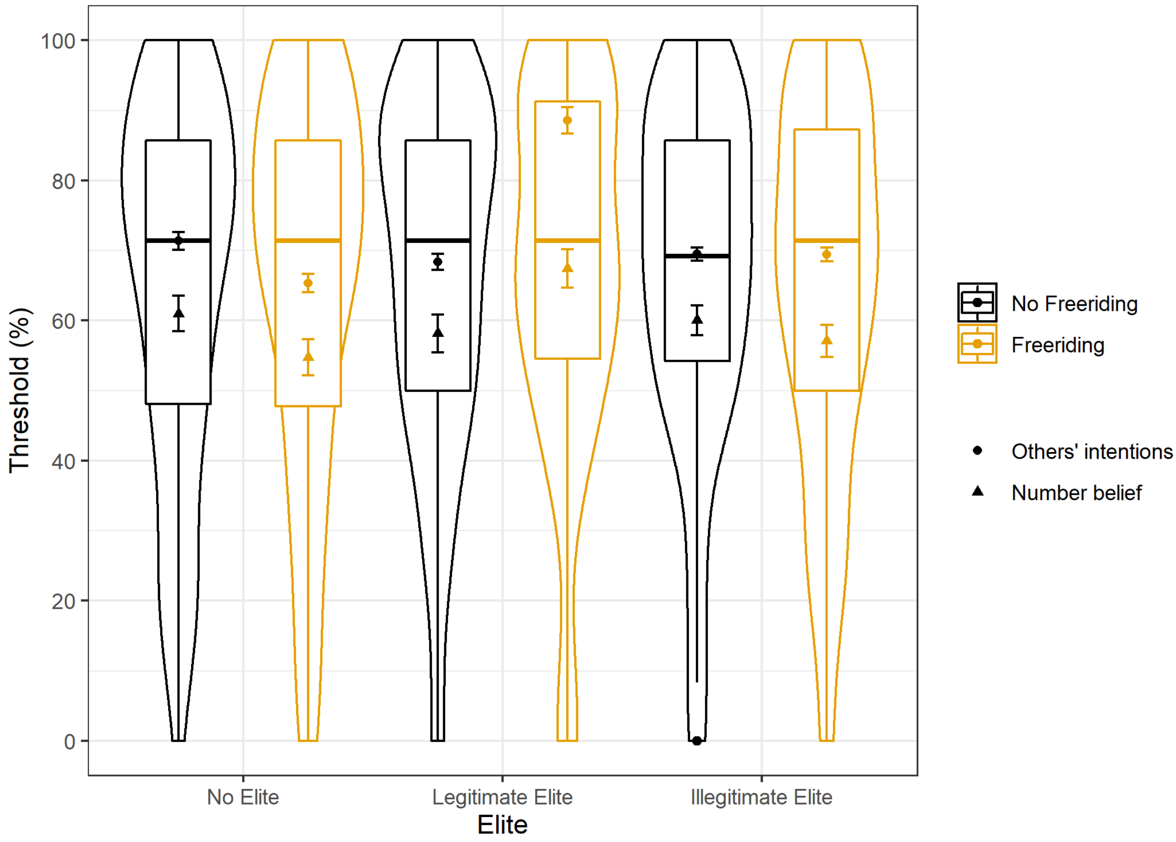


*Figure 2.15*. Spread, central tendency, and density statistics for Threshold values in each Freeriding and Elite condition. Overlaid are the corresponding mean values of Others' intentions (circles) and Number belief (triangles) and their associated within-subject standard error bars.

***Figure 2.15*** also plots the mean proportion of others communicating an intention to challenge (circles) and the mean belief in the number of others challenging (triangles). The plot shows that in most conditions the mean proportion of intentions is close to the median Threshold requirement. The exception is the Legitimate Elite Freeriding condition, where the mean proportion of intentions is far higher (~ 90%) than the median Threshold (~70%). This also has the effect of raising the average belief in the number that actually challenged. This pattern seems to suggest that, though participants' requirements of others was no different in the Legitimate Elite Freeriding condition compared to other conditions, intentions were notably higher in this condition. This led to a corresponding increase in Number belief, but also produced a greater difference between intentions and Number belief. In combination, these results support the notion that the apparent decreased trust that others would follow through on their intentions in the presence of a Legitimate Elite when freeriding was possible can be better understood as a result of exceptionally high intentions to challenge.

Together, the findings relating to communicated intentions provided little evidence to suggest that participants took advantage of the opportunity to freeride by actively misdirecting one another with their communicated intentions. It also appeared that the threat of freeriding did not undermine participants' confidence in others' decisions. The reduced predictive power of intentions in the Legitimate Elite Freeriding condition might therefore be explained by the particularly strong intentions to challenge in that condition.

## Interim Summary

In Experiment 2 I explored how the opportunity to communicate intentions influenced subjective experiences and decisions in Freeriding and No Freeriding contexts. I found that in the system with the Legitimate Elite, disadvantaged participants identified with one another more strongly, the system was rated as fairer, and group efficacy beliefs were stronger, when Freeriding was possible compared to when it was not. In contrast, none of these differences between Freeriding conditions were observed for the other Elite conditions, and this pattern was not evident when participants were not able to communicate intentions.

The odds of a participant deciding to challenge the system were highest when Freeriding was possible and a meritocratically-designated Elite benefited from the unfairness. As in Experiment 1, efficacy beliefs positively predicted decisions in both Freeriding and No Freeriding contexts. However, in Experiment 2, the influence of Number belief was stronger when freeriding was not possible, compared to when it was possible. Feelings of identification amongst disadvantaged participants also increased the odds of challenging. In contrast to

Experiment 1, where Identification had no effect in the No Elite No Freeriding condition, in Experiment 2 the influence of Identification was relevant in all conditions.

Participants' intentions to challenge followed the same pattern as their decisions; intentions to challenge were highest in the Legitimate Elite condition when Freeriding was possible. As expected, others' intentions positively predicted efficacy beliefs (i.e., Number belief). However, in Legitimate Elite systems others' intentions were significantly less influential in predicting efficacy when Freeriding was possible compared to when it was not possible. Feelings of identification were found to predict the extent to which intentions were believed when Freeriding was not possible but not when it was possible. Overall, comparing data of Experiments 1 and 2, the ability to communicate increased the odds of challenging the system by 2.48 times relative to when communication was not possible. Finally, though in the Legitimate Elite Freeriding condition intentions appeared to be less strongly believed, the inferential statistics showed that decisions were broadly consistent with intentions and across all conditions participants were seven-and-a-half times more likely to challenge the system if they had previously expressed an intention to do so. Thus, there was little evidence to support the idea that participants were motivated to take advantage of one another by reneging on intentions to challenge.

## Discussion

I investigated how the incentive structure of group decision-making and the salience of social inequality influenced individuals' decisions to challenge unfair social systems. I did this by comparing decisions to challenge in a scenario where it was not possible to freeride (i.e., participants had to "be in it to win it") to decisions in a scenario where participants could freeride by allowing others to take risks, but not contributing personally. Scenarios also differed in how salient social inequality was; whereas in some systems nobody benefited from the unfairness, in other systems there was an "Elite" participant who benefited from the unfairness of the system. I found that the overall effect of freeriding was contingent on whether social inequality was emphasised with the presence of an Elite. Specifically, without the opportunity to communicate intentions, the possibility of freeriding reduced the likelihood of challenging the system (i.e., cooperative choices), but only in the absence of social inequality. In the presence of social inequality, cooperation was maintained when freeriding was possible by a salient differential between advantaged and disadvantaged participants. Communication changes this pattern. In Experiment 2, the ability to communicate an intention prior to deciding made participants *more* likely to challenge the system when freeriding was possible when an intergroup inequality was made salient.

Previous research has shown that the way in which a social dilemma is framed can have important implications for how decision-makers will behave. Typically, describing scenarios in collaborative terms (e.g., by using words such as "community", "social exchange") tends to improve cooperation, compared to describing them with more utilitarian language (e.g., "business transaction", "stock market game"; Batson & Moran, 1999; Ellingsen et al., 2012; Kay & Ross, 2003; Liberman et al., 2004). Theorists have pointed to several mechanisms through which these framing effects influence behaviour. One possibility is that the way in which a situation is described provides individuals with an understanding of *what is* *appropriate* behaviour in the current context (Weber et al., 2004). Though this can be likened to a priming effect (Ferraro et al., 2005), it is argued that this process involves changes in social preferences based on normative information, which affect the subjective value of different options. Another possibility is that the framing of a scenario can influence what decision-makers *expect others to do*. Thus, this explanation invokes mentalising processes in which decision-makers consider others' preferences and intentions and make their decisions according to this assessment (Fehr & Gächter, 2000; Fehr & Schmidt, 1999).

Though previous research has suggested that these two mechanisms can be mutually exclusive (e.g., Ellingsen et al., 2012), the finding that when there was no social inequality participants were less likely to challenge the system when freeriding was possible suggests that both mechanisms were present, and further, that they interacted. When social inequality was not emphasised (i.e., the "No Elite" condition), participants were less likely to challenge the system when freeriding was possible. According to the first mechanism, it is possible that participants felt that supporting the system (i.e., not challenging) was appropriate because they had surmised from the instructions that it was always in their financial interest to do so (which was not the case when freeriding was not possible). In addition, supporting the system might have been considered the default option; all scenarios in the experiment were defined by payments to the system, with challenging by definition representing a less orthodox response. This, combined with the above financial incentive of supporting, makes this option all the more preferred. However, determining what is appropriate in a group decision scenario, in which outcomes are contingent on one another's decisions, must also involve consideration of the expected behaviour of others. It is likely that – at least in the present experiment – forming preferences based on normative information was inseparable from considering the intentions of others. Indeed, the finding that cooperation was not affected when social inequality was emphasised provides strong evidence that participants included the expected behaviour of others in their evaluations.

Further insight into these possible mentalising processes comes from the findings of Experiment 2, where participants were able to communicate intentions before a making a decision. When communication was possible (Experiment 2), participants were nearly two-and-a-half times more likely to challenge the system, compared to when it was not (Experiment 1). This is consistent with much of the social dilemma literature (for meta-analysis see Balliet, 2010). As was the case when communication was not possible, when individuals could communicate intentions and it was possible to freeride, individuals were more likely to challenge when social inequality was salient, compared to when it was not. Thus, the overall effect of freeriding when communication was possible was similar to its effect when communication was not possible, but the pattern was shifted towards a greater tendency to challenge (compare ***Figure 2.2*** and ***Figure 2.6***).

Experiment 2 permits a deeper inspection of the possible processes underlying this effect by exploring communicated intentions and how they translated to efficacy beliefs. The overall trend across all conditions was that participants' beliefs about the number of others challenging were lower than the communicated intentions, which demonstrates that in all situations, participants, on average, either believed that some individuals who communicated an intention to challenge would change their mind, or deliberately lied about their intention (or both). Intentions to challenge the system were most likely in the presence of the Legitimate Elite, and participants' beliefs about the number who ultimately challenged were consistent with this pattern. Despite this, in the presence of a meritocratically-designated Elite, intentions were a weaker predictor of beliefs when freeriding was possible compared to when it was not. Thus, although there were strong *intentions* to challenge when freeriding was possible and social inequality was salient, these intentions were less strongly associated with *decisions* to challenge compared to when freeriding was not possible. Though this suggests that freeriding reduced individuals' confidence that others would follow up on their intentions, there was no evidence to suggest that efficacy beliefs were lower in this condition, nor that individuals were less confident in those beliefs.

Perhaps the most parsimonious explanation for this finding is therefore that the greater disconnect between intentions and beliefs was driven by the exceptionally strong intentions to challenge in the presence of the meritocratically-designated Elite when freeriding was possible. Here, the mean proportion of other group members communicating an intention to challenge was nearly 90% (***Table 2.37***). Though feelings of efficacy correspondingly increased and were indeed highest in this condition out of all others, there was still a larger difference between intentions and efficacy beliefs than in other conditions (***Figure 2.14, Figure 2.15***). The effect of apparently less believed intentions in the presence of the meritocratically-designated Elite when freeriding was possible might therefore be understood as an artefact of the exceptionally high frequency of challenge intentions rather than a social psychological effect (e.g., reduced trust). The findings relating to individuals' required action of others (i.e., the number of others challenging that would encourage an individual to also challenge) support this notion; across conditions participants were remarkably consistent in their requirement of approximately 70% of others challenging for them to be willing to challenge as well. The finding that, when freeriding was possible and a merit-worthy Elite was present, participants did not require a greater number of challengers, may indirectly indicate that participants did not perceive this condition as riskier than the other conditions, adding further support to the idea that reduced association between intentions and efficacy beliefs was not due to reduced trust that others would follow up on their intentions.

This begs the question of why intentions were so high in this condition. Whereas intentions to challenge when social inequality was least salient (i.e., the "No Elite" condition) and when there was an Elite who was not there by merit (i.e., the "Illegitimate Elite" condition) were relatively similar when freeriding was possible compared to when it was not, intentions were at their highest in the presence of the meritocratically-designated Elite (i.e., the "Legitimate Elite" condition) when freeriding was possible. This freeriding scenario was identical to the scenario involving the same Elite when freeriding was *not* possible except for its incentive structure. It is therefore highly likely that it was this factor that led to the differences observed. However, it is striking that intentions to challenge *increased* when the incentive to challenge the system was lower. This could have happened either (i) because participants genuinely felt more cooperative when freeriding was possible compared to when it was not, or (ii) because some participants were strategically deceptive and signalled an intention to challenge to encourage others to challenge (and then subsequently supported). Though the former seems unlikely, evidence for the latter is inconclusive. Inspection of ***Figure 2.11*** indicates that typically individuals followed up on their intentions; more people who challenged the system had previously expressed an intention to challenge, and more people who supported had previously expressed an intention to support. However, switching evidently did occur. In all conditions, approximately 20% of those who said they would challenge changed their mind and ultimately decided to support the system. Though it cannot be ruled out that these individuals attempted to manipulate others by encouraging them to challenge and avoid challenging themselves, the fact that these rates were very similar when freeriding was not possible (where there was no incentive – indeed where there was a disincentive – to misdirect others) raises the possibility that this small number of participants were not actively deceiving one another. Furthermore, the finding that an intention to challenge could account for most of the variance in decisions to challenge, irrespective of the type of person benefiting from the unfairness and whether freeriding was possible, suggests that the predictive power of an individual's intention was not reduced when freeriding was possible. This evidence suggests that participants were not using a strategy of deception to capitalise on the opportunity to freeride.

In contrast, though it seems strange to suggest that individuals became more cooperative when there was more risk that they could take advantage of one another, there may be a case for concluding just that. Firstly, if participants were motivated to take advantage of one another by communicating challenge intentions and subsequently freeriding, this should have been evident to some extent in all freeriding conditions. This was not the case. Extremely high intentions to challenge when freeriding was possible were only observed in the presence of the meritocratically-designated Elite. Secondly, it is unlikely to be coincidence that it was this condition that I previously observed to be particularly mobilising for participants in Study 1. Recall that I considered the possibility that this effect may be explained by a strengthening of identification amongst group members when there was a salient outgroup to whom blame could be attributed (i.e., politicisation of identity, Simon & Klandermans, 2001). The finding that mobilisation was even stronger in response to the possibility of freeriding may be interpreted as evidence of strong moral conviction amongst disadvantaged participants (Skitka, 2010). Indeed, moral conviction has been suggested as one of four core motivations for collective action (van Zomeren, 2013, 2019) and has been demonstrated to predict identification and collection action intentions and behaviour (Mazzoni et al., 2015; van Zomeren et al., 2012).

It should be noted, however, that moral convictions have been conceptualised as context-independent values about what is right and wrong (Skitka, 2010; Skitka et al., 2008). According to this conceptualisation, it is problematic to infer that moral convictions relating to the fairness of the present study’s system were present only in one specific context (i.e., when freeriding was possible and there was a meritocratically-designated Elite who benefited from the unfairness), since a moral conviction against distributive injustice should have been present in all conditions. I suggest, however, that whilst individuals and groups in the experiment may be understood to possess a moral standpoint relating to justice in all conditions, moral convictions may have been specifically more *relevant* when freeriding was possible and the Elite earned their position through merit. This was because the saliency of the issue as a moral dilemma was stronger in this scenario. Firstly, as established in Study 1, the perception of a status differential was particularly strong in the presence of an Elite who was meritocratically designated. In groups, this led to a relatively more politicised identity that motivated group members to take action. In the present study, the freeriding scenario posed more of a moral dilemma than the no freeriding scenario because it presented a stronger conflict between the interests of the group and the interests of the individual when confronting the systemic unfairness. When freeriding was not possible, the only way a group member could achieve the best outcome was by taking a personal risk; only by personally challenging the system could an individual benefit from its overturn. In contrast, when freeriding was possible, each individual had an incentive to capitalise on the risk of others without taking a personal risk themselves. This created a high-stakes scenario where each challenger had to trust that other group members would challenge instead of freeride. The increased trust requirement of the freeriding scenario, combined with finding that challenging the system was highest in this condition, implies that decision-makers were motivated beyond self-interest and were collectively resolute in challenging. The fact that the moral dilemma posed by freeriding was met with increased cooperation is highly suggestive that individuals drew on a collective moral purpose. This assertion is consistent with van Zomeren et al.'s (2012) argument that moral convictions are effective in encouraging collective action because of the normative fit between the conviction and identity, pointing out that this fit is likely to be strongest when identities are politicised.

Several other results provide support for this picture. First, finding that identification was also higher in the presence of a meritocratically-designated Elite when freeriding was possible, compared to when it was not, is further evidence that the possibility of freeriding galvanised participants. Again, this finding is somewhat surprising; under conditions where participants should have been expected to trust each other less, they appeared instead to feel more, not less, associated. Second, overall, the predictive influence of efficacy beliefs was lower when freeriding was possible, suggesting that expectations about success were less relevant for participants when freeriding was possible compared to when it was not. Though there is a possibility that this was due to decreased confidence in efficacy beliefs under freeriding conditions, as I have considered above, there appears to be little evidence that participants were motivated to behave deceptively or that cooperation was undermined by a fear of freeriding. Instead, this reduced role of efficacy implies that other forces were responsible for the strong mobilisation observed. Combined, these findings further suggest the possibility that moral convictions were involved in the strong mobilisation observed when freeriding was possible in the presence of a meritocratically-designated Elite. Ultimately though, in the absence of explicit measurement of moral convictions these possibilities remain speculative.

Overall, these results present evidence that individuals' collective rejection of inequality is sensitive to the incentives of the decision scenario. Contrary to expectations, when there was a conflict between self- and group-interests, participants were *more* motivated to take personal risks and challenge the system. Interestingly, this was the case only when the beneficiary of the unfairness was meritocratically designated. Assuming the argument from Study 1 that mobilisation against this Legitimate Elite was so great because they were considered an "outsider", the result suggests that if individual decision-makers are sufficiently bound by a shared identity in a collective endeavour, not only may they be resistant to the temptation of freeriding but this possibility may bolster their conviction and induce them to take action on behalf of the group. However, this effect appears to rely on group members' ability to communicate to coordinate their action and on the system having a clear antagonist. Organisers of real-world social movements might therefore be advised to aim to personify objects of protest (i.e., by identifying individual(s) of the opposition, rather than a faceless notion of the system) and to ensure that there are effective communication channels between all members of the movement (and beyond) that will facilitate the expression of intentions and convictions.

## References

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*, *54*(1), 39–57. https://doi.org/10.1177/0022002709352443

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Batson, C. D., & Moran, T. (1999). Empathy-induced altruism in a prisoner's dilemma. *European Journal of Social Psychology*, *29*(7), 909–924. https://doi.org/10.1002/(SICI)1099-0992(199911)29:7<909::AID-EJSP965>3.0.CO;2-L

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology*, *47*(3), 453–480. https://doi.org/10.1348/014466607X244970

Boone, C., Declerck, C., & Kiyonari, T. (2010). Inducing cooperative behavior among proselfs versus prosocials: The moderating role of incentives and trust. *Journal of Conflict Resolution*, *54*(5), 799–824. https://doi.org/10.1177/0022002710372329

Chen, C. W., & Gorski, P. C. (2015). Burnout in social justice and human rights activists: Symptoms, causes and implications. *Journal of Human Rights Practice*, *7*(3), 366–390. https://doi.org/10.1093/jhuman/huv011

Clark, L. P., Millet, D. B., & Marshall, J. D. (2014). National patterns in environmental injustice and inequality: Outdoor no2 air pollution in the united states. *PLOS ONE*, *9*(4), e94431. https://doi.org/10.1371/journal.pone.0094431

Colman, A. M. (2003). Beyond rationality: Rigor without mortis in game theory. *Behavioral and Brain Sciences*, *26*(02). https://doi.org/10.1017/S0140525X0350005X

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*(1), 169–193. https://doi.org/10.1146/annurev.ps.31.020180.001125

Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, *76*(1), 117–130. https://doi.org/10.1016/j.geb.2012.05.007

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*(3), 159–181. https://doi.org/10.1257/jep.14.3.159

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151

Ferraro, F., Pfeffer, J., & Sutton, R. I. (2005). Economics language and assumptions: How theories can become self-fulfilling. *Academy of Management Review*, *30*(1), 8–24. https://doi.org/10.5465/amr.2005.15281412

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9

Hardin, G. (1968). The tragedy of the commons. *Science, 162*, 1243–1248. https://doi.org/10.1126/science.162.3859.1243.

Kay, A. C., & Ross, L. (2003). The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the Prisoner's Dilemma. *Journal of Experimental Social Psychology*, *39*(6), 634–643. https://doi.org/10.1016/S0022-1031(03)00057-X

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, *24*(1), 183–214. https://doi.org/10.1146/annurev.soc.24.1.183

Lee, A. L. Y., Prendergast, G. P., Yim, F. H. K., & Choi, L. (2019). A social dilemma perspective on recycling. *Journal of Applied Social Psychology*, *49*(9), 585–595. https://doi.org/10.1111/jasp.12618

Lenth, R. V., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.6.0) [Computer software]. https://CRAN.R-project.org/package=emmeans

Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin*, *30*(9), 1175–1185. https://doi.org/10.1177/0146167204264004

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Mazzoni, D., Zomeren, M. van, & Cicognani, E. (2015). The motivating role of perceived right violation and efficacy beliefs in identification with the italian water movement. *Political Psychology*, *36*(3), 315–330. https://doi.org/10.1111/pops.12101

McAdams, R. H. (2009). Beyond the Prisoners' Dilemma: Coordination, game theory, and law. *82 Southern California Law Review*, *209*, 51.

Olson, M. (1965). *The Logic of Collective Action*. Harvard University Press.

Patel, J. A., Nielsen, F. B. H., Badiani, A. A., Assi, S., Unadkat, V. A., Patel, B., Ravindrane, R., & Wardle, H. (2020). Poverty, inequality and COVID-19: The forgotten vulnerable. *Public Health*, *183*, 110–111. https://doi.org/10.1016/j.puhe.2020.05.006

Perlin, S. A., Wong, D., & Sexton, K. (2001). Residential proximity to industrial sources of air pollution: interrelationships among race, poverty, and age. *Journal of the Air & Waste Management Association*, *51*(3), 406–421. https://doi.org/10.1080/10473289.2001.10464271

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*(1), 58–92. https://doi.org/10.1177/104346395007001004

Simon, B., & Klandermans, B. (2001). Politicized collective identity: A social psychological analysis. *American Psychologist*, *56*, 319–331. https://doi.org/10.1037/0003-066X.56.4.319

Simpson, B. (2004). Social values, subjective transformations, and cooperation in social dilemmas. *Social Psychology Quarterly*, *67*(4), 385–395. https://doi.org/10.1177/019027250406700404

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, *4*(4), 267–281. https://doi.org/10.1111/j.1751-9004.2010.00254.x

Skitka, L. J., Bauman, C. W., & Mullen, E. (2008). Morality and justice: An expanded theoretical perspective and empirical review. In K. A. Hegtvedt & J. Clay-Warner (Eds.), *Justice* (Vol. 25, pp. 1–27). Emerald Group Publishing Limited. https://doi.org/10.1016/S0882-6145(08)25001-0

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge University Press.

Smith, S. M., Haugtvedt, C. P., & Petty, R. E. (1994). Attitudes and recycling: Does the measurement of affect enhance behavioral prediction? *Psychology and Marketing*, *11*(4), 359–374. https://doi.org/10.1002/mar.4220110405

Thöni, C., & Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters*, *171*, 37–40. https://doi.org/10.1016/j.econlet.2018.06.022

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

Van Lange, P. A. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, *77*(2), 337-349. https://doi.org/10.1037/0022-3514.77.2.337

Van Lange, P. A. M., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, *120*(2), 125–141. https://doi.org/10.1016/j.obhdp.2012.11.003

van Zomeren, M. (2013). Four core social-psychological motivations to undertake collective action. *Social and Personality Psychology Compass*, *7*(6), 378–388. https://doi.org/10.1111/spc3.12031

van Zomeren, M. (2019). Toward a cultural psychology of collective action: Just how "core" are the core motivations for collective action? *Journal of Pacific Rim Psychology*, *13*, e15. https://doi.org/10.1017/prp.2019.7

van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin*, *134*(4), 504–535. https://doi.org/10.1037/0033-2909.134.4.504

van Zomeren, M., Postmes, T., & Spears, R. (2012). On conviction's collective consequences: Integrating moral conviction with the social identity model of collective action. *British*

*Journal of Social Psychology*, *51*(1), 52–71. https://doi.org/10.1111/j.2044-8309.2010.02000.x

Weber, J. M., Kopelman, S., & Messick, D. M. (2004). A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness. *Personality and Social Psychology Review*, *8*(3), 281–307. https://doi.org/10.1207/s15327957pspr0803_4

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

# Study 3. Inconvenient Truths: The approach and avoidance of information about social inequality

## Abstract

Though social equality is desirable, inequality in human society is common. Research suggests that individuals implement strategies for managing this disparity to maintain a positive view of the world. One possible strategy is to avoid information about inequality and thereby maintain reasonable doubt about the presence of inequality. The personality literature further documents how individual difference characteristics may help individuals maintain positive beliefs about the world despite their knowledge that injustice exists within it. An important question therefore concerns the extent to which individuals seek or avoid equality information, and what character traits are important for these decisions. I created an unfair social system by having participants work with a virtual partner who may have been paid more or less than them and asked them to choose either to find out or not find out about the payments. The direction of inequality was determined by income; participants with a lower income than their virtual partner were potentially disadvantaged (i.e., if there was inequality they would receive *less* than their partner) whereas participants with higher income were potentially advantaged (i.e., if there was inequality they would receive *more* than their partner). I found that System Justification and Social Value Orientation were important predictors of deciding to find out about possible inequality and their influence varied depending on whether the individual was potentially advantaged or potentially disadvantaged by inequality. Those potentially disadvantaged by inequality were more likely to find out if they were strong justifiers compared to weak justifiers, but for those potentially advantaged by the inequality system justification had no influence. Greater pro-sociality was associated with a greater likelihood of finding out amongst the potentially advantaged and a lower likelihood of finding out amongst the potentially disadvantaged. The results suggest that the motivations for ascertaining equality

information may depend on the individuals' relative social position, with system justifying ideologies being important for predicting the decisions of the potentially disadvantaged but not of the potentially advantaged, and pro-sociality leading to information-seeking amongst the potentially advantaged, but not the potentially disadvantaged.

## Introduction

Humans typically demonstrate preferences for fairness over unfairness. Experimental evidence for this comes from an abundant literature in behavioural economics. For example, the Ultimatum Game (UG; Güth et al., 1982) is a wealth distribution paradigm that has been used to demonstrate preferences for fairness even when unfair situations still represent net gains. The game involves two individuals: a proposer and a responder. The proposer is endowed with a sum of money and must decide how to split the sum between oneself and the responder. The responder must decide either to accept the proposer's suggested division of the wealth or to reject it. If the responder accepts the proposition, the wealth is divided as proposed. If the responder rejects the proposition, the wealth is not distributed at all and both players receive nothing. In terms of utility, the best action for the responder is to accept any offer. In game theory parlance this is known as subgame perfection; for any offer made by the proposer, it is always beneficial for the responder to accept it, since some money – however it is distributed – is better than none. By the same token, the proposer's best strategy is to offer as little as possible since the responder is expected to accept any gain however small.

Experimental results from the UG show that individuals deviate from the game theoretical predictions. On average, proposers offer 40% of the total amount to responders (Oosterbeek et al., 2004) and responders tend to reject offers of 20% or less (Güth et al., 1982; Sanfey et al., 2003; Thaler, 1988). These findings indicate that responders are willing to forfeit net gains and punish the proposer if the division of wealth is considered unfair (Bolton & Zwick, 1995), and that proposers are conscious that unfair offers will be rejected and propose fairer divisions that are more likely to be accepted. Importantly, the finding that activity in insular cortex – an area of the brain implicated in emotional experience – is observed in response to unfair offers (Sanfey et al., 2003) suggests a moral foundation of this fairness principle.

Though findings from the UG show that responders have a strong fairness motivation, the fairness motives of proposers could still be construed in terms of utility maximisation since they anticipate that unfair offers will be rejected. The Dictator Game (DG) is similar to the UG but differs in that non-proposers have no say in the split of wealth; the proposer decides how wealth is distributed and that decision is final. Results from the DG indicate that proposers, even when there is no risk of punishment, still offer more than zero (Bolton et al., 1998; Camerer & Thaler, 1995; Forsythe et al., 1994; Guala & Mittone, 2010; Henrich et al., 2004). Thus, proposers in the DG make non-zero offers despite there being no financial repercussions for offering nothing. This provides further evidence that there is a non-utilitarian (i.e., morality-based) preference for fairness, and indeed it has been suggested that beyond economic utility, proposers may derive emotional utility from the recipient’s gain (Eckel & Grossman, 1996).

Beyond the context of behavioural economic paradigms, research supports the notion that humans are disposed to equal resource distributions. Most seven- to eight-year olds are found to prefer equal versus unequal division of resources (Fehr et al., 2008). Moreover, children have been found to purposefully discard resources rather than allocate them unequally, providing support evidence of a motivation for *equity* itself and not just a motivation for sharing (Shaw & Olson, 2012). This sensitivity to inequity has also been reported in non-human mammals such as chimpanzees (Brosnan et al., 2010), dogs (Range et al., 2009), and corvids (Wascher & Bugnyar, 2013), indicating an evolutionary heritage to inequity aversion amongst social species (Brosnan & de Waal, 2014).

Observation of the literature therefore indicates that humans possess a preference for fairness. Observation of the world, however, makes it clear that real-world inequality is pervasive. Though apparently few people have a preference for it, systemic societal unfairness remains a salient feature in contemporary human society. An understanding of how inequality comes to be a societal staple can be greatly informed by understanding the psychological forces that encourage or discourage individuals to seek information about the issue. In this study, I explored how several individual difference measures influence whether individuals seek or avoid (in)equality information.

An effective strategy for dealing with potential unfairness is to avoid information that highlights it. Individuals are documented to actively or passively avoid or ignore information that has the potential to cause negative affect. For example, in health settings, some patients prefer to avoid taking or receiving results for diagnostic tests for health conditions such as HIV, Alzheimer's disease, and cancer (Cutler & Hodgson, 2003; Friis et al., 2003; Lyter et al., 1987; Thompson et al., 2002). In the present pandemic, individuals have likewise been documented to avoid information to reduce distress and anxiety even though this information is important for physical well-being (Siebenhaar et al., 2020; Song et al., 2021).

On the flipside, remaining in a state of uncertainty has also been shown to prolong positive affect. For example, Wilson et al. (2005) showed that positive moods after receiving an unexpected monetary gift were higher under conditions of relatively greater uncertainty/ambiguity, suggesting that the uncertainty enhanced positive affect. Similarly, some expectant parents who opt not to discover the sex of their unborn child report their decision as motivated by a desire to maintain the sense of surprise at birth, thus prolonging the sense of positive affect (Shipp et al., 2004).

In the domain of equality and fairness, experimental findings suggest that information avoidance can influence social decision-making. For example, dictators in the dictator game make more self-interested decisions when they can avoid information about how payments will be distributed; by avoiding information they are able to maintain a belief (genuine or fabricated) that the payment distribution that is best for them personally is also best for their partner (Dana et al., 2007). In a similar experiment, varying the probability that the wealth distribution in the dictator game was beneficial to the both parties showed that dictators' moral "wiggle room" (Dana at al., 2007) was particularly evident when the probability of a pro-social distribution was high, suggesting that concern for fairness was still present (Feiler, 2014). These and other studies (e.g., Grossman & van der Weele, 2017; Larson & Capra, 2009) demonstrate that information seeking and avoidance are relevant to issues of social inequality.

The present study aimed to add to the existing experimental literature by investigating decisions to seek out or avoid information about the possibility of real payment inequality experienced in the experiment itself. I postulated that several individual difference characteristics would also be important to the approach or avoidance of such information. These were System Justification, Belief in a Just World, Life Orientation, Intolerance of Uncertainty, and Social Value Orientation. I chose this range of characteristics to establish a comprehensive picture of the possible motivations that individuals may have for seeking or avoiding equality information. As I describe below, these characteristics describe individuals' tendency to perceive the world as stable rather than chaotic (i.e., Belief in a Just World, General System Justification), individuals' dispositions toward the world and the unknown (i.e., Life Orientation, Intolerance of Uncertainty), and characteristics explicitly concerning capital and the division of wealth (i.e., Social Value Orientation, Economic System Justification).

**System Justification**

Social systems are complex and vary in the extent to which ideals of fairness are objectively implemented and subjectively experienced. System Justification Theory (SJT; Jost & Banaji, 1994) aims to explain how individuals manage the reality that equality – whilst apparently intrinsically valued – is often lacking in social systems. SJT argues that individuals possess (to varying extents) a tendency to justify the status quo so that existing social systems are perceived to be fair and legitimate (Jost & Banaji, 1994; Jost & Hunyady, 2003). This tendency is motivational in nature (Kay et al., 2009) and serves a palliative function (Jost & Hunyady, 2003) insofar as it increases satisfaction with the existing social structure, prevents negative affect, and preserves well-being (Hammond & Sibley, 2011; McCoy et al., 2013; Osborne & Sibley, 2013). System justification is posited to address a triad of needs (Jost, 2019; Jost et al., 2008; Jost & Hunyady, 2005): epistemic (i.e., needs for certainty), existential (i.e., needs for safety and security), and relational (i.e., needs to maintain a reality shared with others).

The epistemic motive of system justification has been evidenced by studies that demonstrate that conservative political values are more strongly endorsed under conditions of impaired deliberative thought. For example, alcohol intoxication, cognitive load and time pressure are found to positively correlate with endorsement of conservative values (Eidelman et al., 2012), and low-effort thinking is associated with greater support for hierarchy (Van Berkel et al., 2015). Of particular relevance to the present study, intolerance of uncertainty and ambiguity are positive predictors of political conservatism (Jost et al., 2003, 2007). In addition, low need for cognition (i.e., low motivation to cognitively engage with an issue; Cacioppo et al., 1984) is associated with greater system justification (Hennes et al., 2012). This research shows that ideological conservatism and support for the status quo are closely linked with needs to control or reduce uncertainty and the way in which individuals approach and use information.

Evidence for the existential motive of system justification is provided by studies demonstrating that feeling threatened can induce greater endorsement of conversative, system-justifying ideologies. For example, liberals induced to think about their own mortality become conservative in their views on social issues such as capital punishment (Nail et al., 2009). Priming people with threatened mindsets has been found to promote close-mindedness and conservatism (Thórisdóttir & Jost, 2011), and people who experience traumatic threatening events are found to become more conservative (Bonanno & Jost, 2006).

Desires to share reality with others has been shown to be positively related with system-justifying ideologies (Hennes et al., 2012; Jost et al., 2017). In response to social exclusion, individuals adopt stronger meritocratic beliefs (Hess & Ledgerwood, 2014). Indeed, there is evidence to suggest that political ideological beliefs are malleable and can change depending on the present company and context, with close relationships possibly exerting a particularly strong influence on the extent to which conversative versus liberal ideals are endorsed (Jost et al., 2008). This research supports the notion that needs and desires to maintain a shared reality with others can lead individuals to adjust their ideologies towards the status quo.

It has been suggested that, amongst advantaged individuals, system justification motives are consistent with ego and group justification motives, such as motives to maintain personal and group self-esteem and to favour one's ingroup (Jost & Thompson, 2000), whilst the same is not true amongst disadvantaged individuals (Zimmerman & Reyna, 2013). Though the psychological consequences of system justification amongst disadvantaged individuals can therefore be negative, net positive effects have also been documented (e.g., Bahamondes-Correa, 2016; Suppes et al., 2019), indicating that justifying motives amongst disadvantaged individuals can paradoxically cause harm and provide relief contemporaneously.

A specific form of justifying ideology, particularly relevant in Western, industrialised societies, is Economic System Justification (Jost & Thompson, 2000). Economic justifiers believe strongly that capitalism provides equal opportunities for all. It follows that economic justifiers endorse the belief that economic status is a product of personal merit; that is, successful individuals are successful because they worked for it, and unsuccessful individuals are unsuccessful because they didn't. Put succinctly, this amounts to believing that people deserve what they get and get what they deserve. Economic System Justification therefore represents a specific form of justifying belief – conceptually distinct from General System Justification – that is particularly relevant and salient in Western capitalist society.

In the context of seeking information about payment equality, the above findings suggest that the extent to which individuals tend to legitimise and justify existing social arrangements may be an important predictor of whether they seek equality information. For example, conservative viewpoints may encourage individuals to avoid information that could present the system in a bad light (Hennes et al., 2012; Jost et al., 2017). Conversely, since strong justifiers tend to have stronger epistemic needs (e.g., Jost et al., 2003, 2007), they may be instead be *more* likely to seek information about inequality, even if that information might reflect negatively on the system. Additionally, because advantaged and disadvantaged individuals may have differing motives for justifying the status quo (Bahamondes-Correa, 2016; Jost, 2017; Jost & Thompson, 2000), their attitudes toward seeking information may also diverge.

**Belief in a Just World**

Conceptually similar to System Justification Theory, the Just World hypothesis (also known as Belief in a Just World; BJW) argues that individuals have an intrinsic need to believe that people tend to get what they deserve. Such a belief permits them to believe in turn that the world is inherently stable and not chaotically random. This serves a palliative function insofar as it bestows predictability upon the world and thus affords individuals a sense of control (delusory or illusory) over what happens to them (Dalbert, 1999; Furnham, 2003; Lerner, 1980). Because of this palliative effect, individuals are motivated to defend their BJW by interpreting events and information in a way that conforms with it (Bartholomaeus & Strelan, 2019; Lerner & Miller, 1978).

Converging evidence indicates that BJW can be conceptualised as bidimensional (Dalbert, 1999); as well as the general component (BJW-general) that was identified in early research as involving mostly negative appraisals of others to justify the occurrence of unfortunate events (Lerner & Simmons, 1966), a personal component (BJW-personal) involving positive appraisals of the justice of the world as it applies to oneself was identified (Dalbert, 1999). The two components are consistently found to be two distinct aspects of BJW (Dalbert, 1999; Sutton & Douglas, 2005; Sutton et al., 2008), and BJW-personal is frequently reported as the stronger of the two (Bègue & Bastounis, 2003; Sutton & Douglas, 2005; Sutton et al., 2008).

BJW-personal has been linked with a number of positive effects that appear to help individuals cope with adversity. For example, individuals high in BJW-personal score more favourably on well-being measures such as psychological distress and life satisfaction in the wake of natural disasters (Otto et al., 2006; Şeker, 2016; Wu et al., 2011). BJW-personal can also buffer individuals against the negative psychological effects of longer-term challenges such as unemployment (Dzuka & Dalbert, 2002) and career insecurity (Nudelman et al., 2016) and is positively associated with job satisfaction and mental health (Otto et al., 2009). BJW-personal is positively related to pro-social behaviours such as social affiliation (Nudelman, 2013; Sutton et al., 2017), altruism (Bègue et al., 2008), and forgiveness (Strelan, 2018) and negatively related to anti-social behaviours such as criminal behaviour (Sutton & Winnard, 2007) and bullying (Correia & Dalbert, 2008). In the present experiment, it was therefore hypothesised that individuals high in BJW-personal would be more likely to find out about possible inequality, and that this would be the case for both potentially advantaged and potentially disadvantaged individuals.

BJW-general has been explored in relation to inequality-justifying beliefs, collective efficacy, and justice-seeking behaviour. Beierlein et al. (2011) found that BJW was positively related to justification of inequality and negatively related to justice-seeking behaviours; individuals more likely to defend the notion that the world is a just place (high BJW) are also more likely to appraise inequalities as justified and less likely to engage in justice-oriented activities. Interestingly, the positive influence of BJW on justification of inequality was present only for individuals scoring low in perceptions of collective political efficacy; for individuals who believed more strongly in the collective efficacy of their group to enact justice-promoting behaviours, BJW was unrelated to justification of inequality. This supports the notion that when it is felt that seeking justice is impossible, BJW acts as a palliative buffer to render the unchanging injustice more acceptable (Lerner & Simmons, 1966). I therefore predicted that individuals high in BJW-general, like strong justifiers and in contrast to high BJW-personal individuals, would be less likely to find out about (in)equality in the present experiment.

**Social Value Orientation**

An individual's Social Value Orientation (SVO) describes the extent to which they are concerned with others' outcomes relative their own (Liebrand, 1984; Messick & McClintock, 1968; Murphy & Ackermann, 2014; Van Lange et al., 1997). Modern measurements of SVO characterise it as a continuum ranging from a "purely competitive" orientation to a "purely altruistic" one, with certain ranges characterised as representing "purely individualistic" and "purely prosocial" orientations (Murphy & Ackermann, 2014; Murphy et al., 2011). In most research, the predominant distinction is between "pro-social" and "pro-self" individuals. Pro-socials value equality or sometimes even value others' outcomes more than their own, whereas pro-selfs prioritise their own outcomes, either through indifference to the outcomes of others or by actively striving to maximise their own gain relative to others' (Smeesters et al., 2003). As might be expected, pro-socials are consistently documented to be more cooperative than pro-selfs across a range of experimental social dilemma scenarios, such as public goods games (De Cremer & Van Lange, 2001; Smeesters et al., 2003), resource dilemmas (Kramer et al., 1986), and coordination dilemmas such as the Prisoner's Dilemma (Balliet et al., 2009). In the context of real-world dilemmas, pro-socials are found to be more willing to adopt pro-environmental behaviours than pro-selfs (Joireman et al., 2001), such as taking public transport (Van Vugt et al., 1995).

There is evidence to suggest that SVO shapes how individuals attend to information. Pro-social individuals tend to search for information consistent with their cooperative beliefs, whilst pro-self individuals tend to seek information consistent with their competitive beliefs (Camac, 1992; Van Kleef & De Dreu, 2002). Evidence from functional magnetic resonance imaging (fMRI) suggests that pro-social individuals engage in more information processing (as indexed by stronger activation in dorsolateral prefrontal cortex, posterior superior temporal sulcus, and precuneus) than pro-selfs (Haruno & Frith, 2010). Using an eye-tracking paradigm, Fiedler et al. (2013) showed that deviation from pure individualism predicted greater consideration of information, greater attention to others' outcomes, and longer decision times. This was the case for both deviation toward cooperativeness and deviation toward competitiveness, demonstrating that the action of cooperative and competitive SVO lies in the consideration of information about others.

This research demonstrates that an individual's SVO is relevant to the type of information they seek and the extent to which they consider it. Because pro-socials appear to attend to and process information more than pro-selfs (e.g., Haruno & Frith, 2010), it might be hypothesised that prosocial individuals will be more likely to seek information about inequality compared to pro-self individuals. Again, an interesting question concerns whether pro-sociality has a difference influence on information-seeking depending on whether the individual is potentially advantaged or potentially disadvantaged by the inequality.

**Life Orientation**

Life Orientation (LO) is a measurement of an individual's dispositional optimism, which itself can be understood as the extent to which individuals believe that the future is positive (Carver et al., 2010). Generally, optimistic individuals report higher levels of subjective well-being (Scheier et al., 2001) and are better able to adjust to changing or stressful situations (e.g., Brissette et al., 2002; Carver et al., 2005; Carver & Gaines, 1987; Daukantaite & Bergman, 2005) compared to pessimistic individuals. Scholars of optimism have conceptualised it within a framework of self-regulation of behaviour based on expectancies of a given scenario. Positive expectancies (i.e., positive predictions about how a situation can develop) lead to increased effort and engagement, whilst negative expectancies (i.e., appraisals that predict a negative outcome in a situation) lead to reduced effort and disengagement (Carver et al., 1983; Nes et al., 2005). The behavioural response elicited from positive expectancies is termed *approach coping* – the goal of which is to directly address the demands of a challenging situation through explicit action – and the behavioural response elicited from negative expectancies is termed *avoidance coping* – the goal of which is to manage the situation by avoiding or ignoring the problem (Endler & Parker, 1990; Suls & Fletcher, 1985). Optimism is typically found to be positively related to approach strategies and negatively related to avoidance strategies (Carver et al., 2010; Carver & Scheier, 2014). Thus, in general, optimists deal with challenging situations by addressing them directly (approach coping), whilst pessimists tend to adopt avoidance coping strategies such as choosing not to acknowledge problems (Carver et al., 1989; Nes & Segerstrom, 2006). In the context of information-seeking, a clear prediction is therefore that more optimistic individuals will be more likely than less optimistic individuals to approach information rather than avoid it.

**Intolerance of Uncertainty**

The way in which individuals are disposed to react in uncertain situations is likely to influence the likelihood that they will choose to seek further information in a given scenario.

Intolerance of uncertainty (IU) has been defined as “a dispositional characteristic that arises from a set of negative beliefs about uncertainty and its connotations and consequences” (Koerner & Dugas, 2008, pp. 631). Individuals high in IU experience uncertainty as more stressful and threatening than those low in IU (Greco & Roger, 2003; Koerner & Dugas, 2008). This can lead to “uncertainty paralysis” (Birrell et al., 2011), but can also lead high IU individuals to engage in information-seeking behaviours to decrease the stressful effects of uncertainty (Rosen et al., 2007; Rosen & Knäuper, 2009). Moreover, there is evidence to suggest that the extent to which IU motivates information seeking versus avoidance may depend on situational uncertainty, which occurs when the context is poorly defined, ambiguous, unpredictable and/or lacking information (Brouwers & Sorrentino, 1993; Rosen & Knäuper, 2009). In the context of information about health, individuals high in IU under conditions of high situational uncertainty tend to seek the most information, whereas those high in IU under conditions of low situational uncertainty tend to seek relatively less information (Rosen & Knäuper, 2009). This finding demonstrates an interesting duality whereby those intolerant of uncertainty are more likely to seek information if the situation is perceived to be uncertain but less likely to seek information if the situation is perceived to be certain. In the present experiment, a natural hypothesis is that individuals highly intolerant of uncertainty will be more likely to seek fairness information, compared to those more tolerant of uncertainty.

My aim in the present study was to identify what proportion of individuals seek information about whether they had been paid more or less than another individual who completed the same task, and to explore whether and how these decisions were influenced by a selection of individual difference characteristics: System Justification, Belief in a Just World, Life Orientation, and Intolerance of Uncertainty. As discussed above, these individual difference characteristics were chosen because they measure motivations and ideologies thought to be relevant for information-seeking in the social justice domain. Namely, these characteristics speak to generalised concerns for justifying the way things are as the way they should be, belief in the world as inherently just, acceptance of uncertainty, dispositional optimism/pessimism, and attitudes towards the division of wealth and meritocracy. I also asked participants to provide several ratings of their experience, such as how comfortable they were not knowing and the extent to which they felt associated (i.e., identified) with others in the experiment, and related these ratings to their individual difference scores. The results revealed a constellation of findings that provide insight into which character traits are particularly relevant for seeking and avoiding information about social outcomes and how potential advantage or potential disadvantage moderates the influence of these traits.

# Method

The experiment was initially carried out in person. Due to the onset of the coronavirus pandemic this first experiment was discontinued after collecting data for 64 participants, and an online version of the experiment was subsequently developed. The in-person data, though not analysed in the present study, was used to inform power analyses to determine an appropriate sample size for the present study. The smallest effect of interest observed in the in-person data was for the Economic System Justification scale, with odds ratio 0.53. Using GPower (Faul et al., 2007), I determined that to achieve an effect of this size with 80% power, at least 84 participants would be required. Because I was initially interested to explore effects for disadvantaged and advantaged participants separately, I also identified the size of this effect for each sub-group. The smallest effect was an odds ratio of 0.4, and power analysis indicated that to detect this effect with 80% power, at least 62 participants would be required in each sub-group (i.e. 124 participants total). I thus aimed to recruit at least 130 participants for the online version of the experiment presented in this thesis.

## Participants

One-hundred-and-thirty-five participants (66 female, 68 male, 1 undisclosed) were recruited via Prolific (https://www.prolific.co/) and the experiment was administered on Qualtrics (https://www.qualtrics.com/) during the period March - June 2020. The following filters were implemented to recruit participants:

- Age: 18 years old to 64 years old. This range was chosen to capture data for working-age individuals (a minimum age of 18 was necessary because ethical approval for the experiment was for testing adults and not minors).
- Nationality: United Kingdom

- Maximum income: £50,000.00. This value was chosen as a cut-off because it approximates the upper boundary for the basic income tax rate of 20% (£50,270). Higher rate incomes were excluded to avoid differences between individuals in the perceived value of the payments in the experiment.

The mean age was 32.44 years (standard deviation [SD] = 11.71). Participants' average income was £22,678.93 (SD = £10545.33). Participants varied in ethnicity, though the majority reported themselves as 'White' (***Table 3.1***).

**Table 3.1.** Participant ethnicity information

| *Ethnicity* | *Frequency* | *Percentage* |
|---|---|---|
| Undisclosed | 21 | 15.56 |
| African | 3 | 2.22 |
| Arab | 1 | 0.74 |
| Asian | 8 | 5.93 |
| Bangladeshi | 1 | 0.74 |
| Black | 1 | 0.74 |
| Black African | 3 | 2.22 |
| Black Caribbean | 1 | 0.74 |
| White | 92 | 68.15 |
| Chinese | 1 | 0.74 |
| Mixed | 2 | 1.48 |
| Pakistani | 1 | 0.74 |
| ***Total*** | **135** | |

*Note.* Listed ethnicities are those disclosed by participants and have not been categorised by the researcher

## Materials

### *Individual Differences Questionnaires*

I took measurements of several individual difference measures, which served as the primary independent variables of interest in the experiment. The full measurement scales can be found in ***Appendix A*: Individual difference measurement scales used in Study 3**

**System Justification Scale – General.** The general system justification scale (SJG; Kay & Jost, 2003) measures the extent to which individuals support the way that their society currently is. It consists of 8 questions and asks participants to indicate on a scale of 1 (*strongly agree*)-9 (*strongly disagree*) their agreement with statements such as "In general, the political system of my country operates as it should."

**System Justification Scale – Economic.** The economic system justification scale (SJE; Jost & Thompson, 2000) measures the extent to which individuals endorse the ideology that the capitalist economic system provides equal opportunities for all citizens, and that individuals' economic status is a product of personal merit. It consists of 17 questions and asks participants to indicate using the same 1-9 scale as above the degree to which they agree with statements such as "It is virtually impossible to eliminate poverty." This measure was included in addition to the General System Justification scale in order to delineate general justifying ideologies that relate to managing experiential challenges from meritocracy-focused ideologies that relate more strongly to individual effort and monetary value.

**Life Orientation Revised.** The Life Orientation scale (LO; Carver et al., 2010; Scheier et al., 1994) measures individuals' trait optimism across 10 questions. Participants rate their agreement on a scale of 1 (*I disagree a lot*) to 5 (*I agree a lot*) with items such as "In uncertain times, I usually expect the best."

**Intolerance for Uncertainty.** Intolerance of Uncertainty (IU; Carleton et al., 2007) is a 12-question scale that measures the degree to which participants accept uncertainty in their lives. Participants were asked to respond on a scale of 1 (*not at all characteristic of me*) to 5 (*entirely characteristic of me*) to statements such as "I can't stand being taken by surprise."

**Social Value Orientation.** Social Value Orientation (SVO; Murphy et al., 2011) measures preferences for how wealth is distributed between oneself and another to provide a measure of individuals' concern for others. Each item of the scale asks participants to choose one option out of nine for dividing wealth between oneself and an imagined other person. For example, one item asks participants to choose one option for sharing money from the following choices: option 1: 85 to self and 85 to other, option 2: 85 to self and 76 to other, option 3: 85 to self and 68 to other, option 4: 85 to self and 59 to other, option 5: 85 to self and 50 to other, option 6: 85 to self and 41 to other, option 7: 85 to self and 33 to other, option 8: 85 to self and 24 to other, option 9: 85 to self and 15 to other. There are 15 items in total, and the range of distributions to choose from varies between items (see ***Appendix A*** for the full scale).

**Belief in a Just World.** Belief in a Just World (BJW; Dalbert, 1999) measures participants' endorsement of beliefs about fairness in the world. Participants respond on a scale of 1 (*strongly disagree*) to 6 (*strongly agree*) to statements such as "Overall, events in my life are just." There are two subscales of BJW; a 7-item general subscale (BJW-General) that measures individuals' beliefs about the fairness of the world generally, and a 6-item personal subscale (BJW-Personal) that measures individuals' beliefs about the fairness of things as they apply to them as an individual.

***Experimental Tasks***

**Face-rating 'Work' Task.** I created a face-rating task designed to make participants feel that they were engaging in work and not just a research experiment. The task consisted of rating a series of faces on how trustworthy and dominant they were thought to be. Participants were told that the task should be considered work because their ratings would contribute to stimulus selection in a later experiment conducted by the lab. Participants were presented with 24 different identities, each of which was rated once for trustworthiness and once for dominance. For each rating, participants were presented with the text "Rate how dominant [trustworthy] you feel this person is on a scale of 1 (extremely submissive/not dominant [extremely untrustworthy]) to 7 (extremely dominant [extremely trustworthy])." The face was presented below this text with resolution 800 x 800 pixels. Participants indicated their rating using a 7-point slider positioned below the face image. For the dominance ratings, text at the left extremity (corresponding to a rating of '1') read "*Extremely submissive*" whilst text at the right extremity (corresponding to a rating of '7') read "*Extremely dominant*". For the trustworthiness ratings, the text at the left extremity read "*Extremely untrustworthy*" and the text at the right extremity read "*Extremely trustworthy*." The order in which faces were rated was randomised for each participant and each face was rated separately for trustworthiness and dominance (i.e., both traits were not rated in the same trial, resulting in 48 distinct rating trials).

**Post-work Ratings.** Following the work task participants progressed through a series of questions and information screens in the following order.

***Payment information.*** Participants were told how much money they would receive at the end of the experiment as payment for their work. This could either have been the payment stated in the recruitment materials (£8.00), or double this amount (£16.00; see *"Information about possible unfairness"* below).

***Work effort – Time 1.*** Using a Visual Analogue Scale (VAS) slider, participants were asked "*How much effort was required to complete the work?*" The left extremity of the slider read "*No effort*" and the right extremity read "*A lot of effort*."

***Satisfaction with payment – Time 1.*** Participants used a VAS slider to answer the question "*How satisfied are you with the amount you will be paid for the work?*" The left extremity of the slider read "*Completely dissatisfied*" and the right extremity read "*Completely satisfied*."

***Information about possible unfairness.*** This information screen introduced participants to the possibility of unfair payment. The information differed slightly, depending on whether the participant was advantaged or disadvantaged (indicated in square brackets below). The information read:

> "You and your partner have completed the same work. It is possible that you will both be paid the same amount for your work. Your partner has a lower [higher] household income than you. Based on this fact, it is also possible that you are [your partner is] advantaged and will be [has been] paid twice as much as your partner [as you] for the work (i.e., your partner is [you are] disadvantaged and has been [will be] paid half as much as you [your partner]). You are not able to change anything. However, if you want you can find out whether you and your partner are paid the same or different amounts. **Note that it is impossible for you to know based on your payment alone whether you and your partner are paid equally**."

It was important to emphasise to participants that it was impossible to deduce the fairness of the system based on the payment information they received earlier.

***Decision.*** Participants were then instructed to decide either to find out or not find out about whether they had been paid the same as their virtual partner or differently. The text read:

> "You can now decide either to find out whether you and your partner are paid the same or differently, or to not find out. Before making your decision, take a moment to think about how you would you feel if you choose to find out, and how you would feel if you choose not to find out. When you are ready, make your decision. (Your decision has no impact on how long the experiment will take to complete.)"

Participants could select one of two radio buttons to indicate their choice. The first read "*Find out*" and the second read "*Don't find out*." Once the decision was made, those who decided to find out were told that the information about equal pay would be provided after some more ratings. Those who decided not to find out were simply asked to continue with some more ratings.

***Satisfaction with payment – Time 2.*** This was the same VAS scale as at Time 1. The difference between this rating and the Time 1 rating provided a measurement of the degree to which satisfaction changed between having just been paid and having found out about the possibility of unequal payment.

***Comfort not knowing about unfairness.*** Participants were asked "*How comfortable do you feel about not knowing whether you and your partner are paid the same or differently?*" and indicated their response using a VAS slider. The left extremity of the slider read "*Extremely uncomfortable*" and the right extremity read "*Perfectly comfortable*."

***Comfort with unfairness.*** Participants were asked "*How comfortable would you be to know that you and your partner are paid differently?*" and indicated their response using a VAS slider. The left extremity of the slider read "*Extremely uncomfortable*", and the right extremity read "*Perfectly comfortable*."

***Reasons for decision.*** A series of VAS sliders were implemented to gather information on the extent to which participants' decisions were motivated by reasons that were hypothesised to be important. A separate VAS slider was provided for each hypothesised motivation. These motivations were: "*Reduce anxiety from not knowing*", "*Reduce anxiety from unequal pay*", "*Reassurance of being well paid*", "*Reassurance of equal pay*", and "*Reassurance that the situation was fair*." Disadvantaged participants also had the option "*Avoid feeling disadvantaged*" and Advantaged participants had the extra option "*Avoid feeling guilty*."

Following these VAS sliders, participants were invited to provide any other reasons that were important for their decision. They did this with a text entry box where they were asked to list as many reasons as they liked, and to provide a number between 1 and 100 to indicate the extent to which their decision was motivated by that reason.

***Willingness to pay.*** Participants were asked to indicate how much of their payment they would be willing to sacrifice in order to fulfil their decision. The question read "*How much of your payment would you be willing to pay to [not] find out?*" Participants responded by adjusting a VAS slider to the desired position. The left extremity read "*£0.00*" and the right extremity read either "*£8.00*" or "*£16.00*" depending on how much the participant was paid.

***Identification – Time 1.*** Participants were then asked the extent to which they experienced a shared identity with the other participants in the experiment. The question text read:

> "Every participant taking part in this experiment experiences the same thing as you; in every pair there is one potentially advantaged participant and one potentially disadvantaged participant. Below, you are asked to indicate how much you identify with the other participants in the experiment (i.e., the degree to which you feel associated - in any way - with the other participants in the experiment)."

Participants responded to three statements using 7-point Likert scale from ("*Strongly disagree*" – "*Strongly agree*"). The statements were: "*I identify with my partner*", "*I identify with the disadvantaged participants in the experiment*", and "*I identify with the advantaged participants in the experiment*."

***Belief in the probability of fairness.*** Participants indicated what they thought was the likelihood that payments were equal on a VAS scale. The left extremity read "*0%*", the right extremity read "*100%*", and the mid-point read "*50%*".

***Preferred distribution of payments.*** Participants were asked "*If you could choose the relationship between your payment and your partner's, what would it be?*" and responded using a VAS slider. The left extremity read "*I receive half the amount my partner receives, i.e., my partner receives double the amount I receive*", the right extremity read "*I receive double the amount my partner receives, i.e., my partner receives half the amount I receive*" and the mid-point read "*We both receive the same amount*."

***Fairness reveal.*** At this point, if the participant decided to find out, information about the payments was revealed. If payments were equal, the text read "*You and your partner will be paid the same for the work. You will receive £8.00 and your partner received £8.00*." If payments were unequal, the text for advantaged participants read "*You will be paid more than your partner for the work. You will receive £16.00 and your partner received £8.00*." and the text for disadvantaged participants read "*You will be paid less than your partner for the work. You will receive £8.00 and your partner received £16.00*." Participants who chose not to find out were not provided with any information.

***Satisfaction with payment – Time 3***. Participants used the same VAS scale as at Times 1 and 2. For participants who chose to find out, the difference between this rating and the Time 1 rating provided a measurement of the degree to which satisfaction changed from having been paid to having found out about (un)equal payment, and the difference between this rating and the Time 2 rating provided a measurement of the degree to which satisfaction changed from having found out about possible unequal payments to finding out about (un)equal payments. For participants who chose not to find out, the difference between this rating and the Time 1 rating should theoretically have been the same as the difference between the Time 2 and Time 1 ratings, since these participants received no further information relating to payments.

***Satisfaction with decision.*** Participants responded to the question "*How satisfied are you with your decision to [not] find out?*" using a VAS slider. The left extremity of the slider read "*Completely dissatisfied*" and the right extremity read "*Completely satisfied*."

***Identification – Time 2.*** Participants were asked again to indicate their identification with others in the same way as at Time 1. This measure was designed to capture any possible changes in identification having found out about the payments.

***Concern with social inequality.*** Finally, participants were asked "*How often do you think about social inequality?*" and responded on a 7-point Likert scale ("*Never*", "*At least once a year*", "*Several times a year*", "*At least once a month*", "*At least once a week*", "*Several times a week*", and "*At least once a day*"). Following this, they were asked "*How much does social inequality bother you when you think about it?*" and responded using a VAS slider. The left extremity of the slider read "*It doesn't bother me at all*" and the right extremity read "*It bothers me a lot*."

## Design

The design was primarily trait-based to explore how individual difference characteristics relate to decisions and experiences in the experiment. In addition, I implemented a two-level factor called "Status", which described whether participants were potentially advantaged or potentially disadvantaged relative to their virtual partner.

**Procedure**

Once recruited via Prolific, participants were redirected to the Qualtrics platform to complete the experiment. An instructional video explained the experiment and introduced the first part, in which participants first completed the individual differences questionnaires. Once this was complete, participants were introduced to the 'work' task with another video. It was emphasised to participants that this work was non-trivial as the ratings provided would help guide the lab's decisions on stimuli to use in future experiments. Participants started the work when they were ready to.

Following the work, another instructional video explained the final part of the experiment. It was explained to participants that they would next answer some questions relating to the work. Following the initial ratings, participants were informed about the possible payment inequalities (see *Experimental Tasks* section above). They were told whether, based on their reported household income, they had been classified as advantaged or disadvantaged. Participants were told that they had been paired with a participant who had already completed the experiment; for approximately half the participants this previous participant had a lower income than the participant, for the other half the previous participant had a higher income than the participant. If the current participant's income was the lower income, they were classified as disadvantaged. If the current participant's income was the higher income, they were classified as advantaged. To maintain the deception-free ethos of the experiment, efforts were made to make this pairing technically true. This was achieved by taking as potential partners one previous participant who had a very low income (£5,000.00) and one previous participant with a very high income (£300,000.00 – from the pilot study). For each participant, there was a 50% chance of being paired with the low-income partner and a 50% chance of being paired with the high-income partner. The current participant's income was then compared to the income of the partner and the participant was thus classified as either disadvantaged or advantaged. In this way, it was possible to obtain approximately equal numbers of disadvantaged and advantaged participants whilst maintaining the veracity of the information communicated to participants. Once this section was complete, participants were redirected to Prolific and the experiment was concluded.

## Results

### *Note on terminology*

For ease of reading, in this Results section I sometimes refer to participants who were potentially advantaged (i.e., those who, if the payments were unequal, would be paid *more* than their partner) as “Advantaged”. I refer to participants who were potentially disadvantaged (i.e., those who, if payments were unequal, would be paid *less* than their partner) as “Disadvantaged”. The variable that distinguishes participants based on their potential advantage or disadvantage is referred to as “Status”. The reader is encouraged to keep in mind that these terms are used for convenience, and that when “Advantaged” or “Disadvantaged” is used it should always be understood to mean “potentially advantaged” or “potentially disadvantaged”. In addition, the term “Status” only ever refers to this categorisation, and never refers to other connotations of status, such as socioeconomic status.

***Table 3.2*** presents frequency data for participants’ decisions. Overall, 56% of participants chose to find out and 44% of participants chose not to find out. ***Figure 3.1*** shows that, within each Status, decisions to find out were more common than decisions not to find out. ***Table 3.3*** shows the descriptive statistics and correlations amongst the continuous variables measured.

**Table 3.2.** *Frequency data showing the number and percentage of decisions to find out and not find out for potentially advantaged and potentially disadvantaged participants*

| | Find out | | Don't find out | | Total | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *n* | % |
| Potentially advantaged | 33 | 24% | 25 | 19% | 58 | 43% |
| Potentially disadvantaged | 43 | 32% | 34 | 25% | 77 | 57% |
| Total | 76 | 56% | 59 | 44% | 135 | |

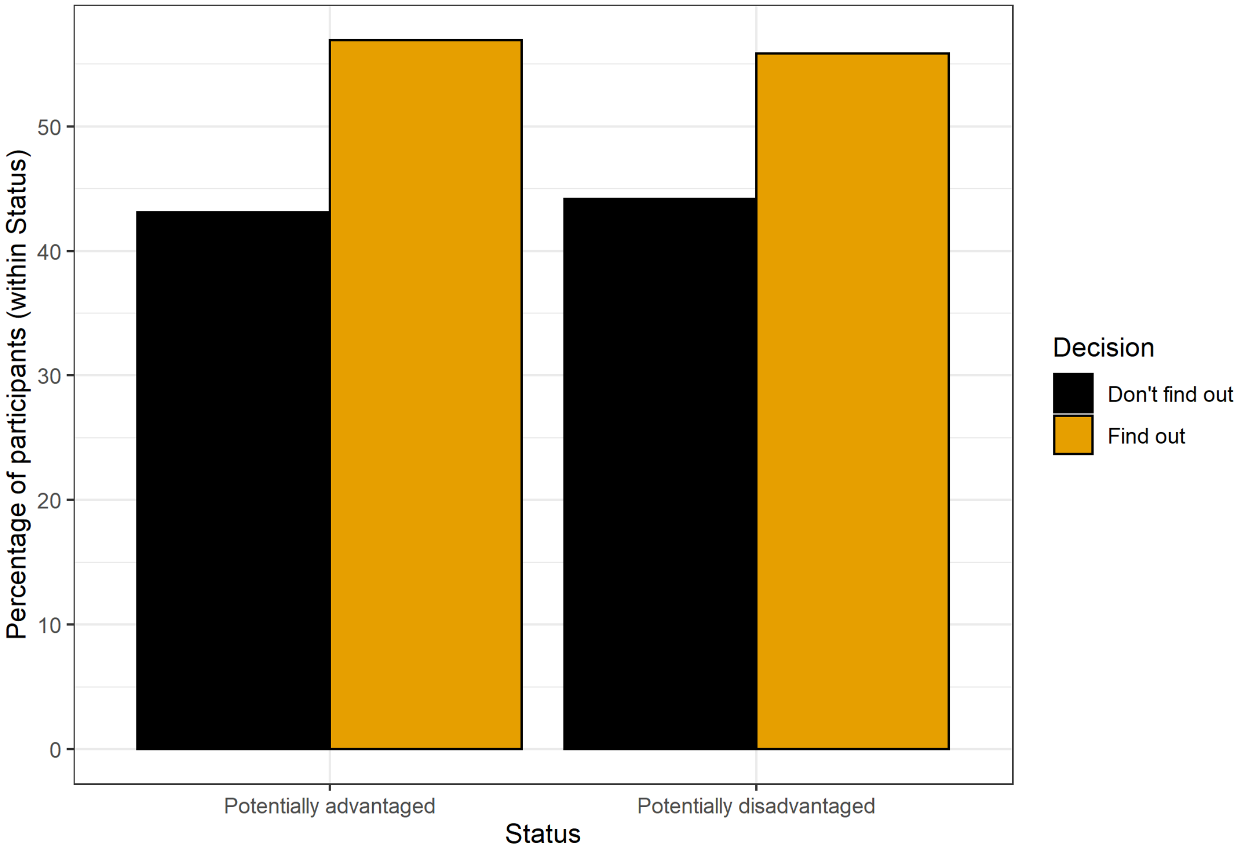


*Figure 3.1.* Percentage of participants within each Status who decided to find out and not find out.

I report the results in sections, owing to their volume. In **Section 1** I report the results of analyses predicting decisions to find out or not find out from individual differences and situational factors (i.e., "Comfort not knowing" and "Comfort with Unfairness"). In **Section 2**

I report the results of analyses predicting Comfort not knowing from individual difference measures. In **Section 3** I report on participants' reported motivations for their decisions, including qualitative comments provided by a subset of participants.

**Table 3.3.** *Correlations amongst individual difference measurements*

| | *Age* | *SJG* | *SJE* | *SVO* | *BJW-G* | *BJW-P* | *LO* | *IU* |
|---|---|---|---|---|---|---|---|---|
| *Age* | | | | | | | | |
| *System Justification General (SJG)* | 0.068 *(.433)* | | | | | | | |
| *System Justification Economic (SJE)* | -0.048 *(.581)* | 0.571 *(<.001)* | | | | | | |
| *Social Value Orientation (SVO)* | -0.089 *(.306)* | -0.022 *(.800)* | -0.176 *(.041)* | | | | | |
| *Belief in a Just World General (BJW-G)* | -0.099 *(.254)* | 0.659 *(<.001)* | 0.551 *(<.001)* | -0.075 *(.389)* | | | | |
| *Belief in a Just World Personal (BJW-P)* | -0.131 *(.129)* | 0.516 *(<.001)* | 0.308 *(<.001)* | 0.106 *(.221)* | 0.484 *(<.001)* | | | |
| *Life Orientation (LO)* | 0.084 *(.333)* | 0.376 *(<.001)* | 0.065 *(.457)* | 0.006 *(.942)* | 0.251 *(.003)* | 0.504 *(<.001)* | | |
| *Intolerance of Uncertainty (IU)* | -0.193 *(.025)* | -0.069 *(.429)* | 0.131 *(.131)* | -0.066 *(.444)* | 0.113 *(.194)* | 0.023 *(.794)* | -0.229 *(.008)* | |
| Mean | 32.44 | 31.89 | 74.26 | 29.72 | 18.64 | 26.92 | 17.40 | 34.86 |
| SD | 11.71 | 11.42 | 15.92 | 12.26 | 5.67 | 6.51 | 5.34 | 10.04 |
| Minimum | 18.00 | 8.00 | 36.00 | -12.23 | 7.00 | 11.00 | 6.00 | 13.00 |
| Maximum | 64.00 | 60.00 | 110.00 | 61.39 | 30.00 | 42.00 | 30.00 | 60.00 |

The goal of analyses was to characterise the best fits of a variety of individual differences to behaviour. This was achieved through a process of fitting a saturated model including all posited terms (including interaction terms) and in a stepwise manner reducing the model until a best-fitting model was identified. Happily, the *buildmer* function in the "buildmer" package in R allows the user to specify a model with all posited terms and identifies the highest-saturated model that converges. From this point the function removes terms from the model until it arrives at the best-fitting model. In all analyses presented here, the criterion used to determine model fit was the Akaike Information Criterion (AIC).

## Section 1: Discovery Decisions

### *Individual Differences*

I started by investigating the factors that directly influenced individuals' binary decision to find out using generalised linear models. The terms passed to *buildmer* were:

**Table 3.4.** *Terms entered into buildmer function*

| |
|---|
| Age |
| Sex |
| Status |
| System Justification General (SJG) |
| System Justification Economic (SJE) |
| Belief in a Just World General (BJW-general) |
| Belief in a Just World Personal (BJW-personal) |
| Life Orientation (LO) |
| Intolerance of Uncertainty (IU) |
| Social Value Orientation (SVO) |
| Age x Sex |
| Age x Status |
| Sex x Status |
| Age x SJG |
| Age x SJE |
| Age x BJW-general |
| Age x BJW-personal |
| Age x LO |
| Age x IU |
| Age x SVO |
| Sex x SJG |
| Sex x SJE |
| Sex x BJW-general |
| Sex x BJW-personal |
| Sex x LO |
| Sex x IU |
| Sex x SVO |
| Status x SJG |
| Status x SJE |
| Status x BJW-general |
| Status x BJW-personal |
| Status x LO |
| Status x IU |
| Status x SVO |

The best-fitting model explained 42.49% of the variance in decisions to find out, $\chi^2$ (22, 135) = 78.61, $p$ < .001. As can be seen in ***Table 3.5***, this model contained many terms.

**Table 3.5.** *Parameter estimates for the best-fitting model resulting from the modelling procedure*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| (Intercept) | 0.00 | 0.00 – 0.06 | **0.027** |
| Age | 1.81 | 1.17 – 3.03 | **0.014** |
| System Justification General (SJG) | 0.72 | 0.58 – 0.85 | **0.001** |
| Life Orientation (LO) | 0.79 | 0.67 – 0.90 | **0.001** |
| System Justification Economic (SJE) | 1.02 | 0.96 – 1.08 | 0.547 |
| Intolerance for Uncertainty (IU) | 1.13 | 0.92 – 1.41 | 0.262 |
| Social Value Orientation (SVO) | 1.57 | 1.23 – 2.10 | **0.001** |
| Belief in a Just World – Personal (BJW-p) | 0.64 | 0.43 – 0.90 | **0.016** |
| Belief in a Just World – General (BJW-g) | 3.29 | 1.80 – 7.01 | **0.001** |
| Status† | 3026806.99 | 1526.86 – 3099434<br>1011.19 | < **0.001** |
| Sex†† | 4220427.28 | 305.28 – 26792492<br>5984.02 | **0.003** |
| Age x IU | 0.99 | 0.98 – 1.00 | **0.037** |
| Age x SVO | 0.99 | 0.99 – 1.00 | **0.012** |
| Age x BJW-p | 1.02 | 1.01 – 1.03 | **0.006** |
| Age x BJW-g | 0.98 | 0.96 – 0.99 | **0.003** |
| Status x SVO | 0.77 | 0.66 – 0.87 | < **0.001** |
| Status x Age | 0.77 | 0.66 – 0.87 | < **0.001** |

| | | | |
|---|---|---|---|
| Status x SJG | 1.63 | 1.32 – 2.14 | **< 0.001** |
| Status x BJW-g | 0.47 | 0.30 – 0.68 | **< 0.001** |
| Sex x SJG | 1.42 | 1.21 – 1.73 | **< 0.001** |
| Sex x Age | 0.74 | 0.62 – 0.86 | **< 0.001** |
| Sex x SJE | 0.85 | 0.75 – 0.94 | **0.003** |
| Sex x SVO | 0.85 | 0.74 – 0.95 | **0.009** |
| Observations | 135 | | |

*Note.* † Reference group = Advantaged; Treatment group = Disadvantaged

†† Reference group = Female; Treatment group = Male

The model indicated many effects relevant to the decision to find out. Since each effect identified in this model was statistically contingent on the presence the other effects, interpretation and inference from this model alone was not appropriate. Accordingly, I investigated whether each effect identified in the overall model existed independently by conducting a series of regression analyses containing only the effect of interest. This was achieved by starting with a base model containing Age, Sex, and Status, and then adding a term and its interaction with each of Age, Sex, and Status in separate models.

The results of this procedure showed that the main effect of Age (***Table 3.6***), the interactions of Status with each of Social Value Orientation (***Table 3.7***), System Justification General (***Table 3.10***), and System Justification Economic (***Table 3.11***) were independently influential in predicting decisions to find out about unequal pay.

**Age.** The base model explained 5.57% of the variance in decisions to find out, $\chi^2(3, 135) = 10.3$, $p = 0.016$. Age was a significant negative predictor of the likelihood of finding out (*OR* = 0.95, $z$ = -3.07, $p$ = 0.002); older participants were significantly less likely to find out compared to younger participants (***Table 3.6***).

**Table 3.6.** *The base model predicting decisions to find out from Age, Sex, and Status*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| (Intercept) | 6.29 | 1.91 – 22.30 | **0.003** |
| Age | 0.95 | 0.92 – 0.98 | **0.002** |
| Sex† | 1.24 | 0.61 – 2.56 | 0.550 |
| Status†† | 0.88 | 0.42 – 1.81 | 0.728 |
| Observations | 135 | | |

*Note.* †Reference group = Female; Treatment group = Male

†† Reference group = Advantaged; Treatment group = Disadvantaged

**Status by Social Value Orientation.** The Social Value Orientation (SVO) model explained 8.7% of the variance in decisions to find out, $\chi^2(5, 135) = 16.1$, $p$ = 0.007. Trend analysis of the Status by SVO interaction indicated that the significant difference in the influence of SVO between Advantaged and Disadvantaged participants (*OR* = 1.08, *z-ratio* = 2.31, $p$ = 0.021) was driven by a non-significant negative trend amongst Disadvantaged participants (*OR* = 0.95, *z-ratio* = -1.81, $p$ = 0.07) and a non-significant positive trend amongst Advantaged participants (*OR* = 1.03, *z-ratio* = 1.42, $p$ = 0.156).

Model diagnostics flagged six potentially influential cases (with Cook's Distances > 4 / $n$ = .0296, where $n$ = sample size). These cases were removed and the model was retested; the resulting model showed the same effects as the original model (with a stronger differentiation of the influence of SVO between Statuses), so the original model was accepted (***Table 3.7***).

**Table 3.7.** *The Status by Social Value Orientation interaction model predicting decisions to find out*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| (Intercept) | 3.21 | 0.60 – 17.13 | 0.169 |
| Age | 0.95 | 0.92 – 0.98 | **0.002** |
| Sex† | 1.18 | 0.57 – 2.48 | 0.656 |
| Status†† | 8.42 | 1.15 – 76.84 | **0.044** |
| Social Value Orientation (SVO) | 1.03 | 0.99 – 1.07 | 0.156 |
| Status x SVO | 0.93 | 0.87 – 0.99 | **0.021** |
| Observations | 135 | | |

*Note.* †Reference group = Female; Treatment group = Male

†† Reference group = Advantaged; Treatment group = Disadvantaged

The results of this analysis suggest opposing effects of pro-sociality depending on social status; Advantaged pro-socials were more likely to find out compared to Disadvantaged pro-socials, and Advantaged pro-selfs were less likely to find out compared to Disadvantaged pro-selfs. ***Figure 3.2*** illustrates the model-derived predicted probability of finding out based on SVO score and Status.

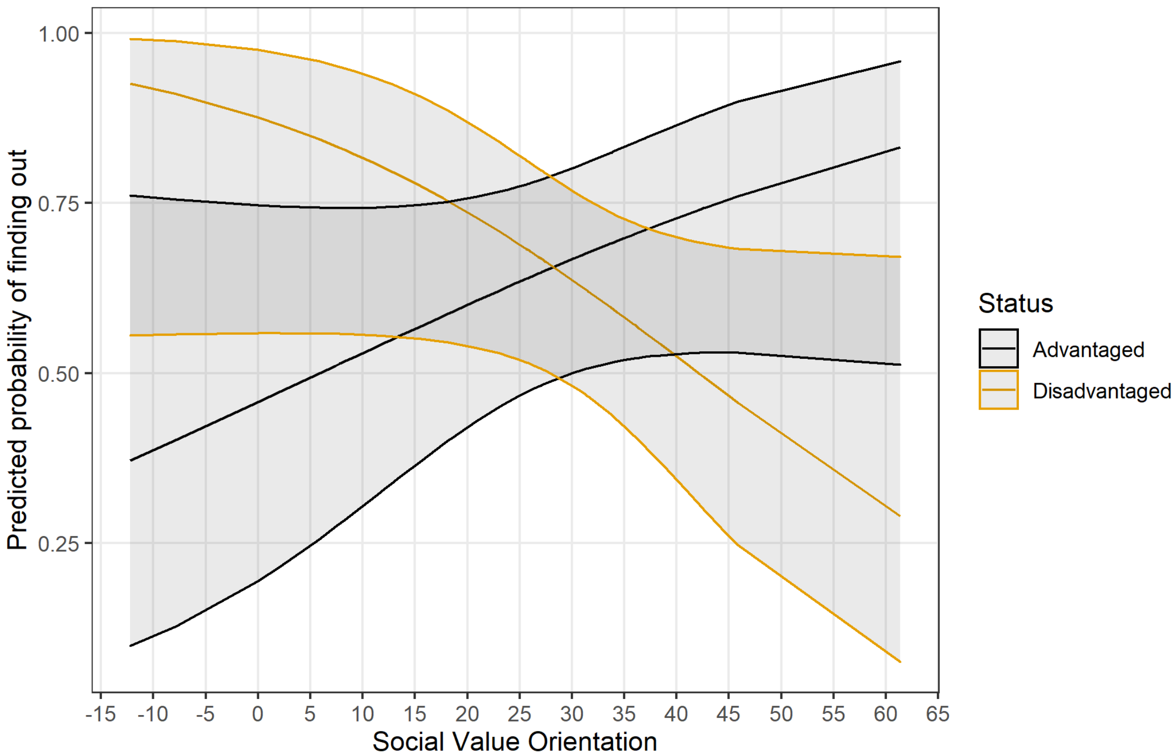


*Figure 3.2*. Predicted probability of a participant deciding to find out based on their Social Value Orientation score and their Status. Shaded areas represent the 95% confidence interval.

Investigation of the secondary measures of SVO permitted deeper exploration into what aspects of pro-sociality were relevant for decisions. The secondary measures of SVO are obtained from participants' SVO responses and describe the extent to which participants' preferences deviate from archetypal inequality aversion (i.e., DIA), archetypal joint gain maximisation (DJG; i.e., maximising the total collective gains), and archetypal altruism (i.e., DAL; Murphy et al., 2011). Since the secondary SVO measures are only relevant for prosocial SVO scorers, these analyses were conducted on the prosocial subset of participants ($N$ = 106). I specified three separate models replacing the SVO term in ***Table 3.7*** with each of DIA, DJG, and DAL. For each model I first specified a main effect and then tested whether adding an interaction with Status improved model fit.

The main effects model for DJG was a significant fit, $\chi^2$ (4, 106) = 9.87, $p$ = 0.043, but only Age was a significant predictor. Adding an interaction term improved the model fit, $\chi^2$ (1,

106) = 7.96, $p$ = 0.005. In the interaction model, the interaction between Status and DJG was significant. Removing four potentially influential cases (Cooks distance > 0.038) and retesting the model showed that the Status x DJG interaction was stronger and the main effect of DJG was significant. Because interpretation of effects was clearer without the influential cases (see post-hocs below), this model, $\chi^2(5,102) = 26.78$, $p < .001$, was accepted (***Table 3.8***).

**Table 3.8.** *The influence of Distance from archetypal Joint Gain (DJG) maximisation on decisions to find out amongst pro-social participants*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 6144.48 | 67.80 – 2373731.99 | **0.001** |
| Age | 0.93 | 0.89 – 0.97 | **0.001** |
| Sex† | 0.84 | 0.34 – 2.08 | 0.706 |
| Status†† | 0.00 | 0.00 – 0.02 | **0.001** |
| DJG⁎ | 0.81 | 0.66 – 0.93 | **0.013** |
| Status x DJG | 1.36 | 1.14 – 1.71 | **0.002** |
| Observations | 102 | | |
| $R^2$ Tjur | 0.227 | | |

*Note.* † Reference group = Female; Treatment group = Male.

†† Reference group = Advantaged; Treatment group = Disadvantaged

⁎ Variable scaled by multiplying by 100

Post-hoc trend analysis of the Status by DJG interaction indicated that the significant difference in the influence of DJG between Advantaged and Disadvantaged participants (*OR* = 0.73, z-ratio = -3.07, *p* = 0.002) was driven by a significant negative trend amongst Advantaged participants (*OR* = 0.81, z-ratio = -2.47, *p* = 0.013) and a non-significant positive trend amongst Disadvantaged participants (*OR* = 1.1, z-ratio = 1.89, *p* = 0.058). ***Figure 3.3*** illustrates this effect.

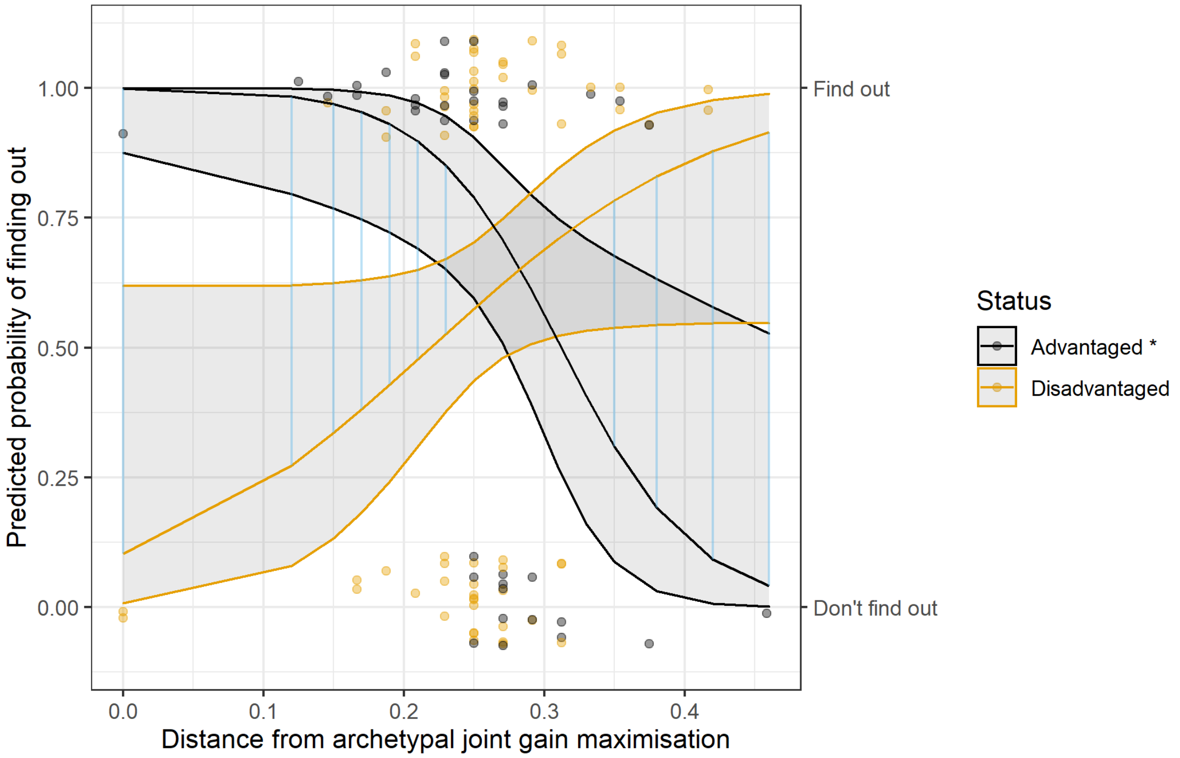


*Figure 3.3.* Predicted probability of a participant finding out about the payments based on their Status and the extent to which they valued maximising joint gains (DJG) in their Social Value Orientation choices. Lower DJG corresponds to greater importance placed on joint gains. The DJG trend was significant amongst Advantaged. Shaded areas are the 95% confidence interval. Blue lines depict significant difference between estimated marginal means. Data points (jittered vertically) correspond to the right vertical axis and represent raw decisions.

This pattern suggests that pro-social Advantaged participants were motivated by the extent to which they valued joint gains. Those who strongly valued joint gains were more likely to find out, whilst those who valued joint gains less were less likely to find out.

The main effect DIA model was a significant fit to the data, $\chi^2$ (4, 106) = 16.82, *p* = 0.002. Adding the DIA x Status interaction term did not improve model fit, $\chi^2$ (1, 106) = 0.61,

$p$ = 0.433. Removing five potentially influential cases (Cook's distance > 0.038) and retesting the model showed the same effects as the original model. The original model was accepted (***Table 3.9***).

**Table 3.9.** *The influence of Distance from archetypal Inequality Aversion (DIA) on decisions to find out amongst pro-social participants*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 6.56 | 1.47 – 33.03 | **0.017** |
| Age | 0.95 | 0.92 – 0.99 | **0.009** |
| Sex† | 0.85 | 0.36 – 1.97 | 0.705 |
| Status†† | 0.57 | 0.23 – 1.34 | 0.204 |
| DIA⁕ | 1.07 | 1.02 – 1.14 | **0.017** |
| Observations | 106 | | |
| $R^2$ Tjur | 0.147 | | |

*Note.* † Reference group = Female; Treatment group = Male

†† Reference group = Advantaged; Treatment group = Disadvantaged

⁕ Variable scaled by multiplying by 100

***Figure 3.4*** shows that, for individuals of both Statuses, those who were more averse to inequality in their SVO choices were less likely to find out about the payments. This may indicate that participants were motivated to ignore payment information to avoid finding out an outcome that conflicted with their preference.

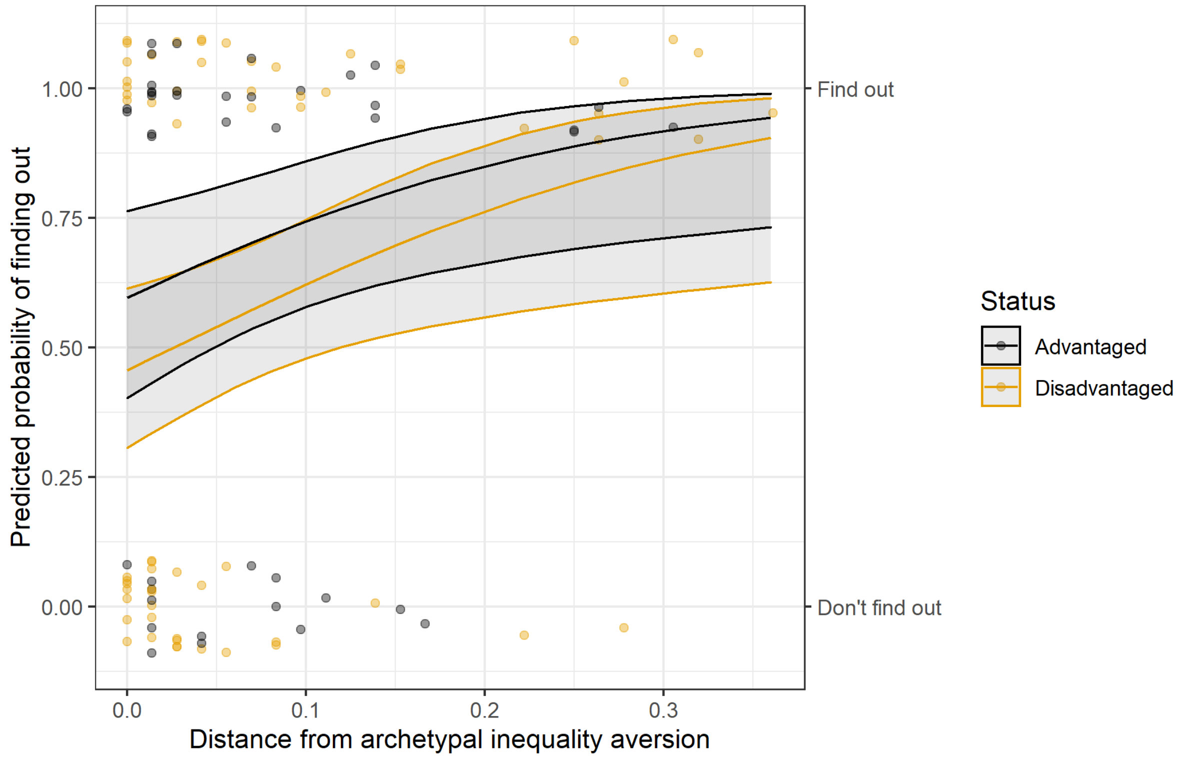


*Figure 3.4.* Predicted probability of a participant finding out about the payments based on their Status and the extent to which they were averse to inequality (DIA) in their Social Value Orientation choices. Lower DIA corresponds to greater aversion to inequality. There was no difference between Statuses in the influence of DIA on finding out. Shaded areas are the 95% confidence interval. Data points (jittered vertically) correspond to the right vertical axis and represent raw decisions.

**Status by System Justification General.** The System Justification General (SJG) model explained 8.5% of the variance in decisions to find out, $\chi^2(5, 135) = 15.72$, $p = 0.008$ (***Table 3.10***). Trend analysis of the Status by SJG interaction indicated that the significant difference in the influence of SJG between Advantaged and Disadvantaged participants (*OR* = 0.93, *z-ratio* = -1.96, *p* = 0.05) was driven mainly by a significant positive trend amongst Disadvantaged participants (*OR* = 1.06, *z-ratio* = 2.17, *p* = 0.03), with Advantaged participants showing a negligible non-significant negative trend (*OR* = 0.99, *z-ratio* = -0.41, *p* = 0.68).

Diagnostics for this model flagged five potentially influential cases (with Cook's Distances > 0.0296). These cases were removed and the model retested. The same effects as the original model were evident, with the addition of a marginally significant Sex term. The original model was accepted.

**Table 3.10.** *The Status by System Justification General interaction model predicting decisions to find out*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| (Intercept) | 10.19 | 1.52 – 79.03 | **0.020** |
| Age | 0.94 | 0.91 – 0.98 | **0.001** |
| Sex† | 1.42 | 0.68 – 3.03 | 0.355 |
| Status†† | 0.10 | 0.01 – 0.98 | 0.054 |
| System Justification General (SJG) | 0.99 | 0.95 – 1.03 | 0.680 |
| Status x SJG | 1.07 | 1.00 – 1.16 | **0.050** |
| Observations | 135 | | |

*Note.* †Reference group = Female; Treatment group = Male.

†† Reference group = Advantaged; Treatment group = Disadvantaged

This analysis showed that the extent to which participants tended to support the status quo influenced the likelihood of finding out about payments if they were Disadvantaged but not if they were Advantaged. ***Figure 3.5*** illustrates the model-derived predicted probability of finding out based on SJG score and Status.

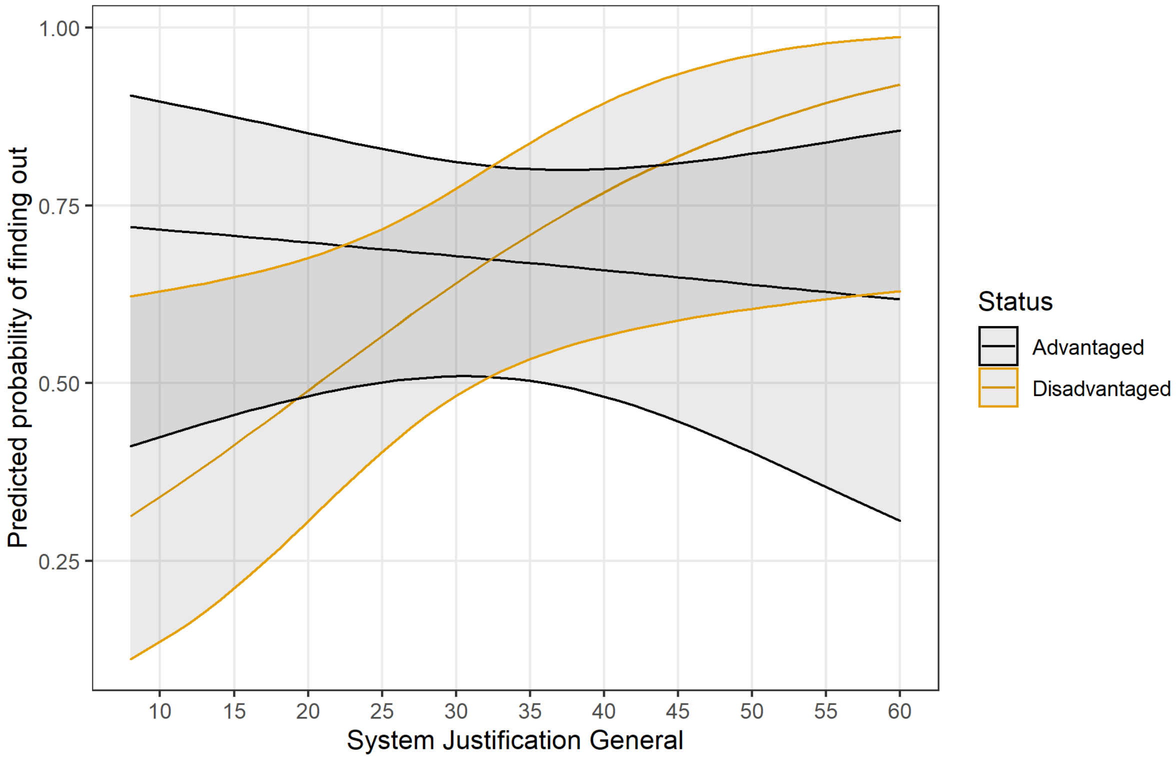


*Figure 3.5*. Predicted probability of a participant deciding to find out based on their System Justification General score and their Status. Shaded areas represent the 95% confidence interval.

**Status by System Justification Economic.** The System Justification Economic (SJE) model explained 7.86% of the variance in decisions to find out, $\chi^2(5, 135) = 14.53$, $p = 0.013$. Trend analysis of the Status by SJE interaction indicated that the significant difference in the influence of SJE between Advantaged and Disadvantaged participants (*OR* = 0.95, *z-ratio* = -1.99, $p = 0.047$) was driven by a non-significant negative trend amongst Advantaged participants (*OR* = 0.97, *z-ratio* = -1.65, $p = 0.099$) and a non-significant positive trend amongst Disadvantaged participants (*OR* = 1.02, *z-ratio* = 1.15, $p = 0.251$).

Model diagnostics flagged four potentially influential cases (Cook's Distances > 0.0296). These cases were removed and the model was retested. The retested model showed the same effects as the original model with the addition that the main effect of SJE became significant. The original model was kept as the accepted model (***Table 3.11***).

**Table 3.11.** *The Status by System Justification Economic interaction model predicting decisions to find out*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| (Intercept) | 91.82 | 3.28 – 4104.57 | **0.012** |
| Age | 0.95 | 0.91 – 0.98 | **0.001** |
| Sex[†] | 1.41 | 0.67 – 3.02 | 0.367 |
| Status[††] | 0.02 | 0.00 – 0.79 | **0.044** |
| System Justification Economic (SJE) | 0.97 | 0.93 – 1.00 | 0.099 |
| Status x SJE | 1.05 | 1.00 – 1.11 | **0.047** |
| Observations | 135 | | |

*Note.* [†]Reference group = Female; Treatment group = Male

[††] Reference group = Advantaged; Treatment group = Disadvantaged

This analysis indicated that the extent to which participants endorsed meritocratic ideologies regarding the distribution of wealth led to a lower likelihood of finding out about amongst Advantaged participants compared to Disadvantaged participants. ***Figure 3.6*** illustrates the model-derived predicted probability of finding out based on SJE score and Status.

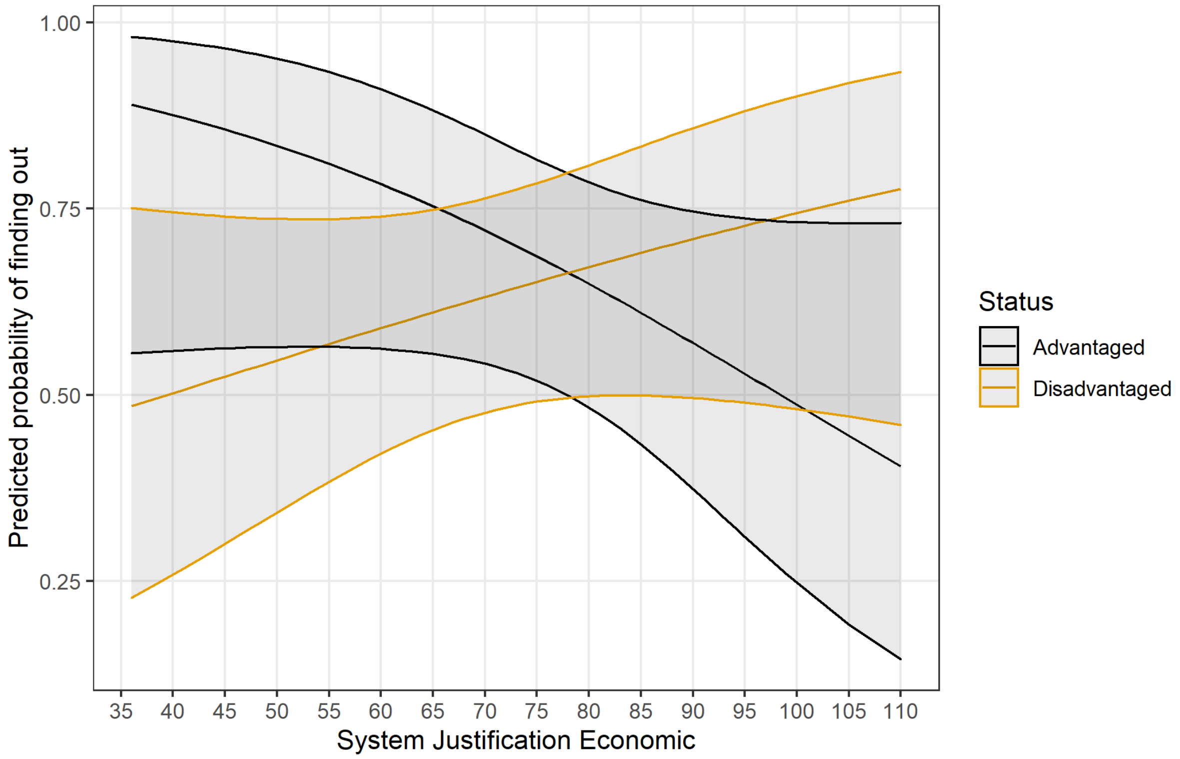


*Figure 3.6.* Predicted probability of a participant deciding to find out based on their System Justification Economic score and their Status. Shaded areas represent the 95% confidence interval.

### *Situational Factors*

I next investigated how situational factors (i.e., participants' subjective feelings in the scenario) influenced Discovery decisions. These situational factors were: (i) Payment satisfaction, (ii) Comfort not knowing about the payments, (iii) Comfort with Unfairness, and (iv) Partner identification. These terms were passed to the *buildmer* function along with Age, Sex, and Status and their interaction with each of above-listed factors. ***Table 3.12*** lists these terms.

**Table 3.12.** *Terms entered into buildmer function*

| |
|---|
| Age |
| Sex |
| Status |
| Payment satisfaction |
| Comfort not knowing about payments |
| Comfort with unfairness |
| Partner identification |
| Identification with Disadvantaged |
| Identification with Advantaged |
| Age x Sex |
| Age x Status |
| Sex x Status |
| Age x Payment satisfaction |
| Age x Comfort not knowing about payments |
| Age x Comfort with unfairness |
| Age x Partner identification |
| Age x Identification with Disadvantaged |
| Age x Identification with Advantaged |
| Sex x Payment satisfaction |
| Sex x Comfort not knowing about payments |
| Sex x Comfort with unfairness |
| Sex x Partner identification |
| Sex x Identification with Disadvantaged |
| Sex x Identification with Advantaged |
| Status x Payment satisfaction |
| Status x Comfort not knowing about payments |
| Status x Comfort with unfairness |
| Status x Partner identification |
| Status x Identification with Disadvantaged |
| Status x Identification with Advantaged |

The best-fitting model identified by *buildmer* contained just Comfort not knowing and Age. As in previous analyses, I used this procedure as a starting to point but did not accept the resultant model as final. Instead, I went on to test each situational factor in independent models alongside Age, Sex, and Status. Doing so I confirmed the effect of Comfort not knowing (controlling for Age, Sex, and Status), and that Comfort with unfairness and Payment satisfaction were non-significant predictors. The final Comfort not knowing model (***Table 3.13***) explained approximately 12% of the variance in Discover, $\chi^2(4, 135) = 23.38$, $p < .001$.

**Table 3.13.** *Best-fitting model predicting finding out from situational factors*

| *Predictors* | *Odds Ratios* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 23.94 | 5.41 – 126.38 | **< 0.001** |
| Age | 0.96 | 0.93 – 0.99 | **0.011** |
| Sex† | 1.95 | 0.89 – 4.45 | 0.101 |
| Status†† | 0.96 | 0.45 – 2.04 | 0.907 |
| Comfort not knowing | 0.98 | 0.96 – 0.99 | **0.001** |
| Observations | 135 | | |
| $R^2$ Tjur | 0.157 | | |

*Note.* † Reference group = Female; Treatment group = Male

†† Reference group = Advantaged; Treatment group = Disadvantaged

**Section 2: Predictors of Comfort not knowing**

Having identified that Comfort not knowing had a negative influence on the likelihood of deciding to find out about pay inequality, I next aimed to explore how individual difference measures predicted Comfort not knowing. To do this, I tested models predicting Comfort not knowing from each individual difference measure in its own model alongside Age, Sex, and Status, including an interaction term between the individual difference of interest and each of Age, Sex, and Status.

The analysis procedure showed that the System Justification General (SJG) and System Justification Economic (SJE), in their separate models, each significantly influenced Comfort not knowing about the payments. Diagnostic tests indicated that the influence of SJG on Comfort not knowing was quadratic in nature. Though a quadratic relationship was not hypothesised, there is some previous research that has likewise observed quadratic effects of system justification (Cichocka et al., 2018). I therefore tested and confirmed the presence of a quadratic influence of SJG. The final model explained 15.95% of the variance in Comfort not knowing, $F(5, 129) = 6.09$, $p < .001$ and contained significant linear ($b = 69.93$, $z = 2.42$, $p = 0.017$) and quadratic ($b = -67.85$, $z = -2.27$, $p = 0.025$) effects of SJG, in addition to significant effects of Age and Sex (***Table 3.14***).

**Table 3.14.** *Model predicting Comfort not knowing about the payments from System Justification General scores*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
|---|---|---|---|
| Intercept | 44.31 | 27.95 – 60.66 | < **0.001** |
| Age | 0.49 | 0.07 – 0.91 | **0.023** |
| Sex[†] | 19.94 | 9.97 – 29.91 | < **0.001** |
| Status[††] | 2.20 | -8.03 – 12.43 | 0.672 |
| System Justification General - Linear[ᴦ] | 69.93 | 12.65 – 127.21 | **0.017** |
| System Justification General - Quadratic[ᴦ] | -67.85 | -127.08 – -8.63 | **0.025** |
| Observations: 135 | | | |

*Note.* † Reference group = Female; Treatment group = Male

†† Reference group = Advantaged; Treatment group = Disadvantaged

ᴦ Orthogonalised estimates

These results suggest that whilst overall stronger justifiers typically felt more comfortable not knowing than weaker justifiers, there was a limit point of justification beyond which strong justifiers felt less comfortable than moderate justifiers (***Figure 3.7***).

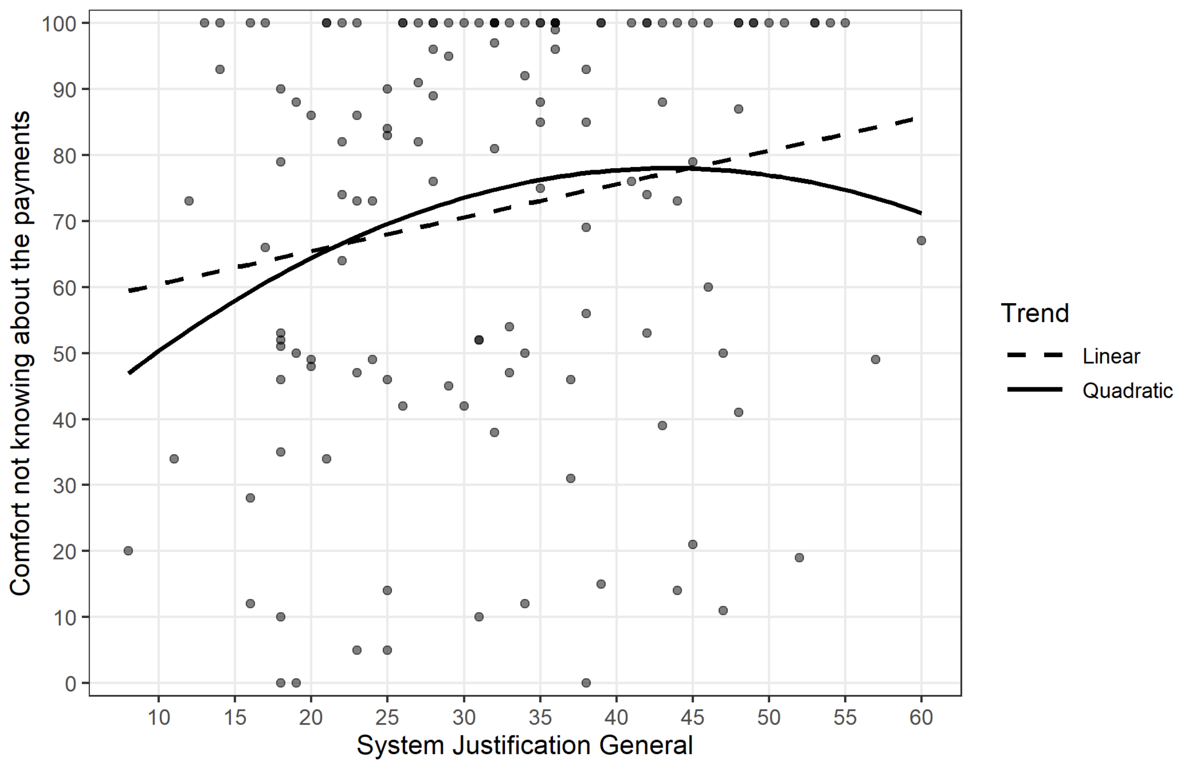


*Figure 3.7.* Linear and quadratic trends of System Justification General predicting Comfort not knowing whether payments were equal

To further illustrate the above model, ***Figure 3.8*** delineates the influence of System Justification General on Comfort not knowing for male and female and younger and older participants (note however that there were no significant interaction effects in the model). The figure suggests that the quadratic trend was particularly evident for younger participants.

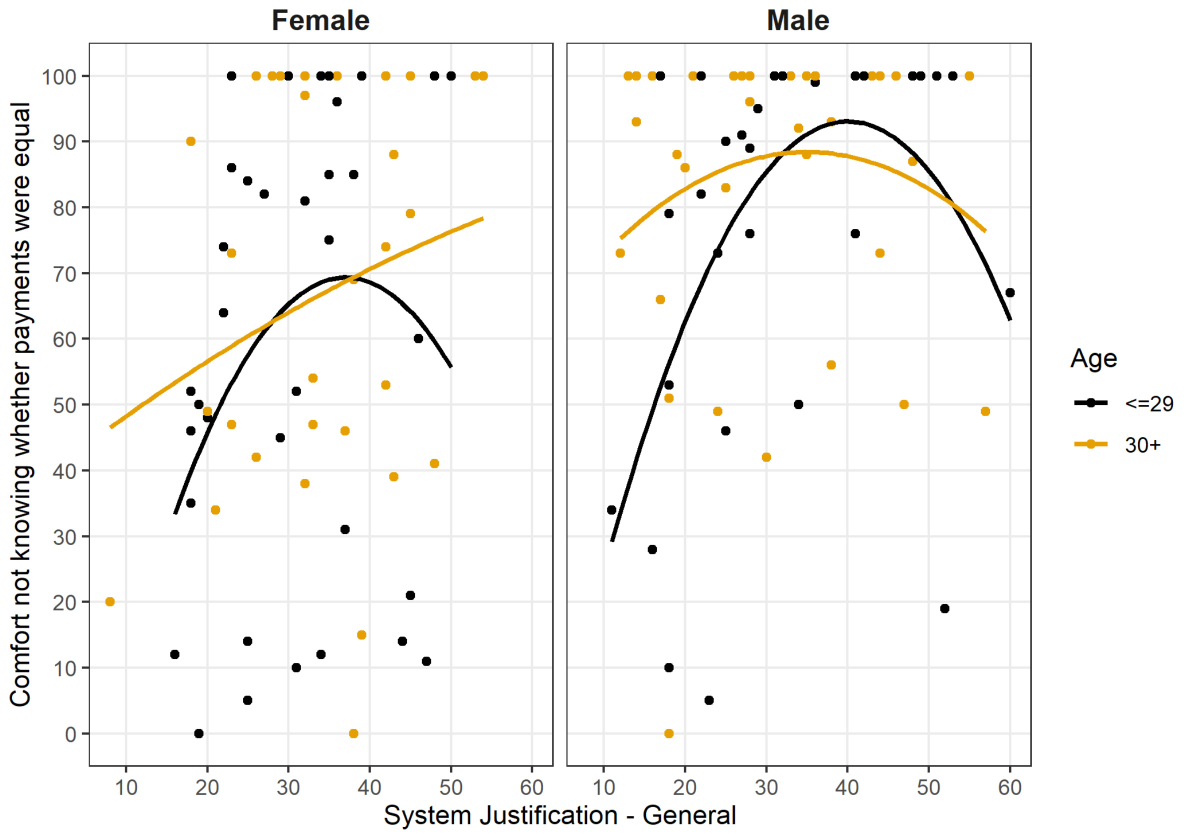


*Figure 3.8.* Comfort not knowing whether payments were equal based on System Justification General score, Sex (Female N = 67; Male N = 68), and Age. Participants were binned into a younger age group up to and including the median age (N = 68) and an older age group above the median age (N = 67).

The SJE model explained 12.36% of the variance in Comfort not knowing, $F(4,130) = 5.72$, $p < .001$ (***Table 3.15***). This model showed that older compared to younger ($b = 0.51$, $z = 2.33$, $p = 0.021$), male compared to female ($b = 18.78$, $z = 3.67$, $p < .001$), and high justifying compared to low justifying ($b = 0.35$, $z = 2.13$, $p = 0.035$) participants were more comfortable not knowing. In contrast to SJG, there was no evidence of a quadratic relationship between SJE and Comfort not knowing.

**Table 3.15.** *Model predicting Comfort not knowing about the payments from System Justification Economic scores*

| *Predictors* | *Estimates* | *CI (95%)* | *p* |
| --- | --- | --- | --- |
| Intercept | 17.15 | -13.91 – 48.21 | 0.277 |
| Age | 0.51 | 0.08 – 0.94 | **0.021** |
| Sex† | 18.78 | 8.66 – 28.90 | < **0.001** |
| Status†† | 4.90 | -5.33 – 15.14 | 0.345 |
| System Justification Economic | 0.35 | 0.03 – 0.67 | **0.035** |
| Observations: 135 | | | |

*Note.* † Reference group = Female; Treatment group = Male

†† Reference group = Advantaged; Treatment group = Disadvantaged

**Section 3: Motivations for the Decision**

I next turned the focus of analysis onto what reasons participants endorsed as motivators of their decisions. The goal of this procedure was to explore, in a descriptive (i.e., non-inferential) way, the reasons participants reported for their decisions. I did this by computing mean endorsement of each reason for each of four cells: Advantaged who found out, Disadvantaged who found out, Advantaged who did not find out, and Disadvantaged who did not find out. In this way, the motivations behind decisions to find out and not find out could be compared between Advantaged and Disadvantaged participants. Note that for decisions not to find out fewer reasons were relevant and so only those relevant for the decision were explored. The results of these comparisons can be seen in ***Figure 3.9***.

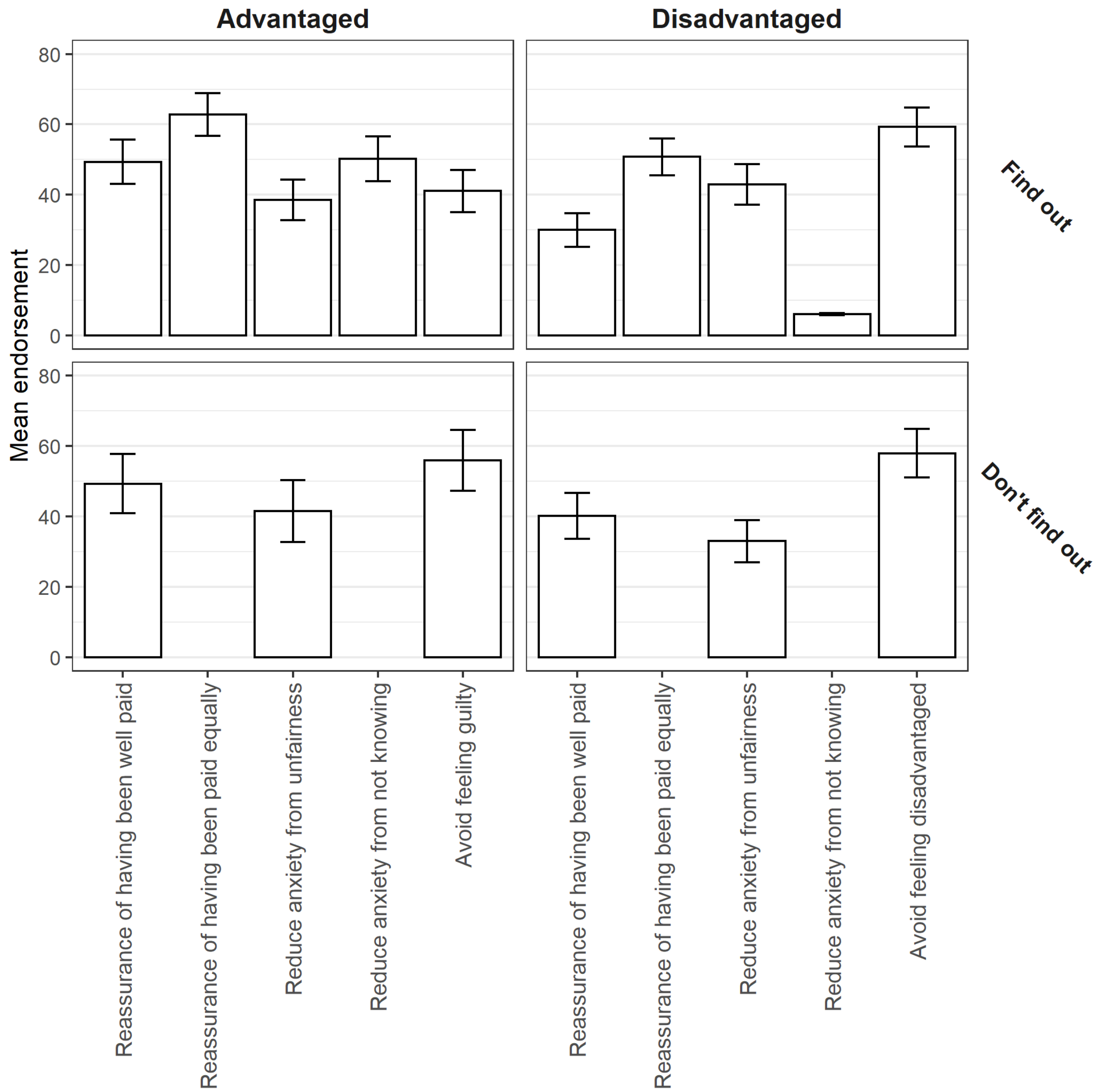


*Figure 3.9.* Mean endorsement of reasons motivating Decisions to find out and not find out amongst Advantaged and Disadvantaged participants. Error bars represent standard error of the mean.

***Figure 3.9*** shows that for Advantaged participants who chose to find out, "Reassurance of having been paid equally" was the highest-endorsed motivation, followed by "Reduce anxiety from not knowing." Amongst Disadvantaged participants who chose to find out, the highest-endorsed reason was "Avoid disadvantage", followed by "Reassurance of having been paid equally." Choices not to find out amongst Advantaged participants were mainly motivated by a desire to avoid feelings of guilt, whilst for Disadvantaged participants a choice not to find out was motivated mainly be a desire to avoid feelings of disadvantage.

***Qualitative reports of motivations for decision***

Participants were also invited to provide any other comments detailing reasons for their decisions. Eighty-eight participants provided such comments. These comments were insightful and provided more detail into participants' decision-making process than the rating questions alone. I categorised each comment into the principal theme it expressed. The identified themes are listed in ***Table 3.16*** (see ***Appendix B*****: Participant comments from which qualitative themes were derived in Study 3** for the raw comments). I then calculated the number of times each theme was mentioned as a proportion of the total references to themes for each cell of the Status-by-Decision contingency table. The results of this process are represented in ***Figure 3.10***.

**Table 3.16.** *Themes identified from qualitative participant reports (N = 88) of the motivations behind their decisions*

| Theme | Description |
|---|---|
| Avoid affective reaction | Motivations to avoid negative feelings associated with inequality (e.g., feelings of guilt or disadvantage). Mostly associated with not finding out. |
| Curiosity | Information-seeking motivations exclusively associated with finding out. |
| Equality concerns | Discomfort with inequality, or suspicions about the presence of inequality. |
| Expectations | Appreciation of the experimental context (e.g., awareness of the fact that all participants made informed decisions to participate). |
| Fairness concerns | Importance placed on participants having been paid a fair amount for their participation. |
| Indifference | Disinterest in the situation either way; the choice is random. |
| Intolerance of ignorance | A feeling that choosing not to find out is unacceptable. |
| Payment satisfaction | Positive expressions of satisfaction with the amount received. Frequently expressed in combination with motivations to avoid ruining this satisfaction with negative emotions associated with finding out. |
| Personal irrelevance | Expressions that since partners are anonymous their outcomes have no relevance or importance to oneself. Exclusively associated with not finding out. |
| Transparency | A desire that information about the payments should be available as a matter of principle. |
| Unchangeability | Feelings arising from the fact that nothing can change the situation. Almost exclusively reported as a justification not to find out, though also expressed as a reason to be indifferent. |
| Other | Reports that did not clearly fit a specific theme. |

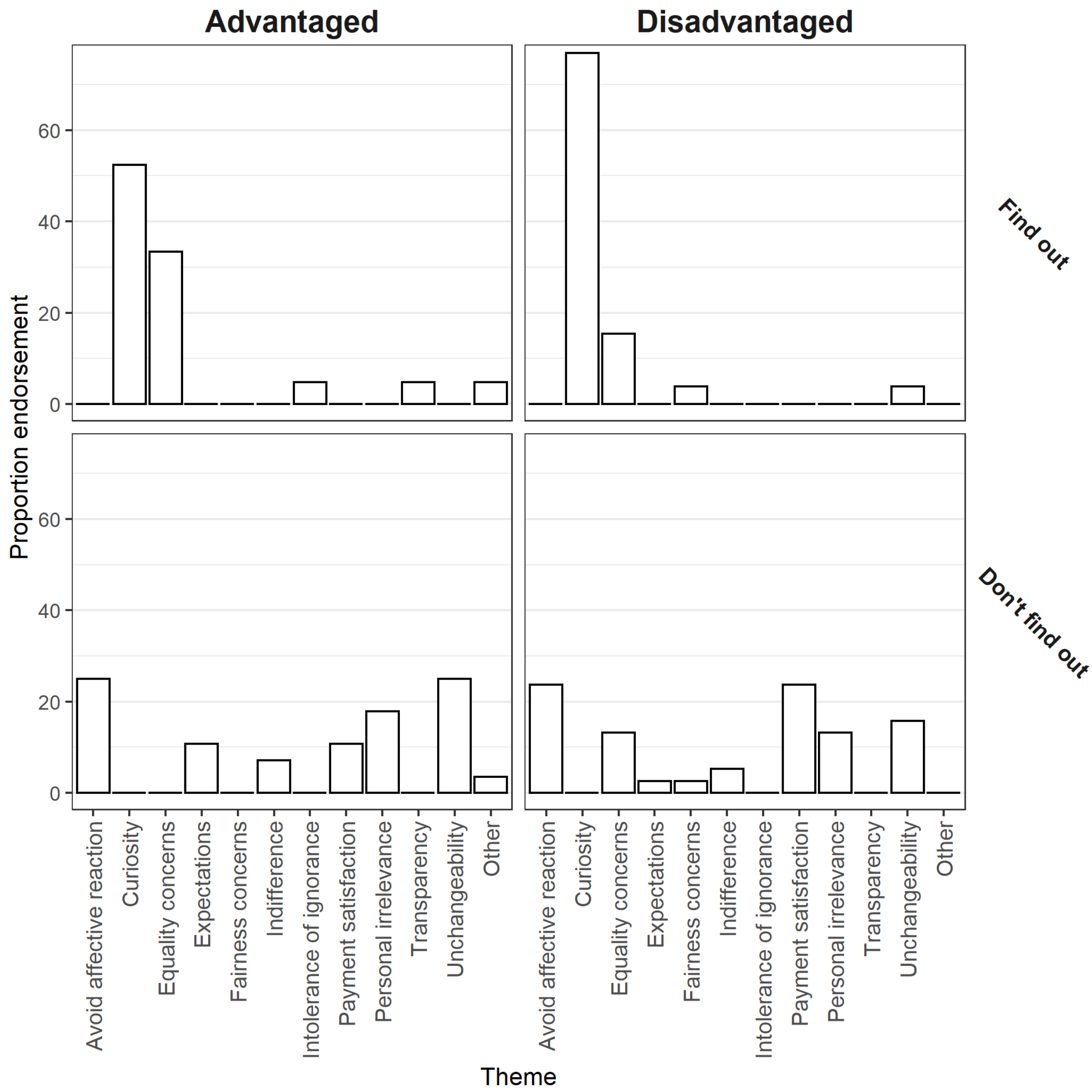


*Figure 3.10.* Proportion endorsement of the themes mentioned in participants' comments regarding the motivations behind their decisions.

***Figure 3.10*** shows that Curiosity was a dominant motivation to find out for both Advantaged and Disadvantaged participants. Indeed, curiosity was the strongest-endorsed theme in the entire subset, reported by 31 of the 88 participants that provided comments. Fifty-two percent of Advantaged – and 77% of Disadvantaged – participants who found out cited curiosity as an important motivator for their decision. Amongst those who chose to find out, Equality Concerns also stood out as important, particularly for Advantaged participants; seven out of 19 (33%) Advantaged participants who found out mentioned Equality Concerns as a motivator, typically expressing the sentiment that payments for the same work should be equal.

Amongst the 4 out of 24 (15%) of Disadvantaged participants that mentioned this theme, only one expressed the sentiment of equal pay for equal work; the other comments instead appeared to reflect a motivation to confirm suspicions about inequality (e.g., “to see if the study will follow the economic structure of society – some being more rewarded than others”).

Avoiding Affective Reactions and Unchangeability were the two highest-endorsed themes amongst Advantaged participants that chose not to find out, each endorsed by 25% (7 out of 19) of this subset. These participants expressed a desire to avoid experiencing negative emotions such as anxiety and guilt that they expected would come with deciding to find out. The unchangeable nature of the situation was expressed as a sound justification not to find out, and in several cases this reasoning was combined with the avoidance of negative affect reasoning, with participants expressing that they felt justified to spare themselves the negative affect associated with finding out since they could not change it in any case.

Disadvantaged participants who did not find out also strongly endorsed avoidance of affective reactions as a reason for their decision (9 out of 25 participants – 24%). These comments tended to reflect an “ignorance is bliss” attitude whereby positive emotions could be maintained by avoiding potentially negative information that could provoke feelings of injustice or jealousy. Payment satisfaction was endorsed equally as strongly amongst these participants; contentment with payment was mentioned by 24% of the subset and indeed several participants explicitly expressed that they wanted to maintain their positive affect from satisfaction with pay by avoiding information that might induce negative feelings. The unchangeability of the situation was also cited by several participants, who expressed the sentiment that since nothing can change there is no use in finding out.

## Discussion

I investigated the individual differences and situational factors that influence whether individuals seek or avoid information about equality in an artificial social system. The results provide considerable insight into the extent to which individuals seek (in)equality information, and the personality characteristics that are important for motivating their decisions to seek or avoid information. Age was consistently found to be influential; older individuals were less likely to find out about payments than younger individuals. An individual's potential advantage or disadvantage was found to be a consistent moderator of the influence of individual difference characteristics. System justification featured prominently. Amongst potentially disadvantaged participants, strong general justifiers (System Justification General; SJG) were more likely to find out compared to weak justifiers, whereas amongst potentially advantaged participants SJG had no influence. Similarly, strong potentially disadvantaged economic justifiers (System Justification Economic; SJE) were more likely to find out compared to strong potentially advantaged justifiers. High pro-sociality (as measured by Social Value Orientation; SVO) amongst potentially disadvantaged participants was associated with a lesser likelihood of finding out compared to high pro-sociality amongst potentially advantaged participants. The finding that the relative payment position was an important moderator is highly significant because it implies that the motivations underlying decisions to seek or avoid equality information vary depending on whether the individual is potentially advantaged or disadvantaged by inequality.

Further analyses showed that individuals who were more comfortable not knowing about the payments were less likely to find out. Age, Sex, and both measurements of system justification predicted how comfortable individuals were not knowing. Older, compared to younger, female, compared to male, and high, compared to low, SJE participants were more comfortable not knowing. The relationship between SJG and individuals' comfort with not knowing about the payments was more complex; I found a quadratic relationship in addition to a linear one. The linear effect showed that stronger justifiers were more comfortable not knowing than weaker justifiers. However, the quadratic effect showed that whilst moderate justifiers were more comfortable not knowing than weak justifiers, strong justifiers were less comfortable not knowing than moderate justifiers (***Figure 3.7***).

**The prominent role of System Justification**

The system justification findings were perhaps the most prominent of the study. The influence of system justification on whether individuals sought information about their payments was contingent on their Status, and intriguingly the direction of this difference was different for the General (SJG) and Economic (SJE) scales of system justification.

The SJG findings showed that amongst potentially disadvantaged participants, increases in SJG led to an increased likelihood of finding out, whereas amongst potentially advantaged participants, increases in SJG had no influence. That is, strong potentially disadvantaged justifiers were more likely to find out compared to their non-justifying peers, whereas there was no difference between strong and weak potentially advantaged justifiers (***Figure 3.5***). These findings suggest that the extent to which individuals justify existing social arrangements is particularly important for information-seeking if those individuals believe they are potentially disadvantaged.

According to System Justification Theory (SJT), justifying the status quo serves a palliative function by reducing the negative affect that can arise from the current state of affairs (e.g., the presence of systemic inequality) for *both* advantaged and disadvantaged individuals (Jost & Hunyady, 2003; Jost et al., 2008). Since humans typically demonstrate preferences for fairness (e.g., Güth et al., 1982; Lerner, 1980; Oosterbeek et al., 2004), SJT holds that the experience of systemic unfairness is aversive because it creates dissonance and other negative feelings for both those who do not benefit from unfairness and those who do, for example feelings of anger amongst disadvantaged individuals and feelings of guilt amongst advantaged individuals (e.g., Hegtvedt, 1990). In the present experiment, however, dispositional SJG was only relevant for potentially disadvantaged participants' decisions. This suggests that whereas potentially disadvantaged participants' focus drew on their attitudes towards the justifiability of the status quo, potentially advantaged participants' focus did not.

A possible explanation for this divergence may be drawn from the suggestion that, in some cases, those disadvantaged by social inequalities have an enhanced need to justify the system (Jost & Burgess, 2000; Jost et al., 2003). In such scenarios, greater justification on the part of disadvantaged individuals is suggested to arise from the fact that they experience cognitive dissonance as a result of a need to believe in the inherent fairness of society on the one hand and the self-evident disadvantage they are subject to on the other. In contrast, the advantaged experience little dissonance because their experience in the system is more consistent with their self-esteem motives. Because of this enhanced dissonance experienced specifically amongst the disadvantaged, it is suggested that under some circumstances they may have a greater need to legitimise the system in order to justify the malalignment of self- and system-justifying motives (Jost et al., 2003). Though this phenomenon is not a central aspect of SJT (contrary to what some have asserted, e.g., Brandt, 2013) it has nonetheless been acknowledged that it is *sometimes* the case that disadvantaged participants are particularly motivated to justify the social systems that disadvantage them (Jost, 2017, 2019).

I suggest that the context of the present experiment may be one such occasion where the disadvantaged are more sensitive to processes of system justification that the advantaged. Here, it may be that there was no conflict between potentially advantaged participants' self-esteem and the possible states of the system; regardless of the state of the system, they understood that they would be paid well. In contrast, the conflict did exist for potentially disadvantaged participants; since there was a possibility that they could be paid less, concerns for their self-esteem (i.e., being paid well/fairly) were at odds with seeing the system as legitimate. This may be why potentially disadvantaged, but not potentially advantaged, participants were influenced by their SJG in their decision-making. This suggestion gains credence when one considers the influence of SJG on participants' beliefs about the probability of fairness. Both potentially advantaged and potentially disadvantaged strong justifiers believed in a higher probability that the system was in fact fair compared to weaker justifiers. The finding that this translated into a greater likelihood of finding out only for the potentially disadvantaged is telling. The picture emerges that strong justifiers have greater confidence in the fairness of the system (as indicated by their belief in the probability of fairness), but that potentially advantaged and disadvantaged individuals use that confidence in different ways.

Because of the asymmetry, only potentially disadvantaged strong justifiers were in a position of choosing between the protracted uncertainty of motive dissonance on the one hand and certain information (which they had a strong belief would disaffirm dissonance) on the other. Put succinctly, since strong justifiers had greater confidence that the system was fair, potentially disadvantaged justifiers judged resolving uncertainty to be relatively risk-free and preferable to remaining ignorant. For these individuals, confirming fairness was more valuable than enduring the possibility of unfairness, even though there was some risk (believed to be low) that they would discover unfairness with certainty.

In contrast, potentially advantaged justifiers experienced no (or at least less) conflict between self- and system-justifying motives, and as a result their decision was of lesser import to their beliefs about their own deservingness and the legitimacy of the system. Whether the system was fair or unfair, because these individuals were either paid well or very well, there was little inconsistency between their treatment by the system and a disposition to support it. This explains why potentially advantaged justifiers had no more motivation to find out than potentially advantaged non-justifiers (indeed, one could speculate that they should even be motivated to avoid information to maintain their confidence in the system, since a situation in which they are paid well and believe the system to be fair would be maximally consistent; see below discussion of SJE).

The notion that disadvantaged individuals are typically more likely than advantaged individuals to legitimise the system has been dubbed the "status-legitimacy hypothesis" (Brandt, 2013). It is a controversial argument that has received inconsistent support (Brandt, 2013; Caricati & Lorenzi-Cioldi, 2012; Henry & Saul, 2006; Jost et al., 2003; Sengupta et al., 2015), leading some to question its relevance. However, it has been emphasised that the status-legitimacy argument is not a central tenet of SJT and in fact is not even a logical prediction of the theory (Jost, 2017, 2019). What has been argued by SJT theorists is that in certain circumstances the disadvantaged may be more sensitive to and benefit from the processes of legitimising existing social arrangements (Jost, 2017). For example, Sengupta et al. (2015) showed that low-status ethnic groups perceived ethnic-group relations – but not the overall political system – to be fair to a greater extent than did high-status ethnic groups. Thus, under certain circumstances disadvantaged individuals may exhibit greater justification of specific aspects of a social system. However, Jost (2017) emphasises that focusing on which group justifies more or less is missing the point. Instead, the focus should be on the fact that there is justification to any extent of a system by those who are disadvantaged by it. My findings add to the abundant literature demonstrating that disadvantaged groups are consistently found to legitimise existing social arrangements to some extent (Durrheim et al., 2014; Godfrey & Wolf, 2016; Richter & König, 2017; Wiederkehr et al., 2015), and provide the novel insight that disadvantaged justifiers may be particularly motivated to seek justice information.

The observed influence of SJE on decisions lends credibility to the argument that potentially advantaged and disadvantaged individuals may be sensitive to different aspects of social arrangements. Here, I observed that potentially advantaged participants who were stronger endorsers of the meritocratic ideology of Western society (i.e., high SJE) were less likely to find out about the payments than their potentially disadvantaged colleagues. Whereas potentially advantaged participants were not influenced by their general endorsement of society (SJG), their endorsement of the economic and meritocratic aspects of society (SJE) appeared to be more relevant for their decisions (compare ***Figure 3.5*** and ***Figure 3.6***; though recall that both SJE trends were in themselves non-significant). A clearer divergence between the motives of potentially advantaged and disadvantaged strong economic justifiers was thus evident. As alluded to above, the attraction of potentially advantaged justifiers towards not finding out may be understood to derive from a motivation to maintain consistency between their payment and their beliefs about meritocracy. For strong endorsers of meritocracy, it is difficult to ameliorate the view that people's economic outcomes are directly related to their effort with the fact that they have just been paid more than others for objectively the same work. To minimise the discord of these incompatible states, potentially advantaged justifiers were likely to avoid finding out and their strong belief about the fairness of the system allowed them to maintain, in ignorance, consistency between their payments and their ideology.

Again, the difference between the potentially advantaged and disadvantaged can be understood in terms of what each group stood to gain or lose by seeking information. I suggest that potentially advantaged participants by default adopted a loss frame (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), whereby a decision to find out could only reduce the subjective value of their position in terms of consistency between beliefs and outcomes. Instead, potentially disadvantaged justifiers were in a position of greater uncertainty regarding the consistency between their payment and their system justification outlook. These participants, believing in the fairness of the system, stood to gain from finding out insofar as – driven by their confidence in the system – a decision to find out was believed to be likely to release them from the uncertainty of payment-belief inconsistency by affirming the fairness of the system. The finding that, amongst the potentially disadvantaged, strong (compared to weak) justifiers reported being less motivated by a need for reassurance of having been well paid further supports the notion that strong justification ideologically bolstered these participants against the negative affect that might come from finding out that they were paid unequally, and so had few reservations about deciding to find out.

### *System Justification and Comfort not knowing*

"Comfort not knowing" describes the extent to which individuals were comfortable remaining in a state of relative information paucity. As might have been expected, I found that an individual's comfort with not knowing about the payments decreased the likelihood that they would choose to find out. This comfort was itself predicted by dispositional system justification. Specifically, SJG exerted both a linear influence – indicating that higher SJG predicted greater comfort with not knowing about payments – and a quadratic influence – indicating that the positive influence of SJG on this comfort became less for higher values of SJG, to the point that very high values of SJG were associated with less comfort than moderate values of SJG. In simple terms, system justifiers were more comfortable not knowing (up to a point) and so were less likely to find out, compared to weak justifiers. Extremely strong justifiers were less comfortable not knowing than moderate justifiers, and so were more likely to find out. I suggest that these findings can be understood as addressing individuals' epistemic needs; that is, their needs for dealing with uncertainty. Moreover, below I argue that the presence of both linear and quadratic effects of system justification on comfort not knowing about payments may indicate that, whereas moderate justifiers were comfortable not knowing because their trait confidence in existing social structures mitigated the negative effects of uncertainty, extremely strong justifiers may in fact have stronger epistemic needs which made them less comfortable not knowing.

The finding that higher SJG was associated with higher levels of comfort provides evidence that justifiers were better able to derive epistemic comfort in an uncertain situation compared to weak justifiers. This is consistent with previous research which shows that Need for Cognition – a measure of the extent to which individuals are drawn to effortful cognition (Cacioppo et al., 1984) – is found to negatively relate to both General and Economic System Justification (Hennes et al., 2012; Jost et al., 2017); that is, stronger justifiers are less prone to seek out information and deliberate over it. However, it should be noted that the claim that increases in system justification are related to greater comfort not knowing is in some ways counter theoretical. Strong justifiers may be understood to have stronger epistemic needs. As Jost and colleagues (2017; 2019) have articulated, delegitimization of the status quo requires resilience against epistemic threats such as uncertainty. This implies that weak justifiers are better able to deal with epistemic ambiguity than strong justifiers, and thus have weaker epistemic needs. Thus, in the present experiment, stronger justifiers might also have been expected to be *less* comfortable not knowing.

The finding of a quadratic influence of SJG on comfort not knowing may illustrate these two possible explanations. I found that weak justifiers were the least comfortable not knowing, moderate justifiers were the most comfortable not knowing, and extreme justifiers were less comfortable not knowing than moderate justifiers. Thus, a moderate degree of system justification appeared to have the effect of allowing individuals to feel comfortable with not seeking information, perhaps due to a reduced desire or need to engage, which would be consistent with Hennes et al. (2012) and Jost et al. (2017). For extreme justifiers the effect was reversed; the highest justifiers were relatively less comfortable not knowing, and this could be attributable to stronger desires to resolve uncertainty, which would be consistent with the notion that strong justifiers have stronger epistemic needs. In this case system justification may have played a dual role by encouraging uncertainty resolution on the one hand, and on the other buffering individuals against the consequences of that resolution. That is, in the case of discovering unfairness, strong justifiers would be best able to assimilate that knowledge through legitimisation. Indeed, a strong justifier may view the option to find out as less threatening since they may already have accepted the legitimacy of the system.

There are, of course, other possible interpretations of the quadratic effect observed. To the best of my knowledge, only a handful of studies has documented curvilinear influences relating to system justification. Cichocka et al. (2018) investigated the influence of system confidence on a range of political engagement metrics (e.g., intentions to vote, actual voting behaviour, intentions to engage in collective action, and actual participation in collective action movements) and consistently identified a negative quadratic relationship between system justification and political engagement such that moderate levels of justification were typically associated with the greatest degree of engagement, whereas low and high levels of system justification were associated with decreased engagement. (It is noted by the authors that "system confidence" refers to a person's current satisfaction with the system, whilst "system justification" refers to a person's motivation to justify the system. The measures used to assess system confidence were nonetheless a subset of items from the System Justification General scale [items 1, 2 and 7] that I used to assess system justification in the present study. For this reason I use the two terms interchangeably, though I acknowledge that the two concepts can be considered distinct; e.g., Banfield et al., 2011.) Cichocka et al. (2018) suggested that the diverging effects of system justification on political engagement may be explained thus: decreased engagement may accompany increased system justification because higher justifiers should typically have a decreased desire for system change, whereas increased engagement from increased system justification may occur because strong justifiers typically have stronger beliefs in the efficacy of such engagement to elicit change.

Caricati (2019) investigated the relation between political orientation and attitudes toward the political system (including system justification) and found that this relationship was curvilinear; system justification was highest for moderate (i.e., central) political orientations and lower for extreme orientations (both left-leaning and right-leaning). This was interpreted as indicating that both extreme right and extreme left ideological orientations could lead to decreased support for the system. Jost and Kende (2020) similarly reported quadratic relationships between political orientation and system justification. However, these effects were very subtle and there was greater evidence for a linear effect supporting the notion that individuals on the right end of the political spectrum are typically stronger supporters of the system than those on the left.

The quadratic effect of system justification in the present experiment (***Figure 3.7***) pertains to the extent to which individuals were comfortable with not acquiring certain information about the system. The increased comfort evident amongst moderate justifiers compared to weak justifiers is entirely consistent with the well-established function of system justification in addressing epistemic needs. It may be argued that those more comfortable not knowing were more comfortable because their epistemic needs were satisfied by their confidence in the legitimacy of social structures (including the artificial experimental context) and the opportunity to believe that payments were made in a way that they preferred. This conclusion is consistent with the information avoidance literature that demonstrates that individuals can use ambiguity in information to maintain beliefs and avoid disproving them in social distribution contexts (Dana et al., 2007; Feiler, 2014; Grossman & van der Weele, 2017; Larson & Capra, 2009). This literature documents that one motivation for remaining in an ambiguous state is to maintain a belief (or excuse) that self-interested decisions are not wholly selfish. In the same way, I suggest that moderate justifiers in the present experiment were more comfortable not knowing because they were motivated to maintain a belief that payments were made in the way they preferred. Such an illusion (or delusion; Lerner, 1980) implies that moderate justifiers may have been prone to self-deception, and this is consistent with previous research documenting a relation between system justification and self-deception (Jost, Blount, et al., 2003).

Implicit also in my argument is that moderate system justification was the driver of the motivation to maintain ambiguity. This suggestion is somewhat at odds with the notion that system justification addresses epistemic needs to *reduce* uncertainty and ambiguity (e.g., Jost & Hunyady, 2005) and the evidence indicating a positive relation between conservatism and intolerance for uncertainty (e.g., Jost, Glaser, et al., 2003). However, I suggest that the line of reasoning is not inconsistent with SJT. Firstly, in the context of the present experiment, uncertainty and ambiguity were in the participants' control. This is conceptually distinct from the uncertainty borne from the chaos of a random and unpredictable world, which is perhaps closer to what measures such as the Need for Closure Scale (Webster & Kruglanski, 1994) and others which have been used to explore epistemic needs are aimed. Secondly, I suggest that, in light of the above, the use or disuse of information in the experiment nonetheless pertains to epistemic motives. In a sense, the overarching motivations relate to the use of information as a means to an end. I argue that justifying tendencies led moderate justifiers to make use of ambiguity to achieve the goal of palliation. The fact that participants were motivated to maintain rather than reduce ambiguity takes nothing away from the tenet that the means and end of this process were epistemic in nature.

Explanation of the subsequent decreases in comfort for particularly strong justifiers involves more speculation. As Cichocka et al. (2018) suggest, the relation between system justification and engagement may involve at least two processes that manifest at different points of the system justification spectrum. Whereas moderate justifiers' epistemic needs were relatively satisfied, it appears that the needs of strong justifiers were less so. One possibility is that extreme levels of system justification represent such a strong tendency to support existing social structures that it resembles unconditional loyalty to the system. In such cases, strong justification affords confidence in the system to such an extent that its policies become relatively less important (an individual representing the utmost extreme of this might be understood as a "fanatic"). This could paradoxically make extreme justifiers less comfortable not knowing about the payments; for these individuals, knowing for certain is a more effective way of satisfying epistemic needs than remaining uncertain but confident in the legitimacy of the system. Put differently, I suggest that moderate justifiers were able to sustain themselves in a state of ambiguity, and that this afforded them the opportunity to believe that payments were made in the way that they preferred. Doing so served a palliative function by allowing individuals to sustain the illusion. In contrast, extreme justifiers placed less value in this ambiguous state and, to the extent that their comfort not knowing about payments influenced decisions, appeared to value knowing relatively more because they would accept the system whichever way it was.

### Interpersonal concerns

The extent to which individuals were concerned with the outcomes of others (i.e., Social Value Orientation; SVO) was an important individual difference characteristic that influenced whether individuals approached information. As was the case with system justification measures, the influence of SVO was moderated by whether the individual was potentially advantaged or disadvantaged; pro-sociality (as indexed by higher SVO scores) was associated with a lower likelihood of finding amongst potentially disadvantaged participants relative to potentially advantaged participants, and there was some evidence to suggest this was mainly driven by a negative influence of SVO amongst potentially disadvantaged participants (***Figure 3.2***). Based on these findings, I suggest that prosocial motivations manifest in different ways depending on the individual's relative social position. In essence, potentially disadvantaged pro-socials were more concerned with avoiding information, whilst potentially advantaged pro-socials were more concerned with seeking information.

Previous research has explored the relationship between SVO and information-seeking. Fiedler et al. (2013) showed that deviation from individualistic SVO predicts greater information search. Using an eye-tracking paradigm, these researchers showed that the extent to which individuals engaged with information was positively related to the extent to which their SVO score deviated from pure individualism – both positively (i.e., towards cooperation) and negatively (i.e., towards competitiveness). In addition, the authors noted that these SVO-related changes were gradual, providing support for the notion that differences in SVO alter the relative weights that individuals apply to cooperative and individualistic goals rather than leading to qualitatively distinct decision strategies.

In the present experiment, I similarly found evidence implicating SVO with information-seeking (note however that almost all participants were either individualistic or cooperative – only one participant fell in the competitive range). The direction of SVO's influence depended on whether the individual was potentially advantaged or disadvantaged. For potentially advantaged participants, pro-sociality was associated with information seeking to a greater extent than an individualistic orientation. For potentially disadvantaged participants this pattern was reversed; greater pro-sociality was associated with information seeking to a lesser degree than individualistic orientations. To the extent that participant decisions were manifestations of SVO-derived goals, it therefore appears that potentially advantaged and disadvantaged participants weighted the value of each decision differently; whereas the "pro-social goal" for potentially advantaged participants was typically pursued by finding out, the pro-social goal for potentially disadvantaged participants was typically pursued by not finding out.

Inferences about the motivations of individuals can be further derived from the secondary SVO measures: Distance from archetypal Inequality Aversion (DIA) – the extent to which participants' SVO choices differed from the choices representing pure inequality aversion – and Distance from archetypal Joint Gain maximisation (DJG) – the extent to which their choices differed from choices that maximised communal wealth, regardless of distribution. Both potentially advantaged and potentially disadvantaged individuals were influenced by DIA. In contrast, the pattern of results for the influence of DJG was very similar to the overall SVO pattern. A negative influence of DJG was observed amongst potentially advantaged participants such that greater concern for joint gains was associated with a greater likelihood of finding out (recall that smaller values of DJG represent greater concern for joint gains). This suggests that the above-discussed concern for finding out amongst pro-social potentially advantaged participants was in some part driven by these participants' concerns for maximising communal wealth; ascertaining how much their potentially disadvantaged partner was paid allowed these participants to know with certainty how much wealth was accumulated by the two of them. In contrast, participants with a strong inequality aversion were no more likely to find out as they were not to find out (***Figure 3.4***), suggesting that inequality aversion was not strongly implicated in potentially advantaged pro-socials' decisions to find out.

Potentially disadvantaged participants who valued joint gains were relatively less likely to find out compared to their Advantaged colleagues. This suggests that concerns for joint gain maximisation amongst these participants were realised by avoiding finding out. In this way, a belief about joint gains having been maximised could be maintained by avoiding information that might contradict this belief. This suggestion is somewhat supported by the finding that the comments of potentially disadvantaged participants regarding their motivations were largely focused on the themes 'Avoid affective reaction' and 'Payment satisfaction' (***Figure 3.10***). As noted in the Results section, some comments explicitly referred to the desire to maintain feelings of payment satisfaction by avoiding information that would make them feel bad about their payments.

Similar distinctions have been noted by others. Mischkowski et al. (2019) showed that the influence of SVO on inequality aversion depended on whether the individual was disadvantaged by the inequality or advantaged by it. By distinguishing between joint gains that advantage the self and joint gains that disadvantage the self, Mischkowski et al. (2019) demonstrated that if a joint gain distribution was disadvantageous (i.e., joint maximisation was achieved by personally receiving less), preference for joint gains increased with SVO; that is, joint gains were valued by disadvantaged pro-socials but not by disadvantaged individualists. In contrast, if a joint gain distribution was advantageous (i.e., joint maximisation was achieved by personally receiving more), preference for joint gains *decreased* with SVO; that is, joint gains were less valued by advantaged pro-socials than advantaged individualists. Though the exact arrangement of variables may differ, these findings are broadly consistent with those reported here, insofar as they support a distinction between the motivations of pro-socials based on their relative social position.

**Age**

Age was found to be an important predictor of several of the outcome measures investigated. First and foremost, it negatively predicted the likelihood of deciding to find out; older individuals were less likely to find out compared to younger individuals. Furthermore, older individuals reported being more comfortable not knowing about the payments. The finding that older individuals were less likely to find out is somewhat consistent with previous findings documenting differences in information search in a distributive justice context between children and adolescents and university aged individuals. In a dictator game in which participants could choose to learn or ignore payoff information for themselves and/or the recipient, older participants sought less information than younger participants (Brocas & Carrillo, 2020). Moreover, the majority of dictators only discovering their own payoff were university students, whereas there were fewer university students than children from the youngest age group who discovered only the recipient's payoff. This provided evidence that older individuals were more prone to strategically avoid information that would indicate whether the distribution of wealth was fair, and in so doing made more selfish distribution decisions, presumably without feeling guilt (since the unfairness was not known).

My finding adds to the above finding by demonstrating that a preference for avoiding fairness information continues to grow throughout adulthood. Whereas Brocas and Carillo (2020) found that the avoidance of fairness information increases in the relatively short timespan of childhood to university age, I found that a similar increase in avoidance is evident from 18 to 64 years of age. Definitive explanation for why this was the case is not possible for the current data, but one can speculate. One possibility is that individuals become more realistic (or cynical) with age. This is consistent with documented age-related decreases in belief in a just world (Dalbert, 2001; Oppenheimer, 2005, 2006). Such findings suggest that with advancing age and growing experience, individuals become better equipped to draw on personal judgement and rely less on general beliefs to derive purpose and predictability. Though in the present study no relation was observed between Belief in a Just World and Age, I did observe that older individuals appear to be more tolerant of uncertainty (***Table 3.3***), which suggests that in one form or another, older individuals were better able to deal with the epistemic challenge of uncertainty than younger individuals. Indeed, this is supported by the finding that older individuals were more comfortable not knowing about the unfairness. It is possible that older individuals were less concerned with others' payoffs, supporting an interpretation of older individuals as more morally cynical, consistent with Brocas and Carillo (2020). However, if this was the case it might have been expected to observe an age-related decrease in Social Value Orientation, which was not the case. It therefore appears that the present results are best interpreted as the result of lower degree of epistemic uncertainty in older individuals.

Since the data here were cross-sectional, there is an important caveat in the age findings. It cannot be ascertained whether the age effects are due developmental processes or to generational differences. This concern is particularly relevant at this point in history, where younger generations are increasingly liberal, perhaps due to greater rates of university education (Stubager, 2008). Understanding why exactly older individuals appear to be less concerned with seeking social fairness information is a priority for future research, not least because globally populations are ageing.

Overall, the findings of the present study add to the literature on the motivated avoidance of information by demonstrating that in the context of remuneration for work several individual difference characteristics predict whether people will seek or avoid social equality information. Though it is possible that some participants were simply indifferent (indeed, many stated simple “curiosity” as an explanation for their decision), the myriad findings relating dispositional tendencies to ignorance strongly implicates *motivation* in avoiding such information. My results therefore accord with studies documenting self-ignorance as strategic (e.g., Dana et al., 2007; Larson & Capra, 2009; Thunström et al., 2016). In addition, the findings presented here provide person- and context-specific conditions under which this motivation is particularly likely to be manifest. Specifically, I have presented evidence to suggest that those who strongly support the meritocratic ideals of capitalism (i.e., high SJE) are less likely to find out about social inequality if they stand to be advantaged by it, but more likely to seek that information if they stand to be disadvantaged by it. In addition, those who are by disposition pro-social (i.e., high SVO) will seek information about social equality if they could be advantaged by it but avoid information if they could be disadvantaged by it. These findings highlight the situational and personal complexity of the processes behind seeking or avoiding social equality information and highlight their highly interactive nature. Beyond self-interest, they point to a constellation of psychological needs and motivations that underlie whether individuals engage with social equality information.

## References

Bahamondes-Correa, J. (2016). System justification's opposite effects on psychological wellbeing: testing a moderated mediation model in a gay men and lesbian sample in chile. *Journal of Homosexuality*, *63*(11), 1537–1555. https://doi.org/10.1080/00918369.2016.1223351

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, *12*(4), 533–547. https://doi.org/10.1177/1368430209105040

Banfield, J. C., Kay, A. C., Cutright, K. M., Wu, E. C., & Fitzsimons, G. J. (2011). A Person by situation account of motivated system defense. *Social Psychological and Personality Science*, *2*(2), 212–219. https://doi.org/10.1177/1948550610386809

Bartholomaeus, J., & Strelan, P. (2019). The adaptive, approach-oriented correlates of belief in a just world for the self: A review of the research. *Personality and Individual Differences*, *151*, 109485. https://doi.org/10.1016/j.paid.2019.06.028

Bègue, L., & Bastounis, M. (2003). Two spheres of belief in justice: Extensive support for the bidimensional model of belief in a just world. *Journal of Personality*, *71*(3), 435–463. https://doi.org/10.1111/1467-6494.7103007

Bègue, L., Charmoillaux, M., Cochet, J., Cury, C., & de Suremain, F. (2008). Altruistic behavior and the bidimensional just world belief. *The American Journal of Psychology*, *121*(1), 47–56. https://doi.org/10.2307/20445443

Beierlein, C., Werner, C. S., Preiser, S., & Wermuth, S. (2011). Are just-world beliefs compatible with justifying inequality? collective political efficacy as a moderator. *Social Justice Research*, *24*(3), 278-296. https://doi.org/10.1007/s11211-011-0139-2

Birrell, J., Meares, K., Wilkinson, A., & Freeston, M. (2011). Toward a definition of intolerance of uncertainty: A review of factor analytical studies of the Intolerance of Uncertainty Scale. *Clinical Psychology Review*, *31*(7), 1198–1208. https://doi.org/10.1016/j.cpr.2011.07.009

Bolton, G. E., Katok, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, *27*(2), 269–299. https://doi.org/10.1007/s001820050072

Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior*, *10*(1), 95–121. https://doi.org/10.1006/game.1995.1026

Bonanno, G. A., & Jost, J. T. (2006). Conservative shift among high-exposure survivors of the september 11th terrorist attacks. *Basic and Applied Social Psychology*, *28*(4), 311–323. https://doi.org/10.1207/s15324834basp2804_4

Brandt, M. J. (2013). Do the disadvantaged legitimize the social system? A large-scale test of the status–legitimacy hypothesis. *Journal of Personality and Social Psychology*, *104*(5), 765–785. https://doi.org/10.1037/a0031751

Brissette, I., Scheier, M. F., & Carver, C. S. (2002). The role of optimism in social network development, coping, and psychological adjustment during a life transition. *Journal of Personality and Social Psychology*, *82*(1), 102–111. https://doi.org/10.1037/0022-3514.82.1.102

Brocas, I., & Carrillo, J. D. (2020). The development of social strategic ignorance and other regarding behavior from childhood to adulthood. *Journal of Behavioral and Experimental Economics*, *85*, 101524. https://doi.org/10.1016/j.socec.2020.101524

Brosnan, S. F., & de Waal, F. B. (2014). Evolution of responses to (un) fairness. *Science*, *346*(6207). https://doi.org/10.1126/science.1251776

Brosnan, S. F., Talbot, C., Ahlgren, M., Lambeth, S. P., & Schapiro, S. J. (2010). Mechanisms underlying responses to inequitable outcomes in chimpanzees, Pan troglodytes. *Animal Behaviour*, *79*(6), 1229–1237. https://doi.org/10.1016/j.anbehav.2010.02.019

Brouwers, M. C., & Sorrentino, R. M. (1993). Uncertainty orientation and protection motivation theory: The role of individual differences in health compliance. *Journal of Personality and Social Psychology*, *65*(1), 102–112. https://doi.org/10.1037/0022-3514.65.1.102

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Camac, C. (1992). Information preferences in a two-person social dilemma. In W. Liebrand, D. Messick, & H. Wilke (Eds.), *Social dilemmas: Theoretical issues and research findings* (pp. 147–161). Pergamon Press.

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives*, *9*(2), 209–219. https://doi.org/10.1257/jep.9.2.209

Caricati, L. (2019). Evidence of decreased system justification among extreme conservatives in non-American samples. *The Journal of Social Psychology*, *159*(6), 725–745. https://doi.org/10.1080/00224545.2019.1567455

Caricati, L., & Lorenzi-Cioldi, F. (2012). Does status matter? Testing hypotheses from strong form of System Justification Theory. *Revue internationale de psychologie sociale, Tome 25*(1), 67–95.

Carleton, R. N., Norton, M. P. J., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, *21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014

Carver, C. S., & Gaines, J. G. (1987). Optimism, pessimism, and postpartum depression. *Cognitive Therapy and Research*, *11*(4), 449–462. https://doi.org/10.1007/BF01175355

Carver, C. S., Peterson, L. M., Follansbee, D. J., & Scheier, M. F. (1983). Effects of self-directed attention on performance and persistence among persons high and low in test anxiety. *Cognitive Therapy and Research*, *7*(4), 333–353. https://doi.org/10.1007/BF01177556

Carver, C. S., & Scheier, M. F. (2014). Dispositional optimism. *Trends in Cognitive Sciences*, *18*(6), 293–299. https://doi.org/10.1016/j.tics.2014.02.003

Carver, C. S., Scheier, M. F., & Segerstrom, S. C. (2010). Optimism. *Clinical Psychology Review*, *30*(7), 879–889. https://doi.org/10.1016/j.cpr.2010.01.006

Carver, C. S., Smith, R. G., Antoni, M. H., Petronis, V. M., Weiss, S., & Derhagopian, R. P. (2005). Optimistic Personality and psychosocial well-being during treatment predict psychosocial well-being among long-term survivors of breast cancer. *Health Psychology*, *24*(5), 508–516. https://doi.org/10.1037/0278-6133.24.5.508

Carver, C. S., Weintraub, J. K., & Scheier, M. F. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology*, *56*, 267–283. https://doi.org/10.1037/0022-3514.56.2.267

Cichocka, A., Górska, P., Jost, J. T., Sutton, R. M., & Bilewicz, M. (2018). What inverted U can do for your country: A curvilinear relationship between confidence in the social system and political engagement. *Journal of Personality and Social Psychology, 115*(5)*,* 883–902. http://dx.doi.org/10.1037/pspp0000168

Correia, I., & Dalbert, C. (2008). School bullying: Belief in a personal just world of bullies, victims, and defenders. *European Psychologist*, *13*(4), 248–254. https://doi.org/10.1027/1016-9040.13.4.248

Cutler, S. J., & Hodgson, L. G. (2003). To test or not to test: Interest in genetic testing for Alzheimer's disease among middle-aged adults. *American Journal of Alzheimer's Disease & Other Dementias, 18*(1)*,* 9–20. https://doi.org/10.1177/153331750301800106

Dalbert, C. (1999). The world is more just for me than generally: About the personal belief in a just world scale's validity. *Social Justice Research*, *12*(2), 79–98. https://doi.org/10.1023/A:1022091609047

Dalbert, C. (2001). *The Justice Motive as a Personal Resource*. Springer US. https://doi.org/10.1007/978-1-4757-3383-9

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80. https://doi.org/10.1007/s00199-006-0153-z

Daukantaite, D., & Bergman, L. R. (2005). Childhood roots of women's subjective well-being: The role of optimism. *European Psychologist*, *10*(4), 287–297.

De Cremer, D., & Van Lange, P. A. M. (2001). Why prosocials exhibit greater cooperation than proselfs: The roles of social responsibility and reciprocity. *European Journal of Personality*, *15*(1_suppl), S5–S18. https://doi.org/10.1002/per.418

Durrheim, K., Jacobs, N., & Dixon, J. (2014). Explaining the paradoxical effects of intergroup contact: Paternalistic relations and system justification in domestic labour in South Africa. *International Journal of Intercultural Relations*, *41*, 150–164. https://doi.org/10.1016/j.ijintrel.2013.11.006

Dzuka, J., & Dalbert, C. (2002). Mental health and personality of slovak unemployed adolescents: The impact of belief in a just world. *Journal of Applied Social Psychology*, *32*(4), 732–757. https://doi.org/10.1111/j.1559-1816.2002.tb00240.x

Eckel, C. C., & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior*, *16*(2), 181–191. https://doi.org/10.1006/game.1996.0081

Eidelman, S., Crandall, C. S., Goodman, J. A., & Blanchar, J. C. (2012). Low-effort thought promotes political conservatism. *Personality and Social Psychology Bulletin*, *38*(6), 808–820. https://doi.org/10.1177/0146167212439213

Endler, N. S., & Parker, J. D. (1990). Multidimensional assessment of coping: A critical evaluation. *Journal of Personality and Social Psychology*, *58*(5), 844–854. https://doi.org/10.1037/0022-3514.58.5.844

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191. https://doi.org/10.3758/BF03193146

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*(7208), 1079–1083. https://doi.org/10.1038/nature07155

Feiler, L. (2014). Testing models of information avoidance with binary choice dictator games. *Journal of Economic Psychology*, *45*, 253–267. https://doi.org/10.1016/j.joep.2014.10.003

Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social Value Orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes*, *120*(2), 272–284. https://doi.org/10.1016/j.obhdp.2012.07.002

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*(3), 347–369. https://doi.org/10.1006/game.1994.1021

Friis, L. S., Elverdam, B., & Schmidt, K. G. (2003). The patient's perspective. *Supportive Care in Cancer*, *11*(3), 162–170. https://doi.org/10.1007/s00520-002-0424-6

Furnham, A. (2003). Belief in a just world: Research progress over the past decade. *Personality and Individual Differences*, *34*(5), 795–817. https://doi.org/10.1016/S0191-8869(02)00072-7

Godfrey, E. B., & Wolf, S. (2016). Developing critical consciousness or justifying the system? A qualitative analysis of attributions for poverty and wealth among low-income racial/ethnic minority and immigrant women. *Cultural Diversity and Ethnic Minority Psychology*, *22*(1), 93–103. https://doi.org/10.1037/cdp0000048

Greco, V., & Roger, D. (2003). Uncertainty, stress, and health. *Personality and Individual Differences*, *34*(6), 1057–1068. https://doi.org/10.1016/S0191-8869(02)00091-0

Grossman, Z., & van der Weele, J. J. (2017). Self-Image and willful ignorance in social decisions. *Journal of the European Economic Association*, *15*(1), 173–217. https://doi.org/10.1093/jeea/jvw001

Guala, F., & Mittone, L. (2010). Paradigmatic experiments: The Dictator Game. *The Journal of Socio-Economics*, *39*(5), 578–584. https://doi.org/10.1016/j.socec.2009.05.007

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7

Hammond, M. D., & Sibley, C. G. (2011). Why are benevolent sexists happier? *Sex Roles*, *65*(5), 332–343. https://doi.org/10.1007/s11199-011-0017-2

Haruno, M., & Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*(2), 160–161. https://doi.org/10.1038/nn.2468

Hegtvedt, K. A. (1990). The effects of relationship structure on emotional responses to inequity. *Social Psychology Quarterly*, *53*(3), 214–228. https://doi.org/10.2307/2786960

Hennes, E. P., Nam, H. H., Stern, C., & Jost, J. T. (2012). Not all ideologies are created equal: epistemic, existential, and relational needs predict system-justifying attitudes. *Social Cognition*, *30*(6), 669–688. https://doi.org/10.1521/soco.2012.30.6.669

Henrich, J. P., Boyd, R., Bowles, S., Fehr, E., Camerer, C., & Gintis, H. (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford University Press on Demand.

Henry, P. J., & Saul, A. (2006). The development of system justification in the developing world. *Social Justice Research*, *19*(3), 365–378. https://doi.org/10.1007/s11211-006-0012-x

Hess, Y. D., & Ledgerwood, A. (2014). Bolstering system-justifying beliefs in response to social exclusion. *Group Processes & Intergroup Relations*, *17*(4), 494–508. https://doi.org/10.1177/1368430213510572

Joireman, J. A., Lasane, T. P., Bennett, J., Richards, D., & Solaimani, S. (2001). Integrating social value orientation and the consideration of future consequences within the extended norm activation model of proenvironmental behaviour. *British Journal of Social Psychology*, *40*(1), 133–155. https://doi.org/10.1348/014466601164731

Jost, J., & Hunyady, O. (2003). The psychology of system justification and the palliative function of ideology. *European Review of Social Psychology, 13*(1), 111–153. https://doi.org/10.1080/10463280240000046

Jost, J. T. (2017). Working class conservatism: A system justification perspective. *Current Opinion in Psychology, 18*, 73–78. https://doi.org/10.1016/j.copsyc.2017.08.020

Jost, J. T. (2019). A quarter century of system justification theory: Questions, answers, criticisms, and societal applications. *British Journal of Social Psychology*, *58*(2), 263–314. https://doi.org/10.1111/bjso.12297

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1–27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x

Jost, J. T., Blount, S., Pfeffer, J., & Hunyady, G. (2003). Fair market ideology: its cognitive-motivational underpinnings. *Research in Organizational Behavior*, *25*, 53–91. https://doi.org/10.1016/S0191-3085(03)25002-4

Jost, J. T., & Burgess, D. (2000). Attitudinal ambivalence and the conflict between group and system justification motives in low status groups. *Personality and Social Psychology Bulletin*, *26*(3), 293–305. https://doi.org/10.1177/0146167200265003

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339Jost, J. T., & Hunyady, O. (2005). Antecedents and Consequences of System-Justifying Ideologies. *Current Directions in Psychological Science*, *14*(5), 260–265. https://doi.org/10.1111/j.0963-7214.2005.00377.x

Jost, J. T., & Kende, A. (2020). Setting the record straight: System justification and rigidity-of-the-right in contemporary Hungarian politics. *International Journal of Psychology*, *55*(S1), 96–115. https://doi.org/10.1002/ijop.12631

Jost, J. T., Langer, M., Badaan, V., Azevedo, F., Etchezahar, E., Ungaretti, J., & Hennes, E. P. (2017). Ideology and the limits of self-interest: System justification motivation and conservative advantages in mass politics. *Translational Issues in Psychological Science*, *3*(3), e1. https://doi.org/10.1037/tps0000127

Jost, J. T., Ledgerwood, A., & Hardin, C. D. (2008). Shared reality, system justification, and the relational basis of ideological beliefs. *Social and Personality Psychology Compass*, *2*(1), 171–186. https://doi.org/10.1111/j.1751-9004.2007.00056.x

Jost, J. T., Napier, J. L., Thorisdottir, H., Gosling, S. D., Palfai, T. P., & Ostafin, B. (2007). Are needs to manage uncertainty and threat associated with political conservatism or ideological extremity? *Personality and Social Psychology Bulletin*, *33*(7), 989–1007. https://doi.org/10.1177/0146167207301028

Jost, J. T., Pelham, B. W., Sheldon, O., & Sullivan, B. N. (2003). Social inequality and the reduction of ideological dissonance on behalf of the system: Evidence of enhanced system justification among the disadvantaged. *European Journal of Social Psychology*, *33*(1), 13–36. https://doi.org/10.1002/ejsp.127

Jost, J. T., & Thompson, E. P. (2000). Group-based dominance and opposition to equality as independent predictors of self-esteem, ethnocentrism, and social policy attitudes among African Americans and European Americans. *Journal of Experimental Social Psychology*, *36*(3), 209–232. https://doi.org/10.1006/jesp.1999.1403

Jost, J. T., Wakslak, C. J., & Tyler, T. R. (2008). System justification theory and the alleviation of emotional distress: Palliative effects of ideology in an arbitrary social hierarchy and in society. In K. A. Hegtvedt & J. Clay-Warner (Eds.), *Justice* (Vol. 25, pp. 181–211). Emerald Group Publishing Limited. https://doi.org/10.1016/S0882-6145(08)25012-5

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kay, A. C., Gaucher, D., Peach, J. M., Laurin, K., Friesen, J., Zanna, M. P., & Spencer, S. J. (2009). Inequality, discrimination, and the power of the status quo: Direct evidence for a motivation to see the way things are as the way they should be. *Journal of Personality and Social Psychology*, *97*(3), 421–434. http://dx.doi.org/10.1037/a0015997

Kay, A. C., & Jost, J. T. (2003). Complementary justice: Effects of 'poor but happy' and 'poor but honest' stereotype exemplars on system justification and implicit activation of the justice motive. *Journal of Personality and Social Psychology*, *85*(5), 823–837. https://doi.org/10.1037/0022-3514.85.5.823

Koerner, N., & Dugas, M. J. (2008). An investigation of appraisals in individuals vulnerable to excessive worry: The role of intolerance of uncertainty. *Cognitive Therapy and Research*, *32*(5), 619–638. https://doi.org/10.1007/s10608-007-9125-2

Kramer, R. M., McClintock, C. G., & Messick, D. M. (1986). Social values and cooperative response to a simulated resource conservation crisis. *Journal of Personality*, *54*(3), 576–582. https://doi.org/10.1111/j.1467-6494.1986.tb00413.x

Larson, T., & Capra, C. M. (2009). Exploiting moral wiggle room: Illusory preference for fairness? A comment. *Judgment and Decision Making*, *4*(6), 467–474.

Lerner, M. J. (1980). The Belief in a Just World. In M. J. Lerner (Ed.), *The Belief in a Just World: A Fundamental Delusion* (pp. 9–30). Springer US. https://doi.org/10.1007/978-1-4899-0448-5_2

Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, *85*(5), 1030–1051. https://doi.org/10.1037/0033-2909.85.5.1030

Lerner, M. J., & Simmons, C. H. (1966). Observer's reaction to the "innocent victim": Compassion or rejection? *Journal of Personality and Social Psychology*, *4*(2), 203. https://doi.org/10.1037/h0023562

Liebrand, W. B. G. (1984). The effect of social motives, communication and group size on behaviour in an N-person multi-stage mixed-motive game. *European Journal of Social Psychology*, *14*(3), 239–264. https://doi.org/10.1002/ejsp.2420140302

Lyter, D. W., Valdiserri, R. O., Kingsley, L. A., Amoroso, W. P., & Rinaldo, C. R. (1987). The HIV antibody test: Why gay and bisexual men want or do not want to know their results. *Public Health Reports*, *102*(5), 468–474.

McCoy, S. K., Wellman, J. D., Cosley, B., Saslow, L., & Epel, E. (2013). Is the belief in meritocracy palliative for members of low status groups? Evidence for a benefit for self-esteem and physical health via perceived control. *European Journal of Social Psychology*, *43*(4), 307–318. https://doi.org/10.1002/ejsp.1959

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, *4*(1), 1–25. https://doi.org/10.1016/0022-1031(68)90046-2

Mischkowski, D., Thielmann, I., & Glöckner, A. (2019). Minimizing inequality versus maximizing joint gains: On the relation between personality traits and different prosocial motivations. *Journal of Research in Personality*, *81*, 153–157. https://doi.org/10.1016/j.jrp.2019.06.001

Murphy, R. O., & Ackermann, K. A. (2014). Social Value Orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, *18*(1), 13–41. https://doi.org/10.1177/1088868313501745

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making*, *6*(8), 771–781. http://dx.doi.org/10.2139/ssrn.1804189

Nail, P. R., McGregor, I., Drinkwater, A. E., Steele, G. M., & Thompson, A. W. (2009). Threat causes liberals to think like conservatives. *Journal of Experimental Social Psychology*, *45*(4), 901–907. https://doi.org/10.1016/j.jesp.2009.04.013

Nes, L. S., & Segerstrom, S. C. (2006). Dispositional optimism and coping: A meta-analytic review. *Personality and Social Psychology Review*, *10*(3), 235–251. https://doi.org/10.1207/s15327957pspr1003_3

Nes, L. S., Segerstrom, S. C., & Sephton, S. E. (2005). Engagement and arousal: Optimism's effects during a brief stressor. *Personality and Social Psychology Bulletin*, *31*(1), 111–120. https://doi.org/10.1177/0146167204271319

Nudelman, G. (2013). The Belief in a Just World and personality: A meta-analysis. *Social Justice Research*, *26*(2), 105–119. https://doi.org/10.1007/s11211-013-0178-y

Nudelman, G., Otto, K., & Dalbert, C. (2016). Can belief in a just world buffer mood and career prospects of people in need of risk protection? First experimental evidence. *Risk Analysis*, *36*(12), 2247–2257. https://doi.org/10.1111/risa.12588

Oosterbeek, H., Sloof, R., & van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: evidence from a meta-analysis. *Experimental Economics*, *7*(2), 171–188. https://doi.org/10.1023/B:EXEC.0000026978.14316.74

Oppenheimer, L. (2005). Justice and the belief in a just world: A developmental perspective. *Personality and Individual Differences*, *38*(8), 1793–1803. https://doi.org/10.1016/j.paid.2004.11.007

Oppenheimer, L. (2006). The belief in a just world and subjective perceptions of society: A developmental perspective. *Journal of Adolescence*, *29*(4), 655–669. https://doi.org/10.1016/j.adolescence.2005.08.014

Osborne, D., & Sibley, C. G. (2013). Through rose-colored glasses: System-justifying beliefs dampen the effects of relative deprivation on well-being and political mobilization. *Personality and Social Psychology Bulletin*, *39*(8), 991–1004. https://doi.org/10.1177/0146167213487997

Otto, K., Boos, A., Dalbert, C., Schöps, D., & Hoyer, J. (2006). Posttraumatic symptoms, depression, and anxiety of flood victims: The impact of the belief in a just world. *Personality and Individual Differences*, *40*(5), 1075–1084. https://doi.org/10.1016/j.paid.2005.11.010

Otto, K., Glaser, D., & Dalbert, C. (2009). Mental health, occupational trust, and quality of working life: Does belief in a just world matter? *Journal of Applied Social Psychology*, *39*(6), 1288–1315. https://doi.org/10.1111/j.1559-1816.2009.00482.x

Range, F., Horn, L., Viranyi, Z., & Huber, L. (2009). The absence of reward induces inequity aversion in dogs. *Proceedings of the National Academy of Sciences*, *106*(1), 340–345. https://doi.org/10.1073/pnas.0810957105

Richter, M., & König, C. J. (2017). Explaining individuals' justification of layoffs. *Journal of Applied Social Psychology*, *47*(6), 331–346. https://doi.org/10.1111/jasp.12442

Rosen, N. O., & Knäuper, B. (2009). A little uncertainty goes a long way: State and trait differences in uncertainty interact to increase information seeking but also increase worry. *Health Communication*, *24*(3), 228–238. https://doi.org/10.1080/10410230902804125

Rosen, N. O., Knäuper, B., & Sammut, J. (2007). Do individual differences in intolerance of uncertainty affect health monitoring? *Psychology & Health*, *22*(4), 413–430. https://doi.org/10.1080/14768320600941038

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758. https://doi.org/10.1126/science.1082976

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063-1078. https://doi.org/10.1037/0022-3514.67.6.1063

Scheier, M. F., Carver, C. S., & Bridges, M. W. (2001). Optimism, pessimism, and psychological well-being. In E. C. Chang (Ed.), *Optimism & pessimism: Implications for theory, research, and practice* (pp. 189–216). American Psychological Association.

Şeker, B. D. (2016). Evaluation of life satisfaction after the 2011 Van (Turkey) earthquake. *Social Behavior and Personality: An International Journal*, *44*(9), 1409–1417. https://doi.org/10.2224/sbp.2016.44.9.1409

Sengupta, N. K., Osborne, D., & Sibley, C. G. (2015). The status-legitimacy hypothesis revisited: Ethnic-group differences in general and dimension-specific legitimacy. *British Journal of Social Psychology*, *54*(2), 324–340. https://doi.org/10.1111/bjso.12080

Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General*, *141*(2), 382-395. https://doi.org/10.1037/a0025907

Shipp, T. D., Shipp, D. Z., Bromley, B., Sheahan, R., Cohen, A., Lieberman, E., & Benacerraf, B. (2004). What factors are associated with parents' desire to know the sex of their unborn child? *Birth*, *31*(4), 272–279. https://doi.org/10.1111/j.0730-7659.2004.00319.x

Siebenhaar, K. U., Köther, A. K., & Alpers, G. W. (2020). Dealing with the covid-19 infodemic: Distress by information, information avoidance, and compliance with preventive measures. *Frontiers in Psychology*, *11*, 2981. https://doi.org/10.3389/fpsyg.2020.567905

Smeesters, D., Warlop, L., Van Avermaet, E., Corneille, O., & Yzerbyt, V. (2003). Do not prime hawks with doves: The interplay of construct activation and consistency of social value orientation on cooperative behavior. *Journal of Personality and Social Psychology*, *84*(5), 972–987. https://doi.org/10.1037/0022-3514.84.5.972

Song, S., Yao, X., & Wen, N. (2021). What motivates Chinese consumers to avoid information about the COVID-19 pandemic? The perspective of the stimulus-organism-response model. *Information Processing & Management*, *58*(1), 102407. https://doi.org/10.1016/j.ipm.2020.102407

Strelan, P. (2018). Justice and forgiveness in interpersonal relationships. *Current Directions in Psychological Science*, *27*(1), 20–24. https://doi.org/10.1177/0963721417734311

Stubager, R. (2008). Education effects on authoritarian–libertarian values: A question of socialization1. *The British Journal of Sociology*, *59*(2), 327–350. https://doi.org/10.1111/j.1468-4446.2008.00196.x

Suls, J., & Fletcher, B. (1985). The relative efficacy of avoidant and nonavoidant coping strategies: A meta-analysis. *Health Psychology*, *4*(3), 249–288. https://doi.org/10.1037/0278-6133.4.3.249

Suppes, A., Napier, J. L., & van der Toorn, J. (2019). The palliative effects of system justification on the health and happiness of lesbian, gay, bisexual, and transgender individuals. *Personality and Social Psychology Bulletin*, *45*(3), 372–388. https://doi.org/10.1177/0146167218785156

Sutton, R. M., & Douglas, K. M. (2005). Justice for all, or just for me? More evidence of the importance of the self-other distinction in just-world beliefs. *Personality and Individual Differences*, *39*(3), 637–645. https://doi.org/10.1016/j.paid.2005.02.010

Sutton, R. M., Douglas, K. M., Wilkin, K., Elder, T. J., Cole, J. M., & Stathi, S. (2008). Justice for whom, exactly? beliefs in justice for the self and various others. *Personality and Social Psychology Bulletin*, *34*(4), 528–541. https://doi.org/10.1177/0146167207312526

Sutton, R. M., Stoeber, J., & Kamble, S. V. (2017). Belief in a just world for oneself versus others, social goals, and subjective well-being. *Personality and Individual Differences*, *113*, 115–119. https://doi.org/10.1016/j.paid.2017.03.026

Sutton, R. M., & Winnard, E. J. (2007). Looking ahead through lenses of justice: The relevance of just-world beliefs to intentions and confidence in the future. *British Journal of Social Psychology*, *46*(3), 649–666. https://doi.org/10.1348/014466606X166220

Thaler, R. H. (1988). Anomalies: The Ultimatum Game. *Journal of Economic Perspectives*, *2*(4), 195–206. https://doi.org/10.1257/jep.2.4.195

Thompson, H. S., Valdimarsdottir, H. B., Duteau-Buck, C., Guevarra, J., Bovbjerg, D. H., Richmond-Avellaneda, C., Amarel, D., Godfrey, D., Brown, K., & Offit, K. (2002). Psychosocial predictors of BRCA counseling and testing decisions among urban African-American women. *Cancer Epidemiology and Prevention Biomarkers*, *11*(12), 1579–1585.

Thórisdóttir, H., & Jost, J. T. (2011). Motivated closed-mindedness mediates the effect of threat on political conservatism. *Political Psychology*, *32*(5), 785–811. https://doi.org/10.1111/j.1467-9221.2011.00840.x

Thunström, L., Nordström, J., Shogren, J. F., Ehmke, M., & van't Veld, K. (2016). Strategic self-ignorance. *Journal of Risk and Uncertainty*, *52*(2), 117–136. https://doi.org/10.1007/s11166-016-9236-9

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Van Berkel, L., Crandall, C. S., Eidelman, S., & Blanchar, J. C. (2015). Hierarchy, dominance, and deliberation: Egalitarian values require mental effort. *Personality and Social Psychology Bulletin*, *41*(9), 1207–1222. https://doi.org/10.1177/0146167215591961

Van Kleef, G. A., & De Dreu, C. K. (2002). Social value orientation and impression formation: A test of two competing hypotheses about information search in negotiation. *International Journal of Conflict Management*, *13*(1), 59–77. https://doi.org/10.1108/eb022867

Van Lange, P. A., Otten, W., De Bruin, E. M., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, *73*(4), 733–746. https://doi.org/10.1037//0022-3514.73.4.733

Van Vugt, M., Meertens, R. M., & Lange, P. A. M. (1995). Car versus public transportation? the role of social value orientations in a real-life social dilemma. *Journal of Applied Social Psychology*, *25*(3), 258–278. https://doi.org/10.1111/j.1559-1816.1995.tb01594.x

Wascher, C. A. F., & Bugnyar, T. (2013). Behavioral responses to inequity in reward distribution and working effort in crows and ravens. *PLOS ONE*, *8*(2), e56885. https://doi.org/10.1371/journal.pone.0056885

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049-1062. https://doi.org/10.1037/0022-3514.67.6.1049

Wiederkehr, V., Bonnot, V., Krauth-Gruber, S., & Darnon, C. (2015). Belief in school meritocracy as a system-justifying tool for low status students. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01053

Wilson, T. D., Centerbar, D. B., Kermer, D. A., & Gilbert, D. T. (2005). The pleasures of uncertainty: prolonging positive moods in ways people do not anticipate. *Journal of Personality and Social Psychology*, *88*(1), 5–21. https://doi.org/10.1037/0022-3514.88.1.5

Wu, M. S., Yan, X., Zhou, C., Chen, Y., Li, J., Zhu, Z., Shen, X., & Han, B. (2011). General belief in a just world and resilience: Evidence from a collectivistic culture. *European Journal of Personality*, *25*(6), 431–442. https://doi.org/10.1002/per.807

Zimmerman, J. L., & Reyna, C. (2013). The meaning and role of ideology in system justification and resistance for high-and low-status people. *Journal of Personality and Social Psychology*, *105*(1), 1-23. https://doi.org/10.1037/a0032967

## General discussion

I have presented three empirical studies which investigated how individuals approach and respond to social inequality. In Study 1, I found that systemically disadvantaged individuals are more likely to challenge the system when their decisions are made as part of a group whose decisions are relevant for one another. Feelings of identification moderated the extent to which participants believed that decisions made in groups were more efficacious than decisions made alone, but only when this efficacy could not be inferred from explicitly-stated intentions. These findings suggest that when communication channels are limited, feelings of group identity are particularly important for encouraging collective-minded decision-making, whereas group identity may be less important if decision-makers are able to coordinate their action through communication. In Study 2, using the same methods as Study 1, I explored how the incentive structure of group decision-making influenced disadvantaged individuals' propensity to challenge inequality. Under conditions of no communication (Study 2, Experiment 1), when it was possible to freeride by avoiding challenging the system and yet still benefiting from the challenge efforts of others, cooperation tended to decrease when class differences were not salient, suggesting that individuals succumbed to the temptation of freeriding in this situation. However, this temptation was not evident when class differences were made salient (i.e., when there was an Elite that benefited from the inequality). When participants were able to communicate intentions (Study 2, Experiment 2), a similar pattern was evident, but this time cooperation slightly increased when freeriding was possible when a merit-worthy Elite benefited from the inequality. This suggests that not only was the temptation to freeride neutralised by the salience of the inter-status differential, but when communication was possible, participants were more effective in coordinating their responses, despite the threat of freeriding. Overall, these results suggest that, whilst the threat of freeriding can undermine collective responses to inequality, if individuals are able to identify an outgroup antagonist, the threat of freeriding can be neutralised and, if they are able to communicate their intentions, the higher stakes environment may actually be more mobilising.

In Study 3, I found that several individual difference characteristics – most notably System Justification and Social Value Orientation (SVO) – are relevant for predicting whether individuals will seek information about social inequality. Moreover, the guiding influence of system justification and SVO was dependent on whether individuals would be advantaged or disadvantaged by the possible inequality. Individuals who were potentially disadvantaged were more likely to seek information about the equality of the system if they were strong justifiers of the status quo, rather than weak justifiers. In addition, potentially disadvantaged individuals who were strong justifiers were more likely to find out than potentially advantaged strong justifiers. The influence of SVO – the extent to which the outcomes of others are valued – had opposing effects depending on whether the individual was potentially advantaged or disadvantaged. “Prosocial” individuals were more likely to find out if they were potentially advantaged than if they were potentially disadvantaged, whereas “pro-self” individuals were more likely to find out if they were potentially disadvantaged than if they were potentially advantaged. These findings highlight the sensitivity of socially relevant motivations to social status. That is to say, for example, that a prosocial motivation is manifested differently depending on the relative social position of the person. Here, I found that for those for whom finding out would mean affirming their disadvantage, it appeared that the prosocial thing to do was to avoid finding out, whereas for those for whom finding out would mean affirming their advantage, the prosocial motivation was to find out. Likewise, for individuals believing strongly in the legitimacy of existing institutions and merit-based distribution of wealth, those who were potentially disadvantaged were more likely to find out than those who were potentially advantaged. This suggests the possibility that justifiers who were potentially disadvantaged were more likely to seek information and legitimise it, whereas the potentially advantaged may have been less likely to seek out information to begin with. These findings strongly suggest that dispositional motivations can be manifested in strikingly different ways, depending on the social context.

The results of these experiments combine to paint a picture of the person-specific factors that encourage equality information-seeking and the context-specific factors that motivate individuals to challenge inequality when it is known. Below, I consider in more detail how these findings contribute to an understanding of when equality information is sought and how individuals respond to inequality when it is known.

**Who seeks equality information?**

The findings of Study 3 provide a kind of "person specification" that details the characteristics of individuals who are more and less likely to seek equality information. A critical insight is that whether the individual stands to be advantaged versus disadvantaged by the inequality is highly important; though it appears that potentially advantaged and potentially disadvantaged individuals may not differ in terms of how likely they are to seek information about systemic inequality, I have presented evidence to suggest that the influences of certain individual difference measures on whether an individual seeks or avoids (in)equality information are sensitive to whether the individual would be advantaged or disadvantaged by the inequality.

If an individual is potentially advantaged by inequality, they are more likely to seek to confirm whether inequality exists if they do not by disposition tend to support and justify the societal status quo and ideals of merit-based class differences. Strikingly, the opposite is true if the individual is potentially disadvantaged by the inequality; that is, potentially disadvantaged individuals are more likely to seek to confirm whether inequality exists if they tend to justify the status quo and meritocratic ideals. This pattern suggests that potentially advantaged system justifiers are inclined to avoid resolving the ambiguity whereas potentially disadvantaged justifiers appear to take a more active role in resolving it. To the extent that these justification-relevant approaches are motivated and serve a palliative function (Kay & Jost, 2003), it might be suggested that those potentially advantaged by inequality are in a position to defer personal judgement and rely instead on the conviction that the situation is just, whereas for potentially disadvantaged justifiers – whose beneficial outcomes are more strongly tied to the question of equality – understanding where one stands in the system is more important.

A strong implication of this finding is that the advantaged in society may be more likely to be indifferent to issues of inequality because there is little incentive for them personally to discover it. This is important because the advantaged are often have better access to the resources that would help elicit social change. A logical deduction from this is that there would be a better outlook for resolving societal inequalities if those better off in society placed greater value in making themselves aware of social issues. One way of achieving this would be to make the outcomes of the disadvantaged more relevant to those of the advantaged. This might be achieved by more strongly instilling fairness norms and/or emphasising the importance of compassion for one another in the fabric of society (Singer & Steinbeis, 2009).

To an extent these processes can already be observed in society. Now more than ever, activists for social justice causes are not uniquely representatives of the group for whom social justice is sought. In broad terms, "allyship" refers to concern on the part of members of advantaged groups for disadvantaged groups (though see Louis et al., 2019 regarding their distinction between allyship and solidarity, and the different motivations for each). Perhaps one of the most striking recent examples of allyship was the global reaction to the killing of George Floyd on 25th May 2020, which saw hundreds of thousands of individuals from various ethnic backgrounds join in protests around the world. A budding literature on allyship explores the psychology behind this behaviour. The issue is complex and multifaceted, and motivations for allyship may not always be as altruistic, and its results may not be as desirable, as might be hoped (e.g., Radke et al., 2020; Russell & Bohan, 2016). For example, allies may act based on personal self-interest, derive a sense of entitlement from their participation, take attention away from a cause, or even reinforce the notion of a prepotent advantaged group (Droogendyk et al., 2016; Russell & Bohan, 2016). However, there is nonetheless evidence to support the notion that the allyship of advantaged individuals can also be attributed to a genuine moral motivation based on beliefs about right and wrong (Louis et al., 2019; Radke et al., 2020; van Zomeren et al., 2011). In both cases, the clear implication – consistent with my argument above – is that advantaged individuals may be more concerned about and engage with issues of social justice if they perceive self-relevant incentives to do so, be they incentives of self-interest or of moral principle.

Another valuable insight provided by Study 3 was that the influence of pro-sociality (as measured by social value orientation) also depended on the individual's relative social position. The finding suggests that prosocial individuals who are potentially advantaged by inequality are more likely to seek equality information, whereas prosocial individuals who are potentially disadvantaged by the inequality are less likely to do so. It further appears that this difference is explained by a 'prosocial' preference for maximising joint gains (i.e., ensuring the highest collective payoffs, regardless of how payoffs are distributed). Individuals who value joint gains will likely seek information if they are could be advantaged but avoid information if they are could be disadvantaged. In essence, these findings suggest that potentially advantaged prosocial, and potentially disadvantaged pro-self, individuals are likely to seek out equality information. Expressed in reverse terms, those least likely to discover inequality are potentially advantaged pro-self and potentially disadvantaged prosocial, individuals.

Though there is evidence to indicate that people are intuitively inequality-averse (Haruno & Frith, 2010), these results suggest that this aversion may be moderated by whether an individual stands to benefit from the inequality. Previous research suggests that when prosocial individuals are disadvantaged by a distribution, they are more prone to favour joint gain maximisation (i.e., they prefer to maximise gains even if this requires unequal distributions that disadvantage themselves and advantage others), whereas when prosocial individuals are advantaged by a distribution (i.e., they gain more than others), they are less prone to maximising joint gains (Mischkowski et al., 2019). The implication is that prosocial individuals tend to be more concerned about others' outcomes when those others are potentially disadvantaged. To the extent that seeking to ascertain (in)equality is considered "concern", I have presented consistent findings which indicate that prosocial individuals who value joint gains are more likely to seek to ascertain whether others have been paid fairly if they are potentially advantaged, whereas they are less likely to seek this information if they are potentially disadvantaged.

This presents a somewhat paradoxical implication whereby pro-sociality amongst the disadvantaged, since it is associated with lower tendencies to seek out equality information, may contribute to the maintenance of existing inequalities. This assertion has some empirical support. Hilbig et al. (2016) used a variation of the ultimatum game, termed the "Uncostly Retaliation Game", to investigate the relevance of the HEXACO personality traits Humility-Honesty and Agreeableness (Ashton et al., 2014) to resource distribution. In this paradigm, a proposer chooses either a fair or unfair division of wealth, and a responder can subsequently choose to retaliate to the proposition by reducing the proposer's wealth (but cannot change their own wealth), for example by reducing the proposer's wealth so that the distribution is equal. Hilbig et al. (2016) found that Agreeableness negatively predicted the extent to which responders retaliated by reducing the wealth of proposers. That is, responders who were more agreeable were more forgiving of proposers' unfair decisions. To the extent that responders in this scenario are considered "disadvantaged" – for example, because they had no control in the initial decision on how to distribute wealth in the scenario – the finding supports the notion that under some conditions disadvantaged individuals may be accepting of inequality. In addition, neuroimaging research suggests that experimentally manipulated social status influences the perception of fairness. Hu et al. (2016) found that low social status participants reject fewer unfair offers in the ultimatum game than high social status participants, and this was positively associated with activity in insular cortex, anterior cingulate cortex, and amygdala – areas implicated in affective responses to fairness and inequity aversion (e.g., Haruno & Frith, 2010; Sanfey et al., 2003). Together, these findings highlight how contextual factors such as social status moderate perceptions of fairness, and are consistent with my argument above that the differing effects for potentially advantaged and potentially disadvantaged individuals may be attributed to differing motivations.

**When do people challenge inequality when it is known?**

Studies 1 and 2 were concerned with decision-making in situations where it was known that systemic inequality was present. These studies paint a picture of when disadvantaged individuals are more and less likely to challenge inequality, and the subjective experiences involved. Firstly, disadvantaged individuals are more likely to challenge inequality when their decisions are interdependent (that is, when individual decisions are collectively relevant), and this appears to be associated with increased feelings of efficacy derived from a strong sense of group identity. This finding clearly accords with the large literature concerning social identity and collective action (e.g., Thomas et al., 2009; van Zomeren et al., 2008). In addition, it appears that the mobilising effect of group decision-making is compounded when there is an out-group to whom blame may be attributed. This aligns with the notion of *politicised* social identity as an important motivator of challenging the social system (Simon & Klandermans, 2001). In Study 2, I further showed that the incentive structure of group decision-making was important for group mobilisation. The results – in addition to reinforcing the notion that attributing blame to an outgroup is particularly mobilising – paradoxically suggest that when there is a strong incentive for group members not to cooperate, under the right conditions this “temptation” can instead galvanise group members to achieve *greater* cooperation.

Of particular note from these findings is the consistency with which they relate to notions of the importance of social identity and its politicisation. Early integrative models of collective action placed social identity as centrally important to collective mobilisation (e.g., van Zomeren et al., 2008), and the findings of Studies 1 and 2 add to research consistent with this notion (Thomas et al., 2009; van Zomeren et al., 2011). In addition, the results I have presented provide a strong case for the importance of politicisation of identity in social mobilisation. Though it is unlikely that the conditions in Studies 1 and 2 led to a full politicisation of social identity in the way outlined by Simon and Klandermans (2001), it is notable that certain aspects align with their outline for the process of politicisation. In particular, they suggest that social identity politicises when there is a collective feeling of group deprivation, which can be attributed to an outgroup, and which is acted on so that members of broader society are brought into the picture and take a stance. However, Klandermans (2014) argues that politicisation of identity is not all or nothing but instead develops in degrees, suggesting that not all the conditions for full pollicisation are necessary for these processes to be socially relevant. Thus, though action in Studies 1 and 2 never involved others beyond the immediate social group, the strong mobilisation occurring when there was a clear ingroup versus outgroup distinction can be understood to be an empirical demonstration of the process of politicisation of social identity.

In the current digital age, the potential for politicisation of identity has reached new heights. The power of “hashtag” memes exemplifies the principles of identity politicisation in social mobilisation, with many successful social campaigns bearing the hallmarks of the politicising process. For example, successful memes often involve a group of interest and an identifiable adversarial outgroup or outgroup member. In 2017 the #dresslikeawoman meme gained significant traction in defiance of then-President Trump's alleged comments that his female staff should "dress like women", and involved thousands of posts from women, often with a photograph showing their workplace attire, which of course was highly variable and typically not gender stereotypical. Here, the ingroup of concern was 'women', and the outgroup adversary was Donald Trump, who could also have been considered to epitomise and represent a category of men who endorse female oppression and sexism. The meme continues to be cited in 2021. The #stopfundinghate movement started in 2016 and campaigned against the way in which certain British newspapers used divisive storytelling and negative headlines about migrants to sell copies. Awareness of the campaign spread and was actively shared by the public, and further induced companies who normally advertised with the newspapers in question to forgo future collaboration. Here, there was perhaps evidence of full politicisation of identity insofar as there was evidence of an initial activist group campaigning against an adversarial outgroup (i.e., the newspapers in question), which led to the involvement of members of broader society, who themselves took a stance on the issue. Finally, as mentioned above, the #blacklivesmatter movement developed enormous momentum in the wake of the killing of George Floyd by a Minneapolis police officer. Though the disproportionate killing of Black men by police is unfortunately not a rare occurrence, George Floyd's killing entered the public consciousness rapidly (perhaps because there was considerable video footage documenting the event), leading to protests worldwide. Here again it is possible to discern a full politicisation of identity, starting from the identity of Black activists versus an oppressive policing system (i.e., an adversarial outgroup), to the involvement and action of broader society on a global scale. Indeed, the extent of the politicisation of this issue is likely one of the most far-reaching of its kind. Thus, though real-world social issues are highly complex – not least because of the historical factors that contribute to and perpetuate them – the findings I have presented in Studies 1 and 2 provide an empirical demonstration of how politicisation of social identity unfolds and contributes to social mobilisation.

The finding that, under conditions of partial politicisation of identity, the threat of freeriding may *increase* cooperation is highly consequential, because research typically documents how the temptation to freeride undermines collective action (e.g., Fischbacher et al., 2001). Moreover, many global issues are characterised as particularly challenging specifically because of the incentive for individuals to avoid cooperation and nonetheless benefit from the efforts of others (social distancing etiquette in the present pandemic is a topical example of this; Cato et al., 2020). The evidence I have presented decentralises the "freerider's temptation" and highlights the importance of considering other moderating factors of the collective problem. Specifically, it appears that emphasising group membership, identifying key antagonists, and ensuring effective communication amongst group members may mitigate the risk of individual temptation in collective endeavour involving social inequality. Indeed, under such conditions, awareness of the possibility of freeriding may even strengthen group members' conviction to cooperate.

Moral conviction has received considerable research interest as a motivator of collective action. Van Zomeren and colleagues (van Zomeren, 2019; van Zomeren et al., 2012) have provided strong accounts for including moral conviction as a core motivation for collective action. Van Zomeren et al. (2012) argued that moral convictions, when contravened, can motivate collective action by increasing feelings of social identity, which in turn influence feelings of group efficacy and anger. Interestingly, these researchers suggest that moral convictions are particularly relevant for action derived from *politicised* social identity (i.e., a collective identity that includes a motivation to act to elicit change; Simon & Klandermans, 2001) because politicised identities are more likely to encapsulate the moral stance on which convictions are based. That is, they argue there is likely to be a stronger "normative fit" (Turner et al., 1987) between moral convictions and the values contained within politicised identities, compared to non-politicised identities. For example, in concrete terms, an individual with a moral conviction that there should be racial equality in society is more likely to engage in collective action if they already possess an equality activist social identity as opposed to a non-activist identity. Other research supports this notion. For example, being primed with injunctive outrage norms (i.e., emphasising injustice) about an issue increases identification with an activist group ('Water for Life' supporters) as well as commitment to its cause (Thomas & McGarty, 2009). Likewise, the Normative Alignment Model proposed by Thomas et al. (2009) posits that sustained activism depends on an alignment of social identity, emotion relating to the issue (i.e., moral conviction), beliefs about efficacy, and action, and suggests that these factors must remain congruent to sustain action. Importantly, in contrast to van Zomeren et al.'s (2012) model, which suggests a causal ordering such that moral convictions contribute to the strength of identity (and subsequent feelings of efficacy and anger), the Normative Alignment Model does not assume a causal ordering and instead emphasises dynamic interrelations such that norms, convictions, identity, and efficacy beliefs can all be shaped by one another. This approach is consistent with social identity approaches that emphasise the possible context-sensitivity of social identity and its potential to evolve within the situation (Drury & Reicher, 2000; Tajfel & Turner, 1979). As a demonstration of this, Zaal et al. (2017; Study 1) showed that individuals holding politicised identities (feminists) identify less with others (women) if those others are perceived to have weaker moral convictions about the cause (gender equality). Taken together, this research highlights the important ways that moral convictions can be understood to motivate collective action. My findings in Study 2 add to this literature by demonstrating that collective action increases when identities are more politicised, and that – because this was the case despite an incentive to avoid collective action – this finding is likely attributable to increased moral convictions related to politicised identity.

A related concept that may help explain this finding is “moral courage”, which has been defined as “morally brave and risky behaviour … that expresses displeasure toward authorities or superiors without weighing potential disadvantages” (Kayser et al., 2010, p. 1137). Moral courage is demonstrated when an individual takes action in support of their moral concepts in situations with potentially negative consequences (Greitemeyer et al., 2007). Thus, in the current discussion, it may be considered as acting on moral convictions. Researchers typically find that morally courageous action is driven by indignation and anger (Greitemeyer et al., 2007; Halmburger et al., 2015; Kayser et al., 2010). For example, Halmburger et al. (2015) constructed a fake theft scenario where participants witnessed a confederate stealing a mobile phone. The extent to which participants felt and displayed anger in response to theft was found to predict the extent to which they intervened; that is, greater anger was associated with a stronger display of moral courage. These findings support the notion of the motivational power of moral convictions and demonstrate that feelings of anger may underly a willingness to take risky action to uphold a moral stance. Though in Study 2, I took no measure of anger, it is possible that the high likelihood of taking action, even when freeriding was possible, may be explained by a similar process of courage derived from moral convictions relating to equality.

## Limitations and Future Directions

It is important to acknowledge the limitations of the research I have presented. Firstly, in Studies 1 and 2 the samples were overwhelmingly comprised of young undergraduate students from Western, Educated, Industrialised, Rich, and Democratic (WEIRD) populations. As a result, generalisations from these data to a broader population are possibly invalid. Indeed, WEIRD populations are suggested to be particularly unrepresentative of the human population (Henrich et al., 2010). Although the sample in Study 3 was drawn from a broader, more representative UK demographic, lending more confidence to the generalisability of its findings to the UK population, it is not clear how applicable these findings are to individuals in non-Western and (especially) non-Democratic societies.

Indeed, recent research suggests that the successful activism in authoritarian contexts may be achieved through different means and processes than those understood to be successful in democratic contexts. For example, Vu (2019) recounts the case of a social movement organisation in Vietnam and highlighted that a possible route to effective outcomes is to present the movement as aligning with state agendas and to place itself within the regulatory bounds of the state. Doing so permits a degree of autonomy to pursue the movement's mission and avoids conflict with the state. This approach, which contrasts with forms of dissenting protest common in Democratic states, demonstrates that there is not a one-size-fits-all formula for activism. In addition, whilst education level is typically found to be a positive predictor of protest participation in Western states (Dalton et al., 2010), Ong and Han (2019) suggest that in the authoritarian context there is evidence that education is a *negative* predictor since, in a state with harsh punishments for dissent, the more educated and richer individuals stand to lose more by participating than their less-educated and poorer counterparts. It is therefore important to keep in mind that activism is likely to take different shapes and forms in societies with different forms of government. Consequently, the findings I have presented can only be interpreted in the context of Western, Democratic society. Future research could aim to address this limitation by considering similar scenarios involving collective responses to social inequality across different types of societal structures. It would be informative to identify what aspects identified as important to collective decision-making in the present studies carry over to non-Democratic contexts. For example, the strong responses against "legitimate elites" that I observed in Studies 1 and 2 may not hold in authoritarian contexts, and this would be important to establish to better understand what types of strategies will be more and less successful in a given society.

A second issue relates to the artificiality of the experimental contexts. The control afforded by the experimental manipulation of contextual factors carries the cost of being quite distinct to real-world contexts of inequality. Inequality was operationalised and experienced in the context of the experiment, rather than the context of society, and it is likely that the experience of inequality in the experiments was quite different to real-world instances of inequality. Thus, though I have made inferences – for example – about the motivating factors behind challenging inequality in the experiment, it cannot be taken for granted that these factors are the same in response to real-world inequality. On this note, it is also important to keep in mind that real inequality is contextualised in complex social systems with historical precedents, making the issue far more complex than the operationalised inequality in the present experiments. This is a critical limitation because the aim of this research is to provide insight into real-world social decision-making. One possibility for remedying the artificial spin of these experiments would be to explore similar concepts in field studies, where the phenomena and behaviours of interest truly have social and societal implications. One possibility would be to study disruptive social manifestations (for example, Extinction Rebellion demonstrations, where the risk of arrest and persecution by the public can be relatively high) and to quantify the costs and the benefits of these activities at the level of both individual and collective. This approach would marry the behavioural economic methodology pursued in this thesis with real-world instances of activism and could help align the theoretical with the practical.

The above considerations suggest that caution should be taken when trying to make claims about the broader population. They also highlight that a productive way forward would be to apply the focal concepts identified here to the study of real-world inequality and in so doing determine which findings receive support and which do not. Such complementary research, combining findings from experimental and field studies, would help build a more comprehensive and valid picture of the processes underlying the acknowledgement of inequality and action against it.

I chose to run Studies 1 and 2 in the same sessions to be able to collect data for two research questions in one sitting with the same participants. Though this was pragmatic, it came at the cost of the two foci of study potentially "contaminating" one another. That is, because the participants were the same across studies, decisions in Study 2 were likely influenced by decisions in Study 1, and vice versa. Though counterbalancing likely mitigated some order effects, the fact remains that decisions in any one context were made with the knowledge of how the context differed to others already completed within the experiment, creating an unavoidable frame of reference for participants. The independence of the first two studies must therefore be acknowledged as a caveat to the findings. Future research could aim to replicate these studies in isolation and see how the results compare to those presented in this thesis. Doing so may highlight which of the findings were potentially influenced by the interdependence within the study. For example, the unexpected finding that freeriding enhanced rather than undermined collective action in groups may be in part attributable to a comparison between the individual and freeriding scenarios, which was not an intended comparison but one that participants could nonetheless make due to the interdependent nature of the studies.

**Practical applications**

The findings of this thesis may have practical applications for activism campaigners and governing bodies. For example, the finding in Study 1 that processes of group identification are less relevant when it is possible to communicate highlights the potential for successful activism without needing to commit time and resources to developing group cohesion. This has strong implications in a digitally connected world, where the limits of social groups are no longer bound by geography. My results suggest that campaigns can make an effective first step through social media by creating a sense of efficacy through communication, and that this needn't be bound to a strong sense of collective identity in the first instance. Indeed, research suggests that social media activism can be a powerful mobiliser that helps strengthen group identities and subsequent action (Foster et al., 2021).

The insights from Studies 1 and 2 may also inform governing authorities on how to reduce citizens' motivation to challenge established social systems or practices. The results suggest that quashing the ability to effectively communicate and fostering a sense of division between individuals can be powerful tools of repression to reduce the likelihood of public dissent. The widespread censorship of media and social media in authoritarian states such as China evidence how effective such strategies can be. Even in Western societies, though explicit censorship of media outlets is less evident, a similar effect is created through the use of media (including social media) by using misinformation to undermine public trust, foster division, and manipulate the public discourse. Indeed, it is important to consider that information needn't be true for it to influence collective action. The attempts of ex US President Donald Trump's administration to overthrow the results of the 2020 US elections bear testament to the impact that consistent communication of misleading information can have on collective action in Democratic society.

A final implication of this research concerns individuals' willingness to engage with social issues in the first place. My results suggest that approximately 50 percent of individuals prefer to ignore inequality if they are given a choice, and that the propensity to do so is strongly linked to motivations to maintain a perception of the world as stable and predictable. Social issues are highly unlikely to be addressed and resolved if people are not willing to engage with them head on. Campaign strategists and governing authorities alike may therefore find it fruitful to force issues of social justice into the public discourse. However, the framing of this discourse is important because acknowledging flaws in society may involve acknowledging uncomfortable truths about one’s own place within it. For example, the term “White fragility” describes the discomfort that can be experienced by White individuals when they find themselves confronting issues of race and racism. This discomfort often leads to defensiveness and a reticence to engage in discussions concerning race, especially if they are required to recognise that they may be complicit in White supremacy (DiAngelo, 2018). A common consequence is to avoid discourse on race, or even deny there is an issue, appealing to justifications such as not “seeing” race or claiming that systemic racism is an issue of the past. This example illustrates the possibility that highlighting social justice issues in public discourse may be counter-productive if it is not done in a way that encourages individuals (including potential allies to those disadvantaged) to engage rather than defensively disengage. Thus, whilst the findings I have reported suggest a need to motivate individuals to seek social justice information, effective practical application of this suggestion by governing authorities or campaign strategists will need to carefully consider the most appropriate ways of doing so.

## Conclusion

The above limitations notwithstanding, my research provides non-trivial insights for guiding social action in the real world. The control provided by the experimental manipulation of contextual factors allow for precise inferences to made about their relative contribution to social decision-making. Firstly, echoing a broad literature, my findings strongly implicate feelings of group identity as a significant contributor to collective action, especially when other forms of affiliative coordination (e.g., explicit communication) are not possible. This is possibly most relevant to extremely large-scale endeavours where communication between all group constituents is unrealistic. Secondly, though action in social dilemmas falls victim to self-interest, the results I have presented suggest that this can be countered by identifying a clear antagonist that epitomises the opposition of collective efforts. Finally, the success of action for a social cause will depend strongly on whether a social cause is acknowledged in the first place. My findings demonstrate that the likelihood that people will seek information about a social issue depends on a combination of their dispositional characteristics and their social position with respect to the issue. For example, prosocial advantaged individuals might seek equality information whereas prosocial disadvantaged individuals might avoid it. It follows that encouraging disadvantaged individuals to learn about equality information might be best achieved by appealing more to self-focused concerns. Likewise, strong justifiers who are in an advantaged position appear to be less motivated to learn about equality information compared to disadvantaged justifiers. This implies that those who believe in the inherent justice of the way the world works and who are relatively unaffected by the inequality are less likely to be concerned with learning the inconvenient truth of their privilege. This is particularly pertinent to issues of social inequality because the "advantaged" groups are often the best positioned to influence how issues unfold. Taken together, these findings provide a foundation on which further research can build in continued efforts to reduce social inequality in society. It would be particularly productive to take the lessons learned from this research and explore whether similar conclusions can be drawn from real-world experiences of social inequality. For example, understanding whether there really exists a tendency for the disadvantaged in society to avoid information about inequality (e.g., through field studies) would be highly important for devising practical solutions to address the persistence of inequality. Likewise, understanding what factors promote allyship in society could help the fight against inequality by tackling the issue from the perspective of the advantaged in addition to that of the disadvantaged. The research I have presented in this thesis provides a map of potential starting points for such endeavours.

## References

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: a review of research and theory. *Personality and Social Psychology Review*, *18*(2), 139–152. https://doi.org/10.1177/1088868314523838

Cato, S., Iida, T., Ishida, K., Ito, A., McElwain, K. M., & Shoji, M. (2020). Social distancing as a public good under the COVID-19 pandemic. *Public Health*, *188,* 51–53. https://doi.org/10.1016/j.puhe.2020.08.005

Droogendyk, L., Wright, S. C., Lubensky, M., & Louis, W. R. (2016). Acting in solidarity: Cross-group contact between disadvantaged group members and advantaged group allies. *Journal of Social Issues, 72*(2), 315–334. https://doi.org/10.1111/josi.12168

Dalton, R., Sickle, A. V., & Weldon, S. (2010). The individual–institutional nexus of protest behaviour. *British Journal of Political Science, 40*(1), 51–73. https://doi.org/10.1017/S000712340999038X

DiAngelo, R. (2018). *White fragility: Why it's so hard for white people to talk about racism*. Beacon Press.

Drury, J., & Reicher, S. (2000). Collective action and psychological change: The emergence of new social identities. *British Journal of Social Psychology*, *39*(4), 579–604. https://doi.org/10.1348/014466600164642

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9

Foster, Mindi D., Adrianna Tassone, and Kimberly Matheson. Tweeting about sexism motivates further activism: A social identity perspective. *British Journal of Social Psychology* 60, no. 3 (2021): 741–64. https://doi.org/10.1111/bjso.12431.r

Greitemeyer, D. T., Osswald, S., Fischer, P., & Frey, D. (2007). Civil courage: Implicit theories, related concepts, and measurement. *The Journal of Positive Psychology*, *2*(2), 115–119. https://doi.org/10.1080/17439760701228789

Halmburger, A., Baumert, A., & Schmitt, M. (2015). Anger as driving factor of moral courage in comparison with guilt and global mood: A multimethod approach. *European Journal of Social Psychology*, *45*(1), 39–51. https://doi.org/10.1002/ejsp.2071

Haruno, M., & Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*(2), 160–161. https://doi.org/10.1038/nn.2468

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hilbig, B. E., Thielmann, I., Klein, S. A., & Henninger, F. (2016). The two faces of cooperation: On the unique role of HEXACO Agreeableness for forgiveness versus retaliation. *Journal of Research in Personality*, *64*, 69–78. https://doi.org/10.1016/j.jrp.2016.08.004

Hu, J., Blue, P. R., Yu, H., Gong, X., Xiang, Y., Jiang, C., & Zhou, X. (2016). Social status modulates the neural response to unfairness. *Social Cognitive and Affective Neuroscience*, *11*(1), 1–10. https://doi.org/10.1093/scan/nsv086

Kayser, D. N., Greitemeyer, T., Fischer, P., & Frey, D. (2010). Why mood affects help giving, but not moral courage: Comparing two types of prosocial behaviour. *European Journal of Social Psychology*, *40*(7), 1136–1157. https://doi.org/10.1002/ejsp.717

Klandermans, P. G. (2014). Identity politics and politicized identities: Identity processes and the dynamics of protest. *Political Psychology*, *35*(1), 1–22. https://doi.org/10.1111/pops.12167

Louis, W. R., Thomas, E., Chapman, C. M., Achia, T., Wibisono, S., Mirnajafi, Z., & Droogendyk, L. (2019). Emerging research on intergroup prosociality: Group members' charitable giving, positive contact, allyship, and solidarity with others. *Social and Personality Psychology Compass*, *13*(3), e12436. https://doi.org/10.1111/spc3.12436

Mischkowski, D., Thielmann, I., & Glöckner, A. (2019). Minimizing inequality versus maximizing joint gains: On the relation between personality traits and different prosocial motivations. *Journal of Research in Personality*, *81*, 153–157. https://doi.org/10.1016/j.jrp.2019.06.001

Ong, L. H., & Han, D. (2019). What drives people to protest in an authoritarian country? resources and rewards vs risks of protests in urban and rural china. *Political Studies*, *67*(1), 224–248. https://doi.org/10.1177/0032321718763558

Radke, H. R. M., Kutlaca, M., Siem, B., Wright, S. C., & Becker, J. C. (2020). Beyond allyship: Motivations for advantaged group members to engage in action for disadvantaged groups. *Personality and Social Psychology Review*, *24*(4), 291–315. https://doi.org/10.1177/1088868320918698

Russell, G. M., & Bohan, J. S. (2016). Institutional allyship for LGBT equality: Underlying processes and potentials for change. *Journal of Social Issues*, *72*(2), 335–354. https://doi.org/10.1111/josi.12169

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758.

Simon, B., & Klandermans, B. (2001). Politicized collective identity: A social psychological analysis. *American Psychologist*, *56*, 319–331. https://doi.org/10.1037/0003-066X.56.4.319

Singer, T., & Steinbeis, N. (2009). Differential roles of fairness- and compassion-based motivations for cooperation, defection, and punishment. *Annals of the New York*

*Academy of Sciences*, *1167*(1), 41–50. https://doi.org/10.1111/j.1749-6632.2009.04733.x

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of inter-group relations* (pp. 33–47). Brooks/Cole.

Thomas, E. F., & McGarty, C. A. (2009). The role of efficacy and moral outrage norms in creating the potential for international development activism through group-based interaction. *British Journal of Social Psychology*, *48*(1), 115–134. https://doi.org/10.1348/014466608X313774

Thomas, E. F., McGarty, C., & Mavor, K. I. (2009). Aligning identities, emotions, and beliefs to create commitment to sustainable social and political action. *Personality and Social Psychology Review*, *13*(3), 194–218. https://doi.org/10.1177/1088868309341563

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

van Zomeren, M. (2019). Toward a cultural psychology of collective action: Just how “core” are the core motivations for collective action? *Journal of Pacific Rim Psychology*, *13*, e15. https://doi.org/10.1017/prp.2019.7

van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin*, *134*(4), 504–535. https://doi.org/10.1037/0033-2909.134.4.504

van Zomeren, M., Postmes, T., & Spears, R. (2012). On conviction’s collective consequences: Integrating moral conviction with the social identity model of collective action. *British Journal of Social Psychology*, *51*(1), 52–71. https://doi.org/10.1111/j.2044-8309.2010.02000.x

van Zomeren, M., Postmes, T., Spears, R., & Bettache, K. (2011). Can moral convictions motivate the advantaged to challenge social inequality?: Extending the social identity model of collective action. *Group Processes & Intergroup Relations*, *14*(5), 735–753. https://doi.org/10.1177/1368430210395637

Vu, A. N. (2019). NGO-led activism under authoritarian rule of Vietnam: Between cooperation and contestation. Community Development, 50(4), 422–439. https://doi.org/10.1080/15575330.2019.1642925

Zaal, M. P., Saab, R., O’Brien, K., Jeffries, C., Barreto, M., & van Laar, C. (2017). You’re either with us or against us! Moral conviction determines how the politicized distinguish friend from foe. *Group Processes & Intergroup Relations*, *20*(4), 519–539. https://doi.org/10.1177/1368430215615682

# Appendices

## Appendix A: Individual difference measurement scales used in Study 3

### System Justification – General

From Kay, A.C., & Jost, J.T. (2003). Complementary justice: Effects of "poor but happy" and "poor but honest" stereotype exemplars on system justification and implicit activation of the justice motive. *Journal of Personality and Social Psychology, 85(5),* 823-837.

Rate agreement with each of the below statements from 1 (Strongly disagree) to 9 (Strongly agree):

1) In general, you find society to be fair.

2) In general, the political system of my country operates as it should.

3) Society needs to be radically restructured. *

4) My country is the best country in the world to live in.

5) Most policies serve the greater good.

6) Everyone has a fair shot at wealth and happiness.

7) Our society is getting worse every year. *

8) Society is set up so that people usually get what they deserve.

Scoring:

Higher numbers indicate greater system justification. * indicates reverse-scored items.

**System Justification – Economic**

From Jost, J. T., & Thompson, E. P. (2000). Group-based dominance and opposition to equality as independent predictors of self-esteem, ethnocentrism, and social policy attitudes among African Americans and European Americans. *Journal of Experimental Social Psychology*, *36*(3), 209-232.

Rate agreement with each of the below statements from 1 (Strongly disagree) to 9 (Strongly agree):

1) If people work hard, they almost always get what they want.

2) The existence of widespread economic differences does not mean that they are inevitable.

3) Laws of nature are responsible for differences in wealth in society.

4) There are many reasons to think that the economic system is unfair. *

5) It is virtually impossible to eliminate poverty.

6) Poor people are not essentially different from rich people. *

7) Most people who don't get ahead in our society should not blame the system; they have only themselves to blame.

8) Equal distribution of resources is a possibility for our society. *

9) Social class differences reflect differences in the natural order of things.

10) Economic differences in the society reflect an illegitimate distribution of resources. *

11) There will always be poor people, because there will never be enough jobs for everybody.

12) Economic positions are legitimate reflections of people's achievements.

13) If people wanted to change the economic system to make things equal, they could. *

14) Equal distribution of resources is unnatural.

15) It is unfair to have an economic system which produces extreme wealth and extreme poverty at the same time.

16) There is no point in trying to make incomes more equal.

17) There are no inherent differences between rich and poor; it is purely a matter of the circumstances into which you are born.

Scoring:

Higher numbers indicate greater system justification. * indicates reverse-scored items.

## Life Orientation Test Revised

From Carver, C. S., Scheier, M. F., & Segerstrom, S. C.(2010). Optimism. *Clinical Psychology Review, 30,* 879-889.

and

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A re-evaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67,* 1063-1078.

Please be as honest and accurate as you can throughout. Try not to let your response to one statement influence your responses to other statements. There are no "correct" or "incorrect" answers. Answer according to your own feelings, rather than how you think "most people" would answer.

A = I agree a lot

B = I agree a little

C = I neither agree nor disagree

D = I disagree a little

E = I disagree a lot

1. In uncertain times, I usually expect the best.

[2. It's easy for me to relax.]

3. If something can go wrong for me, it will.

4. I'm always optimistic about my future.

[5. I enjoy my friends a lot.]

[6. It's important for me to keep busy.]

7. I hardly ever expect things to go my way.

[8. I don't get upset too easily.]

9. I rarely count on good things happening to me.

10. Overall, I expect more good things to happen to me than bad.

----------------------------------------------------------------------

Note:

Items 2, 5, 6, and 8 are fillers. Responses to "scored" items are to be coded so that high values imply optimism. Researchers who are interested in testing the potential difference between affirmation of optimism and disaffirmation of pessimism should compute separate subtotals of the relevant items.

## Intolerance of Uncertainty

From Carleton, R. N., Norton, M. A., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders, 21*, 105-117. doi: S0887-6185(06)00051-X [pii]10.1016/j.janxdis.2006.03.014

Rate the extent to which each of the below statements is characteristic of you on a scale from 1 (Not at all characteristic of me) to 5 (Entirely characteristic of me).

1. Unforeseen events upset me greatly.
2. It frustrates me not having all the information I need.
3. Uncertainty keeps me from living a full life.
4. One should always look ahead so as to avoid surprises.
5. A small unforeseen event can spoil everything, even with the best of planning.
6. When it’s time to act, uncertainty paralyses me.
7. When I am uncertain I can’t function very well.
8. I always want to know what the future has in store for me.
9. I can’t stand being taken by surprise.
10. The smallest doubt can stop me from acting.

11 I should be able to organize everything in advance.

12 I must get away from all uncertain situations.

Prospective anxiety subscale: Sum of items 1, 2, 4, 5, 8, 9, 11

Inhibitory anxiety subscale: Sum of items 3, 6, 7, 10, 12

Total IU score: Sum of all items

## Social Value Orientation

From Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring Social Value Orientation (SVO). *Judgment and Decision Making, 6,* 771-781.

(The below images were obtained from http://ryanomurphy.com/styled-2/downloads/index.html)

### Instructions

In this task you have been randomly paired with another person, whom we will refer to as the **other**. This other person is someone you do not know and will remain mutually anonymous. All of your choices are completely confidential. You will be making a series of decisions about allocating resources between you and this other person. For each of the following questions, please indicate the distribution you prefer most by **marking the respective position along the midline**. You can only make one mark for each question.

Your decisions will yield money for both yourself and the other person. In the example below, a person has chosen to distribute money so that he/she receives 50 dollars, while the anonymous other person receives 40 dollars.

There are no right or wrong answers, this is all about personal preferences. After you have made your decision, **write the resulting distribution of money on the spaces on the right**. As you can see, your choices will influence both the amount of money you receive as well as the amount of money the other receives.

Example:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| You receive | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | You 50 |
| Other receives | 80 | 70 | 60 | 50 | 40 | 30 | 20 | 10 | 0 | Other 40 |

1

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| You receive | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | You ____ |
| Other receives | 85 | 76 | 68 | 59 | 50 | 41 | 33 | 24 | 15 | Other ____ |

2

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| You receive | 85 | 87 | 89 | 91 | 93 | 94 | 96 | 98 | 100 | You ____ |
| Other receives | 15 | 19 | 24 | 28 | 33 | 37 | 41 | 46 | 50 | Other ____ |

3

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| You receive | 50 | 54 | 59 | 63 | 68 | 72 | 76 | 81 | 85 | You ____ |
| Other receives | 100 | 98 | 96 | 94 | 93 | 91 | 89 | 87 | 85 | Other ____ |

4

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| You receive | 50 | 54 | 59 | 63 | 68 | 72 | 76 | 81 | 85 | You ____ |
| Other receives | 100 | 89 | 79 | 68 | 58 | 47 | 36 | 26 | 15 | Other ____ |

5

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| You receive | 100 | 94 | 88 | 81 | 75 | 69 | 63 | 56 | 50 | You ____ |
| Other receives | 50 | 56 | 63 | 69 | 75 | 81 | 88 | 94 | 100 | Other ____ |

6

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| You receive | 100 | 98 | 96 | 94 | 93 | 91 | 89 | 87 | 85 | You ____ |
| Other receives | 50 | 54 | 59 | 63 | 68 | 72 | 76 | 81 | 85 | Other ____ |

7

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| You receive | 100 | 96 | 93 | 89 | 85 | 81 | 78 | 74 | 70 |
| Other receives | 50 | 56 | 63 | 69 | 75 | 81 | 88 | 94 | 100 |

You ________
Other ________

8

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| You receive | 90 | 91 | 93 | 94 | 95 | 96 | 98 | 99 | 100 |
| Other receives | 100 | 99 | 98 | 96 | 95 | 94 | 93 | 91 | 90 |

You ________
Other ________

9

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| You receive | 100 | 94 | 88 | 81 | 75 | 69 | 63 | 56 | 50 |
| Other receives | 70 | 74 | 78 | 81 | 85 | 89 | 93 | 96 | 100 |

You ________
Other ________

10

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| You receive | 100 | 99 | 98 | 96 | 95 | 94 | 93 | 91 | 90 |
| Other receives | 70 | 74 | 78 | 81 | 85 | 89 | 93 | 96 | 100 |

You ________
Other ________

11

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| You receive | 70 | 74 | 78 | 81 | 85 | 89 | 93 | 96 | 100 |
| Other receives | 100 | 96 | 93 | 89 | 85 | 81 | 78 | 74 | 70 |

You ________
Other ________

12

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| You receive | 50 | 56 | 63 | 69 | 75 | 81 | 88 | 94 | 100 |
| Other receives | 100 | 99 | 98 | 96 | 95 | 94 | 93 | 91 | 90 |

You ________
Other ________

13

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| You receive | 50 | 56 | 63 | 69 | 75 | 81 | 88 | 94 | 100 |
| Other receives | 100 | 94 | 88 | 81 | 75 | 69 | 63 | 56 | 50 |

You ________
Other ________

14

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| You receive | 100 | 96 | 93 | 89 | 85 | 81 | 78 | 74 | 70 |
| Other receives | 90 | 91 | 93 | 94 | 95 | 96 | 98 | 99 | 100 |

You ________
Other ________

15

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| You receive | 90 | 91 | 93 | 94 | 95 | 96 | 98 | 99 | 100 |
| Other receives | 100 | 94 | 88 | 81 | 75 | 69 | 63 | 56 | 50 |

You ________
Other ________

## Belief in a Just World

From Dalbert, C. (1999). The World is More Just for Me than Generally: About the Personal Belief in a Just World Scale's Validity. *Social Justice Research, 12(2),* 79–98. https://doi.org/10.1023/A:1022091609047

Please rate your agreement from 1 (Strongly disagree) to 6 (Strongly agree) on the items below.

| Strongly disagree | Disagree | Slightly disagree | Slightly agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

### Belief in a just world – Personal

1) I believe that, by and large, I deserve what happens to me.

2) I am usually treated fairly.

3) I believe that I usually get what I deserve.

4) Overall, events in my life are just.

5) In my life injustice is the exception rather than the rule

6) I believe that most of the things that happen in my life are fair.

7) I think that important decisions that are made concerning me are usually just.

### Belief in a just world – General

1) I think basically the world is a just place.

2) I believe that, by and large, people get what they deserve.

3) I am confident that justice always prevails over injustice.

4) I am convinced that in the long run people will be compensated for injustices.

5) I firmly believe that injustices in all areas of life (e.g., professional, family, politics) are the exception rather than the rule.

6) I think people try to be fair when making important decisions.

## Appendix B: Participant comments from which qualitative themes were derived in Study 3

| Status | Decision | Comment | Endorsement of reason (max 100) | Identified theme |
|---|---|---|---|---|
| Advantaged | Don't find out | It was stated very clearly that finding out could not change anything so I saw no purpose in finding out. | 100 | Unchangeability |
| Advantaged | Don't find out | i got that ammount of money what i did except in the beginning.it was told this survey pays 8 and i got it. | | Expectations |
| Advantaged | Don't find out | I don’t know the other person so I’d rather not know then I essentially have no feelings either way. Good luck to he/she I say. | | Personal irrelevance |
| Advantaged | Don't find out | Mainly because knowing or not there was nothing I could change about it | 100 | Unchangeability |
| Advantaged | Don't find out | I live by the whole 'out of sight, out of mind' motto, so if I don't technically know if my partner has been paid more or less or equal to me it should not matter, its how I will use that money and how they will use that money for themselves. | 100 | Avoid affective reaction |
| Advantaged | Don't find out | It doesn't matter what the other person was paid to me. I do not know the other person. If they were paid less then I do not need to know that as I would just feel guilty. | | Personal irrelevance |
| Advantaged | Don't find out | I didn't care enough to find out because I was satisfied enough with my pay | 100 | Payment satisfaction |
| Advantaged | Don't find out | I couldn't alter the decision anyway | 95 | Unchangeability |
| Advantaged | Don't find out | I genuinely am not bothered about how much the other person is paid for completing the study. What people earn is their business. I'm not sure how I would go about giving this reason a rating between 1 and 100. | | Indifference |

| | | | | |
|---|---|---|---|---|
| Advantaged | Don't find out | household income shouldn't matter if you are doing the same work | 100 | Other |
| Advantaged | Don't find out | I did not want to feel guilty or guilt tripped. | | Avoid affective reaction |
| Advantaged | Don't find out | There is no benefit or sense in stressing over a situation I cannot change - the best I can do is accept the situation or avoid it entirely. As I had the option to avoid, that is what I chose to do. | 100 | Unchangeability |
| Advantaged | Don't find out | People hold their own opinions and interests. I do not personally know the person that I have been paired with, so it does not intrigue me. | | Personal irrelevance |
| Advantaged | Don't find out | I didn't want to find out because whatever I get paid is good for me and my knowing with not change the amount that each of us get paid. | | Payment satisfaction |
| Advantaged | Don't find out | My reason to not find out is that the study is completely volenterily and therefore each participant accepts the reward offered. | 100 | Expectations |
| Advantaged | Don't find out | I did not really want to know how much my partner was been made. | | Indifference |
| Advantaged | Don't find out | I wouldn't like to know they were being paid less and learn there was nothing I could do to change that | 100 | Avoid affective reaction |
| Advantaged | Don't find out | As I couldn't change the amount the other person was being paid, it would just bother me to know if they had received less, i wouldn't be concerned if they received more just less as I would feel terrible guilt and its not something I can change. | | Unchangeability |
| Advantaged | Don't find out | as long as they signed up to it and were happy to participate - i dont see any disconcern in me knowing or not knowing. | | Expectations |
| Advantaged | Find out | I don't like the idea of two people being paid different amounts for doing the same thing. Makes me really uncomfortable. | | Equality concerns |

| | | | | |
|---|---|---|---|---|
| Advantaged | Find out | Wanted to satisfy curiosity | 80 | Curiosity |
| | | Was interested to find more information on why we were possibly paid differently | 90 | |
| Advantaged | Find out | i was just curious to see if my partner got the same, more or less than me. I'm not that bothered if i don't find out i was just curious. id say I'm 50/100 curious to know what my partner got | 50 | Curiosity |
| Advantaged | Find out | I just like people to be treated the same | 90 | Equality concerns |
| Advantaged | Find out | I was curious | 1 | Curiosity |
| Advantaged | Find out | General curiosity | 100 | Curiosity |
| Advantaged | Find out | I'm interested in the equity of the experiment. | 100 | Curiosity |
| Advantaged | Find out | £8.00 feels like a lot for the amount of work/ time I put in. I might donate some of my fee to charity if the partner is paid what I feel to be unfair as in not equal enough. | 99 | Equality concerns |
| Advantaged | Find out | Curiosity | 70 | Curiosity |
| Advantaged | Find out | It would be nice to know everyone was paid the same for the same work. | | Equality concerns |
| Advantaged | Find out | I wanted to find out because I'm hoping we were paid the same | 100 | Equality concerns |
| Advantaged | Find out | Curiosity got the better of me. If I had not been provided with the choice, I probably would not have cared about the outcome. | | Curiosity |
| Advantaged | Find out | I don't like not knowing things. I am just curious | | Intolerance of ignorance |
| Advantaged | Find out | I have issues with unfairness in this situation | 100 | Equality concerns |
| | | I see unfairness in my working life and can at least redress an in balance here because i have the capability to do it | 100 | |
| Advantaged | Find out | High regard for transparency | 80 | Transparency |
| | | Curiosity | 60 | |
| Advantaged | Find out | curiosity really | | Curiosity |
| Advantaged | Find out | you gave me the option | 100 | Other |
| Advantaged | Find out | 1.May not know them, but they completed the same work, so there shouldn't be anything | | Equality concerns |

| | | | | |
|---|---|---|---|---|
| | | stopping us being paid equally.<br><br>2.It's a reward, and not an expectation that either one of us should receive more.<br><br>3. I believe that people on a whole have the same opinion, so there is little risk.<br><br>4. I would never want the other person to feel sad/bad about getting less. | | |
| Advantaged | Find out | "Being nosey" | 70 | Curiosity |
| Disadvantaged | Don't find out | I'd be upset to know that anybody even higher or lower income was paid differently. It wouldn't be fair either way. | | Equality concerns |
| Disadvantaged | Don't find out | was happy not to know..and to just be content with what i was paid | | Indifference |
| Disadvantaged | Don't find out | I did not want to find out incase they got more, I don't want to deliberately feel bad. If I got more, I'd also feel bad, so not knowing was a good choice! | | Equality concerns |
| Disadvantaged | Don't find out | if it won't affect the result i don't need to know | 90 | Unchangeability |
| Disadvantaged | Don't find out | I think there's a certan amount of reassurance that come with ignorance if you can tolerate deliberatley wanting to be ignorant. Over the years, this stance/view has become something of a strategy for me in the sense that I know for sure the world is extremely unequal and yet there is nothing I alone can do about it. Therefore, it is sometimes easier and perhaps somewhat wiser to simply be ignorant of inequality rather than have it trigger anger or anxiety. | 100 | Avoid affective reaction |
| Disadvantaged | Don't find out | id rarther not no if their was an enequility | | Equality concerns |
| Disadvantaged | Don't find out | Finding out would not change my pay, and I felt | 99 | Unchangeability |

| | | | | |
|---|---|---|---|---|
| | | good on the amount I got paid so I didnâ€™t want to change that. | | |
| Disadvantaged | Don't find out | i was happy with the pay but if i turned out someone was paid double for the same work i would not be happy and feel mistreated.<br><br>i chose to be happy with what i have . | | Payment satisfaction |
| Disadvantaged | Don't find out | What others earn does not influence me as long as I feel fairly compensated for my effort | 100 | Personal irrelevance |
| Disadvantaged | Don't find out | I would rather not know to reduce jealously or spite | 60 | Avoid affective reaction |
| | | I would rather not know as there is no need for me to know | 70 | |
| Disadvantaged | Don't find out | I knew how much the financial reward that I received was before commencing the study, it bothers me not that the other earned more or less than me as that was their choice not mine | 100 | Expectations |
| Disadvantaged | Don't find out | I am satisfied with the money I am being paid and comparing it to another person right now would only make me less satisfied for something that is satisfying to me - it'd be a waste of time and energy. | 100 | Payment satisfaction |
| Disadvantaged | Don't find out | it seems relative to the work i did. I am happy with being rewarded with £8. more would be nice but i fear i would not understand why someone else would be paid more without reasoning. | | Payment satisfaction |
| Disadvantaged | Don't find out | I am paid regardless if the other person receives more or not | | Payment satisfaction |
| Disadvantaged | Don't find out | Not knowing removes any chance of future disappointment | 100 | Avoid affective reaction |
| Disadvantaged | Don't find out | I try not to worry about what other people are earning. I am not a jealous person. I am happy with the amount I will receive and will not | 100 | Personal irrelevance |

| | | | | |
|---|---|---|---|---|
| | | think about the other person. | | |
| Disadvantaged | Don't find out | i still get paid | 100 | Payment satisfaction |
| | | i cant change a thing | 100 | |
| Disadvantaged | Don't find out | i did not need to find out, i know what i am being paid and that is all that bothers me. i may have felt that it was unfair if the other person was paid more than me. | | Personal irrelevance |
| Disadvantaged | Don't find out | Won't change outcome so no point knowing and feeling frustrated or annoyed by it | 100 | Unchangeability |
| Disadvantaged | Don't find out | It makes no difference to my pay if i know or not so why bother | 90 | Unchangeability |
| Disadvantaged | Don't find out | I am still being paid £5 which is a good amount | 90 | Payment satisfaction |
| Disadvantaged | Don't find out | I simply don't care what anyone else received, it was expressly stated that my reward wouldn't be changed and as such it's very easy to distance myself from any concerns about what others may or may not have | | Indifference |
| Disadvantaged | Don't find out | I don’t think it would be very fair for someone to get paid more or less than me for doing the same work so I’d rather just not know the outcome so I don’t think about it and get annoyed | 70 | Equality concerns |
| Disadvantaged | Don't find out | would rather not cause any potential negative feelings | | Avoid affective reaction |
| Disadvantaged | Don't find out | Both should be paid for the same amount of work put into the study. | | Equality concerns |
| Disadvantaged | Find out | Simply out of curiosity! | | Curiosity |
| Disadvantaged | Find out | I'm just curious and want to know | | Curiosity |
| Disadvantaged | Find out | Want to know we are all equal for same labour not he favoured more for no reason of substance | | Equality concerns |
| Disadvantaged | Find out | No real reason was just interested to find out,it was a great study and i wish more studies were as good as yours. | | Curiosity |
| Disadvantaged | Find out | My primary reason to find out was to find out if the other participant got fair pay. This was my main concern because I | 80 | Fair payment |

| | | | | |
|---|---|---|---|---|
| | | already believe my pay to be a fair rate. | | |
| | | I did question whether that was information that I really needed to know considering I couldn't do anything about it should the other participant's pay not be fair. As such my secondary reason was curiosity | 20 | |
| Disadvantaged | Find out | There were no other reasons behind wanting to find out. | | Curiosity |
| Disadvantaged | Find out | Just mere curiosity | | Curiosity |
| Disadvantaged | Find out | Out of sheer curiosity | | Curiosity |
| Disadvantaged | Find out | To see if the study will follow the economic structure of society - some being more rewarded than others. | | Equality concerns |
| Disadvantaged | Find out | Wanting to find out how unfairly balanced the payments were | 70 | Equality concerns |
| Disadvantaged | Find out | to see if higher income households got paid more here, like they do get paid more in their real day job. | 80 | Equality concerns |
| Disadvantaged | Find out | it was more for fun to find out, not a competition or anything, just wanted to find out if we got the same or i got more or less, wanted to see how lucky i was. | | Curiosity |
| Disadvantaged | Find out | pure curiosity | 95 | Curiosity |
| Disadvantaged | Find out | i am wondering what difference in pay is | 50 | Curiosity |
| Disadvantaged | Find out | curiosity | 89 | Curiosity |
| Disadvantaged | Find out | Curiosity | 100 | Curiosity |
| Disadvantaged | Find out | just plain old curiosity, no negative feelings if i was paid less, as i was fully aware of the reward when starting the experiment. | | Curiosity |
| Disadvantaged | Find out | Out of curiosity | 70 | Curiosity |
| | | If it was done fairly | 40 | |
| Disadvantaged | Find out | I am just curious | 100 | Curiosity |
| Disadvantaged | Find out | Curiosity | | Curiosity |
| Disadvantaged | Find out | I think it just comes down to the blatancy of being told "you have lower income and so you might be paid less" made it so that I had to know whether it would really | | Curiosity |

| | | | | |
|---|---|---|---|---|
| | | happen. Positive discrimination is publicly so much more common nowadays that openly negative discrimination took me aback, even if it happens all the time discretely. | | |
| Disadvantaged | Find out | curious | | Curiosity |
| Disadvantaged | Find out | Genuine Curiosity, I had not considered until it was mentioned that myself and my partner would have been paid different amounts, I had made an assumption it would be equal. | 80 | Curiosity |
| Disadvantaged | Find out | curiosity | 90 | Curiosity |
| | | why not? | 80 | |
| | | it does not harm me | 90 | |
| | | i can not change it so why not know | 90 | |