“I did it for you!”

Examining the Effect of Norms on Moral Self-Image and Behavioural Spillovers

Fatima Koaik

PhD Thesis, Department of Psychology and Behavioural Science

October 2024

A thesis submitted to the Department of Psychological and Behavioural Science of the London School of Economics and Political Science, for the degree of Doctor of Philosophy, London, 2024

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work.

I confirm that the papers were jointly co-authored with Prof. Paul Dolan and Dr. Dario Krpan.

Acknowledgements

As a mother of two and a full-time employee, completing this thesis was made possible only through the invaluable help, support, and love of the people in my life. I am grateful for you and I dedicate it to each one of you…

Dedicated to my beloved sons, Adam and Jude. **I did it for you!** You have consistently been my primary source of motivation and inspiration and have remained my utmost priority throughout this journey. You were 6 and 2 years old when I started the program, we spent weekends studying together as a team. I hope that this thesis will serve as a source of inspiration for you both, encouraging you to pursue your dreams with perseverance and patience, always striving for improvement. Constant learning and hard work lead to the most satisfaction and sense of purpose in life. Finally, curiosity to learn more about topics that interest me led me to this path. May you never cease to pursue knowledge and experience curiosity!

Dedicated to my husband Ali, whose belief in me, support, and encouragement helped me to embark on and pursue this doctoral journey. He maintained that the seemingly impossible was indeed within reach. Balancing a full-time job and caring for two children under the age of six, I often felt that attaining this degree was a distant dream. Yet, through his continuous support, we were able to transform what seemed impossible, actually conceivable, and attainable. Thank you for always pushing me to become a better version of myself and for taking on more to help me achieve this. I love you!

Dedicated to my dad, my mom, and my siblings. Thank you for always being there to support me. Even from afar, your love and encouragement were instrumental throughout my journey. I learned from my dad's resilience, generosity and curiosity, and from my mom's patience, commitment, and hard work. Mona, we always dreamed big, you have always been my inspiration and idol, this one is for you! Imane, Sarah and Mohamad you are my go-to and the best support and source of encouragement.

Dedicated to my mentors/ supportive managers, and invaluable work colleagues and friends. Throughout the duration of my PhD journey, I navigated the challenging task of balancing full-time employment alongside the demands of academic pursuit. Your support, encouragement, and consistent positive reinforcement have been indispensable throughout this journey. I am also blessed with colleagues and team members who became family, to each one of you thank you!

As I submit this thesis, I am so lucky to be working with the most supporting and caring leader, Dima, who is a source of inspiration and strength, always pushing me to the next level and boosting my confidence and morale, simply bringing the best version of myself. Dima, big thank you and so grateful for you.

I would like to express my heartfelt gratitude to my supervisors, Professor Paul Dolan and Dr. Dario Krpan for their guidance and profound insights throughout my doctoral journey. Their mentorship has been instrumental in shaping my research, refining my ideas, and navigating the challenges of academic inquiry.

Dedicated to every mother and woman out there, you can do it, just find your support system!
Finally, I often get asked why I am doing a PhD. Simply, it brings me happiness and a big sense of purpose!

Table of Content

Tables, Figures and Appended Documents

Figures

Tables

Appended Documents

Document	Pages	Description
1	192	Appendix A: Moral Self Image Scale
2	195	Appendix B: Social Expectations in Study 1 and 2
3	199	Appendix C: Donation Task (Behaviour 2) in Study 2
4	199	Appendix D: Social Expectations in Study 3
5	201	Appendix E: Behaviour 1 in Study 3
6	203	Appendix F: Carbon Emission Task (Behaviour 2) in Study 3 and 4
7	204	Appendix G: Pilot Study 4
8	206	Appendix H: Behaviour 1 in Study 4

I. Introduction to the Thesis

"I did it for you!"… When conformity with the expectation increases moral licensing effect.

Imagine you move to a new residence where most of your neighbours recycle by sorting their domestic waste and expect their community to recycle. You then start recycling because you want to conform with your neighbours' actions. You start feeling good about yourself. This positive feeling about your actions and your moral values justifies your use of plastic bags for your groceries. After all, you did it to conform with your neighbours… you did it for them!

We live in a world where collective action problems are alarming as they relate to complex social issues, environmental and climate change challenges, and public health, among others, and addressing them requires as much structural as behavioural change at the individual level. Despite sharing a common interest, people often face barriers to taking the correct actions and positive behaviour. With the rise of behavioural interventions, a significant concern is the sustainability of the behavioural change to achieve positive, long-lasting impact and solve these systemic challenges rather than short-term quick-win effects. Understanding what influences positive versus negative spillover effects is one of the most exciting pursuits in behavioural economics (Kahneman, 2011). Do we always need intrinsic motivation? Does conformity with social expectations and social pressure exert positive influence and consistent behaviour?

Behavioural spillovers have become a popular area of research since the publication of Dolan and Galizzi (2015), and demonstrations of different spillover effects and theoretical explanations are widespread in the literature, specifically in pro-environmental and sustainability policy areas (Truelove et al., 2014; Nash et al., 2017, 2019, Whitmarsh and O'Neill, 2010). More specifically, there is a considerable debate on what drives positive spillovers (consistency) versus negative spillovers (licensing or balancing) (Monin and Miller, 2001). No research has been conducted with direct empirical support to understand the underlying mechanism and causal pathways (Nash et al., 2017). Evidence is primarily driven by qualitative or correlational analysis rather than empirical and causal relationships (Galizzi and Whitmarsh, 2019). Various situational cues, such as social norms and expectations, influence our self-evaluation and moral self-regard after

engaging in positive behaviour. **How do social norms affect how we perceive our morality? How does this dynamic moral self-image indirectly explain behavioural spillovers?**

In this present research, we aim to answer this question by implementing four quantitative and experimental studies. Through the four empirical papers, the thesis isolates the moderating effect of social expectations on the impact of engaging in one behaviour on moral self-image and the mediating effect of moral self-image on behavioural spillovers. The first paper examines how engaging versus not engaging in behaviours with different social expectations influence our moral self-regard. In the second paper, we replicate the first study and add another behaviour to test the framework's generalizability and the behavioural spillover effects through a novel moderated mediation model. In the third paper, we test again the model in the domain of pro-environmental behaviour related to sustainable transport. In the fourth paper, we explore the concept of virtue signalling in the domain of climate action to examine whether virtue signalling would lead to negative spillover effects compared to pledging in private.

The overall contribution of the thesis is combining three literature strands related to behavioural spillovers, social norms and moral self-image. It provides several theoretical implications for capturing underlying moderating and mediating spillover pathways. It also offers essential practical and policy-relevant implications that aid practitioners in predicting possible behavioural spillovers. Most importantly, it sheds light on the implication of using social norms messaging as behavioural interventions and the importance of intrinsic motivation to induce long-lasting and sustainable behavioural change.

Table 1 below provides the thesis's general outline and presents each one's objective.

Table 1. Overview of the Thesis

Introduction	The introduction to the thesis presents an overview of the research, main findings, and research contribution with a summary of the theoretical and practical implications.
Theoretical Chapter	The theoretical chapter covers an overarching literature review of the three research strands: social norms, morality, and behavioural spillovers. It also presents the literature gaps, expected contribution

	and the two theoretical frameworks with key hypotheses tested in the empirical papers.
Paper 1	Empirical Chapter – It presents an empirical study where we manipulate social expectations and test the effect of the latter on moral self-image and affect for two behaviours related to standing in queue and buying an electric car.
Paper 2	Empirical Chapter – It presents an empirical study where we replicate study 1 with an additional behaviour related to taking the rapid test for COVID-19 and test the effect of moral self-image on behavioural spillovers in the **COVID-19** health behaviour.
Paper 3	Empirical Chapter – It presents an empirical study where we explore the effect of moral self-image and affect on behavioural spillovers in the **environmental domain and sustainable transport.** We also examine whether social expectations affect moral self-image after taking the **train compared to the plane** and how this moral self-image subsequently explains the likelihood of paying for **CO2 offsetting.**
Paper 4	Empirical Chapter – It presents an empirical study where we examine the theoretical framework in the domain of **climate change** and manipulate public observability of the behaviour and its effect on behavioural spillovers. We also test the difference between engaging in a social cause behaviour in **private** by pledging or in **public** through **signalling virtue** on social media.
General Discussion and Conclusion	This chapter presents a general discussion with detailed analysis on how the findings of the present research contribute to the literature. It also presents the practical implications of the findings to inform public policies. The chapter presents the overall limitations of the thesis and suggest ideas for future research.

II. Theoretical Framework

1. Introduction

"Morality dignifies and elevates because it ties us all to something greater than ourselves: each other" (Haidt, 2003).

Being the social animals we are, our self-evaluation and moral self-regard after engaging in a particular behaviour are influenced by various situational cues. With the growing research on behavioural spillovers (Dolan and Galizzi, 2015), there is an ongoing debate on moral consistency and balancing (Miller and Effron, 2010; Blanken et al., 2015). When do people stay consistent with their previous actions, and when do people transgress or compensate (Blanken et al., 2015)? This research offers a model that examines the mechanisms behind moral balancing and how social norms and normative expectations influence behavioural spillovers through self-perceptions and moral self-regard. The psychological science of social norms remains limited (Gelfand et al., 2017); only a few studies have examined how social norms and psychological constructs such as affect, cognitions, and motivation are connected (Cialdini and Trost, 1998). Moreover, social norms have gained momentum since the rise of behavioural economics as a behavioural intervention tool to steer people's behaviour in the desired direction (Reno et al., 1993). However, to the best of our knowledge, we know little about how behaviours driven by social norms impact affect and moral self-image.

This PhD's overarching objective is to understand better how social norms influence behavioural spillovers. To do this, we first examine how social expectations moderate the effect of engaging or not engaging in behaviours on moral self-image and experienced affect. Second, we explore whether moral self-image influences subsequent behaviours (behavioural spillovers). How and in which direction? It attempts to present a new theoretical framework and empirically test the causal relationship using experimental methods. The aim is to empirically isolate the moderating effect of social expectations on the impact of engaging in one behaviour on moral self-image and the mediating effect of moral self-image on behavioural spillovers.

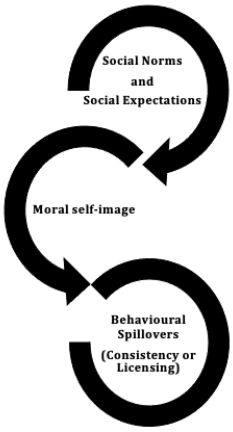
The two research questions are:

Research Question 1: **How do social expectations influence individuals' moral self-image and affect?**

Research Question 2: **How do moral self-image and affect influence subsequent behaviour (behavioural spillovers)?**

This thesis combines literature on behavioural spillovers, social norms, and moral behaviour. In this theoretical chapter, we review three strands of research: 1) behavioural spillovers, 2) moral identity and moral self-image, and 3) social norms and social expectations (see Figure 1). Then, we propose a novel theoretical framework that links the three research areas to answer the research questions.

Figure 1. Three strands of research in the literature review



To date, and to the best of our knowledge, no research has looked at the effect of social norms on behavioural spillovers. We aim to do that by explaining behavioural spillovers through moral self-image. First, the literature review defines behavioural spillovers and provides an overview of the different types, explicitly focusing on the literature debate between moral consistency and balancing. Second, it defines moral self-image and differentiates between moral identity and moral self-regard. Third, it provides an overview of social norms and conformity. Fourth, we lay out the literature gaps and the main theoretical contributions of this research into the three main research strands. Fifth, we present the theoretical framework and a novel statistical model with the predictions and research hypotheses. Table 2 summarises the three research strands and describes the relevant concepts reviewed and how they are linked to each of the two research questions.

Table 2. Summary of the three broad theoretical strands that the thesis aims to integrate, the relevant concepts covered in the literature review, and the justification and contribution of this review to the research questions.

Research strands	**Relevant concepts and references**	**Justification and contribution of the review to the research questions**	**Link to the research questions**
1- Social Norms	- Distinction between independent and interdependent behaviours (Bicchieri, 2006). - Definition of social norms (Bicchieri, 2006). - Measuring social norms (Bicchieri, 2016).	- Define social norms. - Review how to measure social expectations. - Differentiate between social norms and personal norms.	- **Research Question 1: How do social expectations of a behaviour influence individuals' moral self-image?** - This section defines social norms and evidence on social expectations.
2- Moral Behaviour	- Moral Foundations Theory and Social Intuition Model (SIM) (Haidt, 2012) - Differentiation between moral self-image and moral self-identity (Blasi, 1983; Aquino and Reed, 2002) - The dynamic moral self and malleability of moral self-image (Jordan et al., 2015)	- Define moral self-image and understand how to measure it. - Understand whether moral self-image is dynamic, context dependent, and if there is a study on how it influences behavioural spillovers	- **Research Question 1: How do social expectations of a behaviour influence individuals' moral self-image?** - This section defines moral self-image and reviews how to measure it. It also differentiates between moral identity and moral self-image.
3- Behavioural Spillovers	- Definition of types of spillovers (Dolan and Galizzi, 2015) - Measurement of behavioural spillovers (Galizzi and Whitmarsh, 2019) - Systematic review on the debate on moral licensing	- Define behavioural spillovers. - Review the moderators of moral licensing. - Review the debate on moral consistency versus moral balancing	- **Research Question 2: How do moral self-image influence subsequent behaviour (behavioural spillover)?** - This section defines the different types of behavioural spillovers and look for the link between

	and different moderators (Blanken et al., 2015)		moral self-image and moral licensing

2. Behavioural Spillovers

"No behaviour sits in a vacuum, and one behaviour can greatly affect what happens next" (Dolan and Galizzi, 2015). Dolan and Galizzi (2015) developed a conceptual framework with a tri-partition of behavioural spillovers: promoting, permitting, and purging. Nash et al. (2017) define behavioural spillovers as the causal and observable effect of one behaviour on another. The behaviours should be different and sequential and share a motive or have a common link (e.g., pro-environmental behaviours). Behavioural spillovers are measured through various methods, ranging from qualitative self-reports to quantitative empirical experimentation or statistical econometric analyses (Galizzi and Whitmarsh, 2019). Whereas spillover effects could happen from different types of interventions, such as an increase in awareness, technological advancement, or introduction of new infrastructure resources, behavioural spillovers typically result from deliberate behavioural interventions (Nash et al., 2017).
Moreover, Mullen and Monin (2016) also present "Sequential Moral Behavioural Paradigms" that explain how moral behavioural history influences one's subsequent moral decisions. It describes the "experimental situation" (Mullen and Monin, 2016) when a subject chooses to consider past behaviours or actions taken. In the context of moral actions, moral decision-making appears to be influenced by previous behaviours and activities. It is driven by "underlying motives" or factors that affect behaviour (e.g., fundamental values, identity goals, long-term vision). Current literature in this area seems divided on whether past behaviours liberate subjects to act oppositely. On the contrary, they constrain them to be committed and work consistently (Monin and Miller, 2001).

Decades of research in social psychology argue that individuals have a solid drive to remain consistent with their beliefs, intentions, attitudes, past behaviours, and moral self-image (Gawronski and Strack, 2012). This drive towards "consistency," i.e., the tendency to prefer behaving consistently with one's prior beliefs and behaviours, is debated with empirical evidence that a reverse phenomenon is at play (Monin and Miller, 2001). The phenomenon refers to moral balancing; behaving in one direction leads to subsequent opposite behaviour. To inform the empirical debate on the striking contradiction between moral balancing and moral

consistency, emerging research tests different moderators to explain when previous behaviours lead to consistency versus balancing.

2.1 Moral Consistency – Promoting Positive Spillover

Central theories of human behaviour show that people long for consistency in their cognitive thoughts, attitudes, and behaviours (e.g., Beaman et al., 1983; Burger, 1999; Festinger, 1954; Gawronski and Strack, 2012). The main psychological mechanisms at play are the foot-in-the-door effect (Burger, 1999; Freedman and Fraser,1966) and cognitive dissonance (Festinger, 1954). The consistency or constraining effect also applies to promoting negative behaviour after engaging in immoral or wrong actions. This is also known as the what-the-hell effect (Herman and Mack, 1975). For instance, when subjects decide to eat an unhealthy meal, they are more likely to deteriorate from their initial diet motive and engage in more transgressions. They tend to act according to their immediate explanation and satisfy their immediate gratification. These examples and effects promote behavioural spillovers (Dolan and Galizzi, 2015), whether they are all positive or negative behaviours.

2.2 Moral Balancing - Purging versus Permitting Negative Spillovers

When consistency highlights a solid drive to behave the same way, recent empirical evidence shows that past behaviours could liberate people from acting in an opposite way. This is known as moral balancing, which takes two opposite directions: compensatory versus licensing. Theoretically, the literature differentiates between them, although they are both instances of self-regulation and balancing. Compensation highlights a positive goal-striving process; it translates a traditional view of self-regulation. When individuals fall short of a specific goal, they are more likely to regulate and act consistently with their initial motive (Brunstein and Gollwitzer 1996; Tetlock et al., 2000). At the same time, moral cleansing (Zhong and Liljenquist, 2006) is a purging spillover (Dolan and Galizzi, 2015) where people try to restore their moral self-image and integrity after engaging in an immoral act. For instance, in an experiment by Sachdeva et al. (2009), a donation to charity increased after prompting participants to describe their negative traits. However, moral licensing is more controversial as it is unclear why engaging in moral deeds could lead to immoral or unethical behaviour afterwards or refraining from performing positive actions. It is one of the "permitting spillover effects" (Dolan and Galizzi, 2015). Behaving morally could lead to a positive self-image and transgressions in subsequent behaviours.

2.2.1 Moral Cleansing – a purging spillover

Compensation, also known as purging spillover (Dolan and Galizzi, 2015), is a detailed view of self-regulation. People strive to maintain a positive self-image and consistently seek to act positively and morally, making them more likely to take corrective actions after transgressions. Several experiments found that individuals experience moral cleansing to restore their self-integrity in the moral domain. In an investigation conducted by Zhong and Liljenquist (2006), subjects who recalled past unethical acts they performed were more likely to choose cleansing items such as antiseptic wipes.

2.2.2 Moral Licensing – a permitting spillover

To define moral licensing: "Past good deeds can liberate individuals to engage in behaviours that are immoral, unethical, or otherwise problematic, behaviours that they would otherwise avoid for fear of feeling or appearing immoral" (Merritt et al., 2010, p.344). Compared to consistency, moral licensing is somehow counterintuitive. According to self-perception theory, individuals induce their attitudes from observations of their actions, influencing their decisions in subsequent behaviours (Bem, 1972). Mullen and Monin (2016) argue that the driving force for moral licensing is a conflict of motives.

Hence, it has yet to be clear why engaging in positive behaviours could liberate people to do less good. Licensing appeared in various areas and domains, from job hiring (Cascio and Plant, 2015; Monin and Miller, 2001), racism (Choi and Crandall, 2008; Effron et al., 2009), charity donations (Conway and Peetz, 2012; Sachdeva et al., 2009), health behaviours (Mukhopadhyay, Sengupta, and Ramanathan, 2008), dishonesty and cheating (Jordan et al., 2011; Mazar and Zhong, 2010).
The literature differentiates between two types of distinct psychological processes that help interpret moral licensing: the "Moral Credential Model" and the "Moral Credit Model" (Merritt et al., 2010; Miller and Effron, 2010). First, according to the Moral Credential Model, engaging in the initial behaviour alters the interpretation of the subsequent behaviour. In other words, taking positive actions provides a lens through which the second behaviour is justified. So, in this model, immoral behaviour is not perceived as unfavourable, but the mere engagement in initial positive behaviour gives credentials for an individual to behave in an immoral act (Monin and Miller, 2001). Second, according to the Moral Credit Model, people accumulate moral credits in a

hypothetical moral bank account. Moral credits are used to either engage in moral actions or offset immoral actions (Merritt et al., 2010; Miller and Effron, 2010). Moral credits make immoral acts more permissible. Sachdeva et al. (2009) interpreted the licensing effect in terms of having an accrual of "moral currency" (p.527) that could be used to liberate people to engage in less moral acts. In addition, Jordan et al. (2011) and Zhong et al. (2009) suggested a moral equilibrium to justify the licensing effect.

Moral licensing is often perceived as a counterintuitive spillover effect (in the sense that acting positively might lead to transgression); hence, it has taken significant momentum in the literature in the past decade. Various studies examined the moderators that lead to consistency versus licensing after engaging in a positive moral act (Blanken et al., 2015; Mullen and Bonin, 2016). First is the construal level, where recalling abstract past behaviour leads to moral consistency compared to recalling concrete actions (Conway and Peetz, 2012; Cornelissen et al., 2011). Second, commitment versus progress; perceiving a previous act as a commitment to the cause leads to consistency compared to perceiving past actions as progress towards a set goal (Susewind and Hoelzl, 2014). Third, identifying the social cause increases the likelihood of consistency (Meijers and Rutjens, 2014; Effron et al., 2009). Fourth, the value and cost of the initial act (Gneezy et al., 2012); choosing to undertake a costly initial pro-social act increases the probability of consistency compared to engaging in a costless act, which will probably lead to balancing effects (Joosten et al., 2013). This is also true in virtue signalling, known as "slacktivism." It describes engaging in a costless act such as liking or sharing a page on social media or wearing a ribbon to promote a social cause. These behaviours lead to a warm glow effect and positive social desirability, which liberates and licenses people to transgress in subsequent acts. Another value reflection property is the observability of the initial act (Kristofferson et al., 2014). Engaging in pro-social acts in public increases the likelihood of licensing compared to private pro-social behaviours.

2.3 Debate on Moral Consistency versus Balancing

Despite many experiments and studies conducted to look for moderators to explain the consistency versus licensing debate, what drives change in the moral credit balance remains to be determined. Why does performing the same behaviour lead to different subsequent actions? Some people feel entitled to moral credit, whereas others behave consistently.

On the one hand, research on moral identity argues that high moral self-importance predicts positive and consistent pro-social behaviour (consistency theory and identity congruent

motivations (e.g., Blasi, 1983)). On the other hand, moral licensing and moral cleansing research (Jordan et al., 2011; Merritt, Effron and Monin, 2010; Sachdeva et al., 2009; Zhong and Liljenquist, 2006) reveal that individuals with high moral self-importance do not always behave consistently over time.

In addition, despite the vast literature on moral licensing and cleansing (more than 400 articles discuss this topic), replicating the experiments often leads to failure (Rotella and Barclay, 2020). In the meta-analysis review of moral licensing, Blanken et al. (2015) reviewed different theoretical relevant moderators (such as the domain of the behaviour, type of induction (trait versus action), and type of behaviour measured). However, when these effects are measured, little attention is given to the social environment and social expectations. There is no research on how social expectations and norms influence moral licensing. For instance, when replicating a study on cheating and moral cleansing, examining whether cheating is a social norm and how this difference in social expectations influences the likelihood of moral licensing is essential. We propose that the present research contributes to this ongoing debate in the literature on when people are likely to stay consistent with initial behaviours or engage in moral licensing. We do so by examining the role of social norms on the likelihood of experiencing moral licensing or negative behavioural spillovers. Next, we examine the literature on morality.

3. Moral Identity and Moral Self-Image

3.1 Morality and Moral Systems

Philosophy and psychology literature define " morality " differently (Cushman and Young, 2009; Graham et al., 2011). In philosophy, morality is tackled in a “prescriptive and proscriptive” way, pointing out right and wrong (Bentham, 1823/1996; Kant, 1785/2010). Whereas psychologists attempt to investigate when people act in moral and immoral ways and what drives their moral actions (Kohlberg, 1981; Kohlberg and Kramer, 1969; Haidt, 2001, 2012; Narvaez and Lapsley, 2009); how people behave and identify their moral self (Blasi, 1983; Aquino and Reed, 2002; Reed, 2002; Doris, 2002; Graham et al., 2011), how individual characteristics and situations, factors drive these behaviours (Jordan et al., 2015).

What does it mean to be moral and have moral systems? Moral systems are “[…] interlocking sets of values, virtues, norms, practices, identities, institutions, technologies, and evolved psychological mechanisms that work together to suppress or regulate self-interest and make

cooperative societies possible." (Haidt 2012, p. 270). Morality is the basis of social behaviour regulation, as it defines what is perceived to be "right" and "wrong" in a given society or context. According to research, moral self-regard is a crucial component of people's identity and self-evaluation (Aquino et al., 2011; Blasi, 1983), which motivates them to report positive perceptions of how they view their morality (Jordan et al., 2015) and how they compare their morality to others (Barkan et al., 2012; Epley and Dunning, 2000; Tappin and McKay, 2017). To deviate from this positive pattern and engage in immoral action would provide evidence that a person may not be as moral as they would like to think. Consistent with this, research examining moral self-perceptions has found that moral and immoral behaviour directly impacts a person's moral self-image (i.e., their temporary and highly malleable evaluations of how moral they perceive themselves; Monin and Jordan, 2009). Specifically, engaging in moral behaviour leads to a more positive moral self-image, while engaging in immoral behaviour has the opposite effect (Jordan et al., 2015; Perugini and Leone, 2009) and can also lead to negative self-conscious affect just as guilt and shame (Tangney et al., 2007). Thus, moral behaviour is preferable if a person wishes to maintain positive self-perceptions.

An ongoing question in the literature on moral behaviour is when and why people act as they do. The cognitive-developmental model has taken an influential theoretical ground (Kohlberg, 1981), emphasising moral reasoning and its influence on moral behaviour. More recently, moral psychology has started steering the direction of moral reasoning and tapping into moral intuition and social and cultural influences on moral behaviour (Haidt, 2001). According to the **Moral Foundation Theory (MFT**), five moral foundation categories explain the origin and variations of moral reasoning (Haidt and Joseph, 2004; Graham et al., 2001; Haidt, 2001, 2012). These categories are care, fairness, loyalty, authority, and purity. A "prequel of MFT" is the **Social Intuitionist Model (SIM),** which argues that moral judgement results from automatic, fast, emotional, and unconscious moral intuition rather than cognitive reasoning (Haidt, 2001, 2012). These intuitions could be edited and justified afterwards by moral reasoning. SIM resonates with the dual system model that explains how most behaviours result from an intuitive, automatic, and fast system (Kahneman, 2011). Previous research has shown that morality is related to cooperation (Rai and Fiske, 2011; Tomasello and Vaish, 2012). Recently, the theory of "**Morality as Cooperation**" explained that "morality is a collection of biological and cultural solutions to the problems of cooperation and conflict recurrent in human social life" (Curry, 2016; Curry et al., 2019). It identifies moral foundations related to cooperation that are omitted in the MFT. Curry et al. (2019) showed that seven cooperative behaviours are positive moral valence across 60 studies. The seven types of cooperation: (1) the allocation of resources to kin; (2) coordination to mutual

advantage; (3) social exchange; and conflict resolution through contests featuring (4) hawkish displays of dominance and (5) dove-ish displays of submission; (6) division of disputed resources; and (7) recognition of possession (Curry et al., 2019). Typically, behaviours that promote the welfare of others are perceived as positive (e.g., donating to charity, volunteering, saving energy). At the same time, actions that inhibit self-interest are also perceived as moral and positive (e.g., refraining from accepting bribery or not littering). On the other hand, immoral acts describe instances where the self-interested behaviour causes harm or unjustly disadvantages others (e.g., cheating, bribing).

For so long, research focused on outlining the right behaviours, prescribing ethical guidelines, highlighting how people should behave, and defining wrong and immoral actions. However, a recent study realised the importance of understanding actual behaviours and adopting a more descriptive approach to tackling the problem (Shalvi et al., 2015). The field of behavioural economics increasingly attempts to investigate how individuals are tempted to behave in a certain way and how contextual cues, psychological biases, and heuristics influence their actions (Bazerman and Gino, 2012). Unethical and immoral acts, such as cheating, littering, and accepting bribes, to name a few, are not always conducted by "bad people" or "whose moral compasses are internally damaged" (Moore and Gino, 2013). **On the contrary, "ordinary" individuals with high moral identity could transgress and engage in hostile acts. Despite reporting being moral, people might engage in ethical and immoral behaviours** (Gregory-Smith et al., 2013). Whatever we do, we can find ways of justifying ourselves - very few people see themselves as acting immorally. Next, we review what moral identity is.

3.2 **Moral Identity**

Aquino and Reed (2002) build on both the Socio-Cognitive Model (Bandura, 1999; Bandura et al., 1996) (which emphasises the self-regulatory mechanisms and moral standards/ sanctions that predict ethical behaviour) and the Cognitive-Developmental Model (Kohlberg, 1981) to define moral identity. Moral identity is presented as a parameter of social identity (Blasi, 1983; Reed, 2002). According to the Social Identity Theory (Tajfel, 1972; Tajfel and Turner, 1979), people can identify themselves with distinct identities (religious, political, shared traits, and interests, among others...). These diverse self-identities constitute the personal social self-schema (Markus, 1977). Moral identity is then referred to as one type of social schema or another type of social identity (Aquino and Reed, 2002; Reed, 2002). It is also defined as "a self-conception organised around a

set of moral traits" (Aquino and Reed, 2002, p.1424). Since it represents moral traits an individual internalises and identifies, it tends to be stable over time (Damon and Hart, 1992). Aquino and Reed (2002) proposed a measurement scale to assess two dimensions of the self-importance of moral identity: the internalisation dimension (describing how important is possessing those moral traits to the self) and the symbolisation dimension (explaining how important it is to show others that one has those moral traits). The literature differentiates between the moral identity that is more internalised and static over time compared to the moral self-image that is more dynamic and context-dependent.

3.3 Moral Self-Image

The idea that individuals are either moral or immoral is inaccurate (Hofman et al., 2014). Although people identify themselves as honest, they tend to engage in unethical or destructive behaviours, which might influence their moral self-regard, self-concept, or self-image.

The literature differentiates between **moral self-image** and **moral identity**. Most people long to be good, act morally, and project positive moral self-image towards their self-evaluation and others (Barkan, 2015; Jordan et al., 2015; Monin and Jordan, 2009). **However, moral self-image is a dynamic concept that is malleable rather than fixed and is influenced by context and situational cues (Jordan et al., 2015).** It is purely subjective as it measures how (im)moral people perceive themselves at a particular time. moral self-image captures a temporary perception of the moral self that fluctuates based on recent feedback, social comparison, and other situational factors activated in one's working memory (Jordan et al., 2015; Monin and Jordan, 2009). Based on context, social identity, and situational factors, people evaluate their moral self-image and give themselves positive or negative labels (Kernis and Goldman, 2003). Unlike self-esteem and moral identity (which are more stable and less likely to be affected by temporal changes or actions), moral self-image resides in a "working self-concept" (Kernis and Johnson, 1990). Therefore, the literature discusses that moral self-image changes based on different events in the social world; it is also influenced by feedback from the environment. This is reflected in the unpredictability of people's behaviour, where someone with high self-morality could engage in unethical actions.

Moore and Gino (2013) argue that engaging in unethical or immoral behaviour is caused by two reasons: "intrapersonal" and "interpersonal". Intrapersonal is the result of human limited cognitive capacity and mental shortcuts, whereas interpersonal describes the influence of others. Moore and Gino (2013) describe the effect of social interaction on moral behaviour as "magnetic material"

that pulls the moral compass from north to south. This external force deviates the moral compass in the opposite direction, leading to people behaving uniquely. The magnetic moral compass, influenced by external contextual cues, takes us to the literature on "Sequential Moral Behavioural Paradigm" and, more generally, "behavioural spillovers" (Dolan and Galizzi, 2015). After engaging in a particular behaviour, the feedback we receive or give ourselves affects our moral self-concept and leads to several types of spillover.

3.4 Social Norms and Moral Self-Image

The literature argues that moral self-image is malleable and influenced by contextual, social, and situational factors (Jordan et al., 2015). Previous studies have tackled the influence of social norms on altruistic and pro-social behaviours (Berkowitz, 1972; Cialdini and Trost, 1988). These studies examined how social norms could increase the likelihood of engaging in certain pro-social behaviours. However, no analysis in the literature isolates the relationship between social expectations/standards and moral self-image. In this research, we aim to investigate whether there is a difference in moral self-image and experienced affect after engaging in a behaviour that is approved and socially expected compared to engaging in a behaviour that is approved but socially unexpected. This research proposes to fill this gap and combine these two lines of research to inform the debate on when behaviours are consistent with previous actions and when they are not (e.g., moral licensing effects). Next, we discuss social norms and social expectations in detail.

4. Norms and Conformity

Driven by the need to maintain a positive self-concept, people seek approval and reach out to others for social guidance. They strive to conform with others and behave consistently with their actions and beliefs (Sunstein, 2019). The literature differentiates between informational and normative conformity (Deutsch and Gerard, 1955). Informational conformity describes the desire to form an objective reality of what is correct. However, normative conformity expresses the desire to gain social approval from other people. Research in social psychology has tried to disentangle the difference between the two types of conformity and argue that they are independent (Cialdini and Trost, 1998).

4.1 Conformity in Social Psychology

The Asch experiment is a classic social experiment in social psychology demonstrating how people tend to conform with others, especially when exposed to social pressure by the vast majority (Asch,

1956). It presents how social pressure from a majority group could push people to conform. In the experiments, participants were asked to match the lines' length on two cards, which is an easy task to perform. However, in each experiment group, there was one actual participant; the rest pretended to participate and were instructed to provide the incorrect answer. The findings showed that participants were more likely to conform with most participants who reported the wrong answer (Asch, 1956). Another social experiment that highlights the power of conformity and norms is the Milgram experiment (Milgram, 1960). It highlights the critical role of situational factors and authority figures in increasing obedience and compliance. In the experiment, participants continue to administer shocks to conform to the expectations of the experimenter or to avoid not confirming in front of others. Finally, the Bobo Doll experiment (Bandura, 1960) provides insights into how individual behaviour is shaped by social learning and is influenced by observed behaviours in the social environment. Overall, these social experiments from the social psychology literature lay the foundation for the importance of conformity and the interplay between individual behaviour and social norms (Sunstein, 2019). Understanding and diagnosing social norms and social expectations have significant implications for designing effective policy interventions.

4.2 Social Norms in Psychology

To define, norms are “rules or standards that members of a group understand, and that guide and constrain social behaviour without the force of law” (Cialdini and Trost, 1998, p. 152). Norms can take different forms; we know moral, social, and personal norms. Adjusting or altering one’s behaviour to conform with others is a tendency that has been found in different domains (e.g., energy consumption (Schultz et al., 2007), donation, and alcohol consumption (Borsari and Carey, 2003)).

Social norms have always been central in literature, from economic and social psychology to behavioural sciences. One of the questions raised is why people follow the example. The answer includes the desire to comply, avoid social sanctions, gain recognition, identify with a social group, or behave according to self-identity. Moreover, several theories attempted to explore social norms as one of the drivers of understanding human behaviour: the Theory of Reasoned Action (Fishbein and Ajzen, 1975), the Theory of Planned Behaviour (Ajzen, 1991), the Focus Theory of Normative Conduct (Cialdini et al., 1990), the Social Identity Theory (Tajfel and Turner, 1979), the Self-

Categorization Theory (Turner et al., 1987), and the Theory of Normative Social Behaviour (Lapinski and Rimal, 2005; Rima and Real, 2005).

The **Theory of Reasoned Action** (Fishbein and Ajzen, 1975) and the **Theory of Planned Behaviour** (Ajzen, 1991) discussed subjective norms. They argue that attitudes and subjective norms drive behavioural intentions. Beliefs that cause subjective criteria are considered normative beliefs as they reflect what people approve and disapprove of.

In the **Focus Theory of Normative Behaviour**, descriptive norms are one type of social norm (Cialdini et al., 1990; Reno et al., 1993). They distinguish between two kinds of social norms: descriptive and injunctive. On the one hand, descriptive norms describe what people do, signalling prevalence and what is considered "normal." In circumstances where the appropriate behaviour is unclear, what prevails is the expected behaviour or "social reality" of what others are doing (Festinger, 1954). Conforming to descriptive norms results from heuristic and cognitive shortcuts, such as the social proof heuristic and the "common is moral" heuristic (Lindström et al., 2018). Commonness induces the belief that a behaviour is moral and leads to conformity. Also, numerous experiments in behavioural economics show that humans are social animals. When faced with decision choices, they tend to rely on the automatic system and herd rather than engage in rational deliberation (Ariely, 2009).
On the other hand, injunctive norms refer to what is (dis)approved by others or what is appropriate to do "ought to be done," signalling acceptance. Moreover, injunctive norms could be prescriptive and prospective. Prescriptive norms describe actions to "approach" positive outcomes, while prospective norms refer to activities that inhibit adverse effects (Janoff-Bulman et al., 2009). Prescriptive norms motivate the intent to obtain social approval from others, whereas prospective norms avoid social disapproval (Herrmann, 1971).

According to **Social Identity Theory and Self-Categorization Theory** (Tajfel, 1972; Tajfel and Turner, 1979; Tuner et al., 1987; Stets and Burke, 2000), individuals' sense of identity is influenced by the characteristics and norms of the groups that are relevant to them and to which they self-categorize themselves. Their definition of themselves is associated with the group norms.

The **Theory of Normative Social Behaviour** (Lapinski and Rimal, 2005; Rimal and Real, 2005) builds on the distinction between descriptive and normative behaviours and identity's role in explaining the relationship between descriptive norms and behaviour. It looks into the moderators

that show the effect of descriptive norms on behavioural change; these include injunctive norms, group identity, and outcome expectations.

4.3 Defining and Diagnosing Social Norms

The growing literature in psychology and behavioural science produces empirical evidence on the effectiveness of social norms and their role in predicting behaviours (Bicchieri, 2006; Cialdini et al., 1990; Elster, 1989; Lapinski and Rimal, 2005; Mackie, 1996; Miller and Prentice, 2016; Paluck and Shepherd, 2012; Tankard and Paluck, 2016).

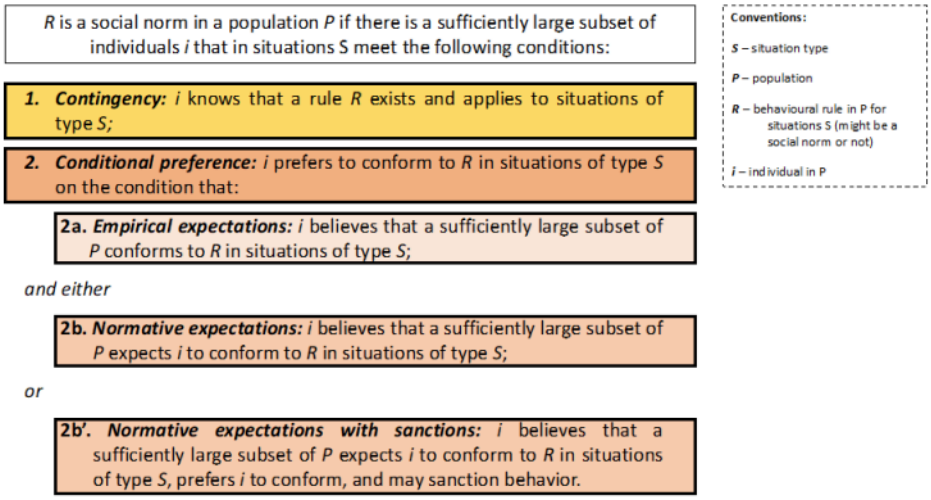
A solid and empirical definition of social norms is presented in the "Grammar of the Society" (Bicchieri, 2006). It starts with differentiating between independent and interdependent behaviours. Interdependent behaviours are based on social preference, as other people's beliefs or actions matter, whereas independent behaviours are not conditional on other people's choices and behaviours. For instance, habits, social customs, religious behaviours, and moral injunctions are independent as they occur regardless of what others believe or do (Bicchieri, 2006). So, independent behaviours are performed irrespective of what others think or do, and interdependent behaviours are based on social expectations.

Social norms are defined as "expectations" and "conditional preferences" to follow them. Bicchieri (2006) explains that one's preference to conform to a social norm is conditional upon others expecting conformity. Both the expectations and conditional preferences lead to collective behaviour that further confirms the existence of the norm itself. Bicchieri (2016) argues that the following conditions should be met in her definition of social norms: contingency and conditional preferences, including empirical and normative expectations.

Figure 2 presents the social norms' condition to exist (Bicchieri, 2006). First, behaviour is considered a social norm if a subject knows that this behaviour exists and applies to a particular situation (a contingency). Second, a topic should believe that most of the population conforms to the behaviour (this represents empirical expectation). Third, a subject believes that most of the population expects them to conform to the conduct (normative expectation), or one supposes that others expect them to conform. Any deviance from the social norm will lead to sanctions (normative behaviour with sanctions). In other words, fulfilling the contingency, empirical, and normative expectations is sufficient for a behavioural rule to become a social norm. To summarise,

social norm requires an individual to expect others to conform to the behaviour and expect others to conform.

Figure 2. Conditions for Social Norms to Exist (Bicchieri, 2006)



Bicchieri (2006, 2016) emphasises the conditional aspect of following a norm as key to diagnosing/identifying and differentiating it from other social or moral preferences in this definition of social norms. "Compliance with a norm is conditional on having the right expectations" (Bicchieri, 2016, p.11). To illustrate, having an unconditional preference (moral preference or obligation) differs from having a conditional preference to be fair (i.e., based on social expectations). Put differently, I can care little about fairness but still act reasonably because I expect others to do the same (empirical expectation or descriptive). Others expect me to be fair (normative expectation or injunctive).

When people have a "*conditional social expectation*" to behave in a certain way (e.g., stop at a red traffic light or accept child marriage), they form beliefs about what is suitable or appropriate to do based on what others are doing and what they believe others expect them to do. These expectations are either factual or normative. On the one hand, empirical expectations are elicited by observing what others are doing. When most people stand in a queue in a public place, we form empirical expectations that this is being done. On the other hand, **normative expectations** are the beliefs of what other people expect us to do. It is simply other people's *"Personal Normative Belief*," also known as personal norms (Schwartz, 1977). For instance, I believe that most people in my city expect me to stand in the queue." Personal normative belief is what we think is right and ought to be done. It is important to note that my personal beliefs could differ from my normative

expectations (i.e., different from other people's normative beliefs). Also, normative expectations cannot be observed; people assume what others think should be done. For example, a husband can beat his wife, thinking or believing that other people in his neighbourhood approve of this behaviour (his personal normative belief is to beat her, and he expects others to believe the same). This distinction in social expectations and their underlying conditional aspects improves social norms' definition, diagnosis, and measurement. More specifically, *Descriptive Norms* (called the same by Cialdini et al.'s Focus theory, 1990) are defined as follows:

"A descriptive norm is a pattern of behaviour such that individuals prefer to conform to it on the condition that they believe that most people in their reference network conform to it (empirical expectation)." (Bicchieri, 2016, p. 12)

Linking a descriptive norm to empirical expectation differentiates descriptive examples from habits, customs, and moral injunctions. Even though these behaviours are expected and describe what other people do. Descriptive norms are conditional on what others do, so what others do affects engaging or avoiding a behaviour. A custom or habit is performed regardless of what others think (they don't entail any conditional social expectation and preference). The other type is **social norms**, subjective norms in the Theory of Reasoned Action (Fishbein and Ajen, 1975); injunctive norms or what others ought to do by Cialdini et al., 1990) contain both empirical and normative expectations.

"A social norm is a rule of behaviour that individuals prefer to conform to on condition that they believe that a) most people in their reference network conform to it, and b) that most people in their reference network believe that they ought to conform to it (normative expectation)." (Bicchieri, 2016, p 22)

Biccheri's theory (2016) focuses on defining and identifying the characteristics of "interdependent behaviours" (i.e., behaviours that are anchored on social expectations and external regulation). To fully understand norm-based behaviours, we will look at the norms that drive "independent behaviours." These behaviours are endorsed regardless of what others think or believe; they are referred to as habits, moral injunctions, or cultural customs. Bicchieri (2006) refer to them as Personal Normative Beliefs, and Schwartz (1977) identifies them as Personal Norms as part of the Norm Activation Model (NAM).

4.4 Personal Norms: Introjected and Integrated Norms

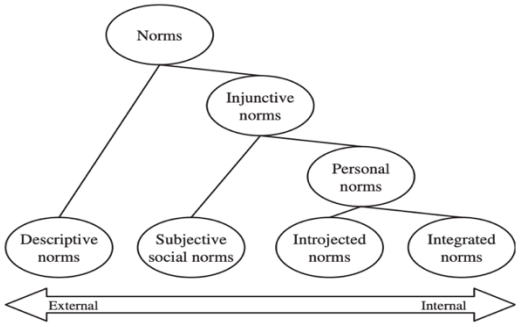
According to the NAM (Schwartz, 1977), social norms are the results of social expectations: i.e., the normative explanation of the behaviour is based on expectations, sanctions, and obligations anchored in a social or reference group. Personal norms are anchored around oneself and are based on self-expectations (Schwartz, 1977). Self-expectations result from moral obligations, which activate an internalised sense of values. Social norms are also driven by socially mediated sanctions that explain why these social expectations result in a particular behaviour or action (Blake and Davis, 1964). Social norms are accompanied by social pressure and fear of social sanctions, whereas personal norms are grounded on an individual's conviction. Bamberg et al. (2007) argue that the main characteristic of subjective norms is internalisation; so, behaving in line with personal norms is to avoid a feeling of guilt and regret rather than fear of social sanctions. The NAM model has emerged and demonstrated altruistic and pro-social behaviours (e.g., donation, volunteering, and helping people in need, among others). However, there is a body of research that works on areas related to pro-environmental behaviours, such as energy conservation (Black et al., 1985), recycling (Guagnano et al., 1995), and green products (Thogersen, 1999).

The proposed taxonomy by Thogersen (2006) is presented in Figure 3 below. The injunctive norms are considered subjective social or personal norms, depending on whether they are anchored on external or internal expectations. *Subjective social norms* are externally regulated and controlled by the expected reward and punishment (e.g., social-informal and formal sanctions). This type of norm is the social norm, as defined by (Bicchieri, 2006; Cialdini et al., 1991; Reno et al., 1993; Kallgren et al., 2000), which is reinforced by external social expectations on what others believe one should or ought to do. People have internalised behavioural regulation by the anticipated degree of guilt and pride (Ryan and Deci, 2000). So, in this case, social pressure becomes "inner pressure" to comply with the internal sense of values and moral obligations (Schwartz, 1977; Biel et al., 1999).

However, this distinction between the two levels of internalisation, i.e., external social norms and internal personal norms, is perceived as oversimplified (Thogersen, 2006). It does not differentiate the underlying motive or drive behind the individual norm behaviours. Thogerson (2006) breaks down the subjective norms' psychological states into two categories that vary depending on their

degree of internalisation. So, the suggested taxonomy includes *Introjected Norms*, defined as one type of personal norm that is superficially internalised and is driven by the anticipated feeling of providing or guilt. In other words, people engage in the behaviour to avoid an anticipated sense of guilt or regret or seek a feeling of pride. The other type of personal norm is Integrated Norms, defined as deeper internalised and integrated norms based on deep evaluation and reflection on complying with one's own values and goals. The differentiation is both internalisation and motivational content (Thogersen, 2006). The proposed norm taxonomy is presented in Figure 3 below.

Figure 3. The proposed norm taxonomy (Thøgersen, 2006)



4.5 Bad Norms or Socially Harmful Norms

"Freedom from a bad norm is a public good that is often very difficult to bring about" (Bicchieri, 2006, p.42). People long to comply with the norm and get distressed when they do not conform (Reno et al., 1993). In some instances, the norm could be socially harmful, such as accepting bribery or crossing traffic lights, discriminating against gender, race, religion, or littering. Such negative norms are inefficient for social and economic welfare and wellbeing. A conflict between injunctive and descriptive norms emerges.

4.6 Social Norms and Behavioural Change

With the rise of behavioural science, norms have been extensively reported in behavioural experiments as a tool to motivate behavioural change (Aarts and Dijksterhuis, 2003; Cialdini and Goldstein, 2004; Cialdini et al., 1991). Using norms triggers automatic, fast, and intuitive reactions and behavioural changes. From a psychological point of view, humans are social animals, automatically wired to use the social norm as a reference point and tend to adhere to what is common (Vlaev and Dolan, 2015). Research in cognitive neuroscience (Klucharev et al., 2009)

shows that a conflict with group opinion leads to neuronal response, and the amplitude of this effect results in behavioural adjustments. This shows a reinforcement learning model where social norms cause conformity (Vlaev and Dolan, 2015). This learning mechanism is also reinforced during evolution, resulting in an automatic valuation process of conforming to the norm (i.e., behaviours in line with the norms are automatically assessed as more rewarding) (Klucharev et al., 2009). A mental heuristic leads to this automatic processing or habit: "If everyone is doing it, then it must be a good action" (Erev and Barron, 2005).

As a powerful tool for behavioural change, social norms have been used in several policy areas: public finance management to increase payment of taxes (Hallsworth et al., 2014); environment to improve the reuse of hotel towels (Goldstein et al., 2008); reduce energy and water consumption (Schultz et al., 2007); health behaviours to improve screening, hand washing (Judah et al., 2009), reduce alcohol (Perkins et al., 2011) and reduce health-risk behaviours. The underlying idea is to highlight the descriptive norm (i.e., what other people do), and in some cases, it is used in conjunction with injunctive norms (what is the right thing to do or not to do). For example, in the pro-environmental domain, engaging in positive behaviour is influenced by intrinsic motivation (related to one's pro-environmental position or sense of identity) and extrinsic motivation about conformity with social expectations (Steg et al., 2015). Both descriptive norm messages as well as feedback information with social comparison showed significant effects in increasing pro-environmental behaviours (Abrahamse and Steg, 2013).

An innovative and trendy type of norm has been explored lately to promote behaviours that are not descriptive; they are referred to as "Dynamic Norms**"** (Sparkman and Walton, 2017; Sparkman and Walton, 2019). Dynamic norms involve informing people about the collective change in behaviour, and it is shown to lead to meaningful behavioural change. It infers behavioural change even though the norm does not yet exist. The underlying mechanism that drives evolution is the increased self-efficacy (by demonstrating that change is possible), highlighted injunctive norms (by showing that others expect to change), and relating with one's social identity (by making change consistent with personal and social identity).

4.7 Rethinking Social Norms

In the last two decades and with the rise of behavioural economics, behavioural interventions using social norms have gained tremendous momentum. Implemented as tools for effective behavioural change, the literature focuses on defining and characterising several norms, measuring norms, designing experiments using norms, and even changing norms (Bicchieri, 2006, 2016; Cialdini et al., 1991). However, **little attention is given to how the prevalence of social norms and conformity to the norms influence people's self-perception, affect, and subsequent behaviours.** To our best knowledge and based on the above literature review on social norms, no study examined how conforming to a social norm versus personal norm influences affect and moral self-image. In this research, we suggest rethinking social norms as an effective tool for behavioural change and a behavioural phenomenon that could influence how people evaluate/ perceive their actions and decide to behave subsequently. For instance, social norms have been used as an effective tool to make people reduce water and energy consumption. However, no studies looked at how people perceive their actions when energy conservation is the social norm (expected) versus when this behaviour is not socially expected. Subsequently, how individuals perceive and evaluate their previous behaviours influences their decisions afterwards. In other words, the prevalence of norms could by itself affect moral self-image and affect. So far, the literature has focused on norms as a tool to change behaviour; we aim to employ social norms to understand the occurrence of behavioural spillovers through morality. The following section will summarise the literature gaps related to the three research strands and clearly outline how this research aims to fill these gaps.

5. Literature Gaps[1] and Research Contributions

So far, we have reviewed the literature on three research strands: behavioural spilloverssocial norms, morality, and. Before presenting the research hypotheses and the suggested framework, the next section will summarise the main literature gaps and how the thesis aims to fill these gaps (Table 3).

To sum up, this present thesis attempts to contribute to the literature in three areas:

1) The first literature gap concerns the effect of social norms on experienced affect and moral self-regard. To our knowledge, no study has investigated how behaviours dependent on social expectations compared to behaviours independent of social expectations influence

[1] Literature related to social psychology, behavioural economics and behavioural spillovers.

moral self-image and affect. This research will connect these two lines of research on social norms and moral self-image.

2) The second literature gap is the effect of moral self-image on behavioural spillovers. To date, evidence on moral licensing and moral cleansing is based on moral credit balance. However, to our knowledge, no study empirically measured how morality through moral self-image mediates the effect of behaviour one on behaviour 2. In addition, no study quantifies and tests the moral credit model quantitatively.
3) The third literature gap is the relationship between social norms and behavioural spillovers. There is a debate in the literature on when people are consistent with their previous behaviours and when they feel licensed to transgress (Blanken et al., 2015). Different possible effects were studied, but social expectations and social norms were not. This research aims to look at whether behaviours based on social expectations will lead to different behavioural spillovers.

Table 3. Summary of the main theoretical concepts and literature gaps

	Literature gaps	**Related research strands from the literature review**	**Contribution of the thesis to fill this gap**
1	Social norms are explored as a tool to change behaviour, and no research looked at how social norms influence moral self-perceptions and affect. The effect of behaving in congruence with a social norm (interdependent behaviour) versus personal normative beliefs (interdependent behaviour) has not been investigated. Is there a difference in interdependent and independent behaviours on moral self-image and affect?	- Social norms and expectations - Morality and Moral Self-Image	- Investigate whether social expectations lead to a significant difference in the self-reported experienced affect and moral self-image. - Social expectations (as a contextual effect) moderate the effect of engaging versus not in behaviour on the experienced moral self-image. In other words, with different social expectation levels, there is a significant difference in moral self-image when engaging or not in certain behaviour.
2	No study looked at how moral self-image (which is dynamic and context-dependent) influence behavioural spillovers (particularly moral licensing	- Morality and Moral Self-Image - Behavioural Spillovers	- The effect (mediator effect) of moral self-image on the effect of engaging in behaviour one behaviour 2 (i.e., does the change in moral self-image influence the likelihood of

	and moral consistency) and on how to quantify the moral credit model		behavioural spillovers (i.e., consistency versus compensatory effects?).
3	The effect of social expectations on behavioural spillovers through moral self-image (no study examined social norms as a moderator and moral self-image as a mediator)	- Social Norms - Morality and Moral Self-Image - Behavioural Spillovers	- The effect of moral self-image on behavioural spillovers is moderated by social expectations.
4	The role of virtue signalling (public) compared to private pledging on spillover effects	- Virtue Signaling - Commitment/ Pledging - Behavioural Spillovers	- The effect of publicly observable pro-environmental behaviour on moral licensing. - The comparison between public and private behaviour on spillover effects.

6. Suggested Theoretical Framework and Model

This section presents a novel theoretical framework that explores how social norms influence MSI[2] and affect, resulting in different behavioural spillovers. The framework shows how differentiating between the "Expectation" (i.e., behaviours that are approved and socially expected) and the "Exception" (i.e., behaviours that are approved but not socially expected) helps us better explain when people experience moral licensing or consistency.

We use descriptive norms or empirical expectations to refer to common behaviours and describe what people do (Cialdini et al., 1990; Reno et al., 1993; Cialdini et al., 2006; Bicchieri et al., 2014). We use injunctive norms or normative expectations to refer to approved behaviours. Injunctive norms could take two forms: prescriptive norms to define the right behaviour (what is "ought to") and proscriptive norms to explain the wrong behaviour (what is "ought not to") (Janoff-Bulman et al., 2009). Prescriptive and proscriptive are mutually exclusive but mutually inclusive of descriptive and non-descriptive norms. They are either the result of an intrapersonal belief (cognitive) or an interpersonal belief of what others believe is the right thing to do (social) (Janoff-Bulman et al., 2009; Pavey et al., 2008).

First, we refer to the "**Expectation**" to describe behaviours that i) have injunctive norms or normative expectations and ii) have descriptive norms or empirical expectations (i.e., we expect others to undertake them, and they are approved). Second, we refer to the "**Exception**" to describe

[2] In this section onwards, we refer to Moral Self-Image by MSI.

behaviours that i) have injunctive norms or normative expectations (they are approved of) and ii) are non-descriptive or do not have empirical expectations (i.e., we do not expect others to undertake them) (Table 4).

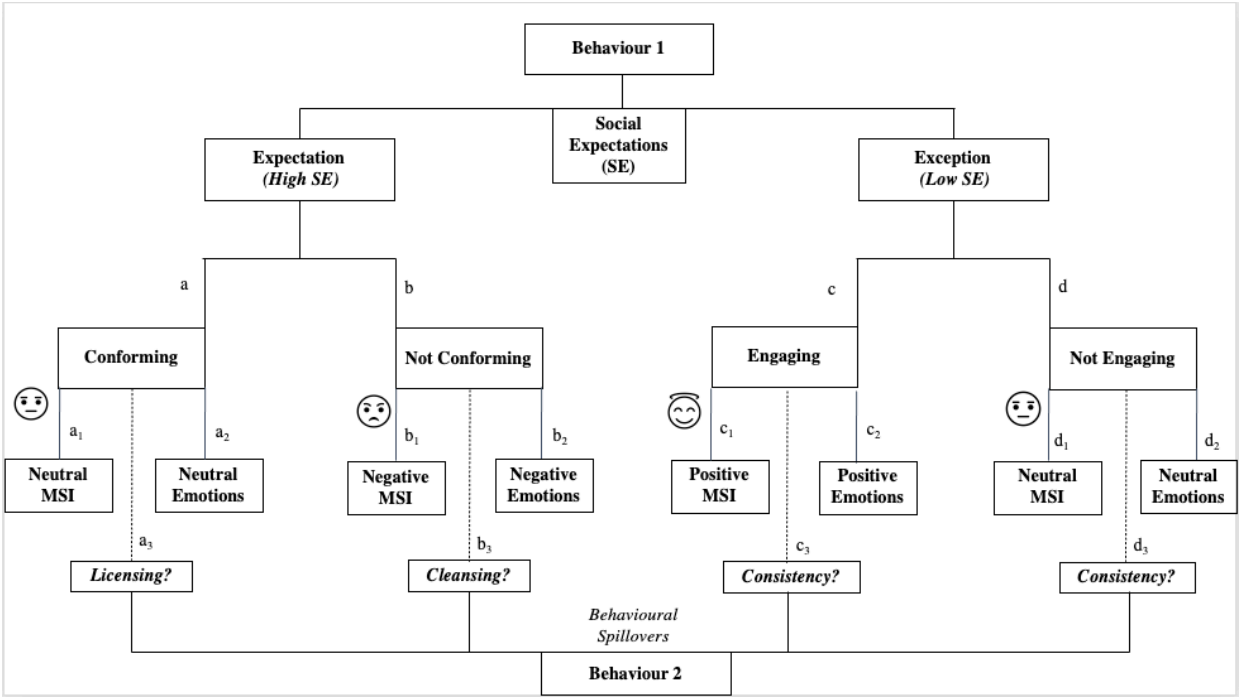
Table 4. Definition of Expectation and Exception

	Expectation	Exception
Injunctive (Normative expectation)	✓	✓
Descriptive (Empirical expectation)	✓	✗

We present the suggested theoretical framework in Figure 4 as a conceptual diagram. It differentiates between four conditions for Behaviour 1:

1- Conforming to the "Expectation": **Acting** according to what is socially approved and expected.
2- Not conforming to the "Expectation": **Not acting** according to what is socially approved and expected.
3- Engaging in the "Exception": **Acting** according to what is approved but not expected.
4- Not engaging in the "Exception": **Not acting** according to what is approved but not expected.

Figure 4. Conceptual Diagram of the Theoretical Framework: Effect of Social Expectations on MSI/ Affect and Possible Behavioural Spillovers. MSI refers to "Moral Self-Image", SE refers to" Social Expectation



The literature discusses that situational factors influence experienced MSI (Jordan et al., 2015). Also, situation-specific experiences influence powerful and anticipatory feelings of guilt, shame, embarrassment, and pride (Tangney et al., 2007). Shame, guilt, embarrassment, and pride work as an emotional, moral barometer, providing immediate and salient feedback on our social and moral acceptability (Tangney et al., 2007). Feelings of shame, guilt or embarrassment are experienced when we sin or transgress. However, feelings of pride and enthusiasm, or self-approval, are experienced when we "do the right thing" (Kroll and Egan 2004). This framework builds on these findings and zooms into social expectations on possible situational factors that influence MSI and moral affect.

In the framework, a positive behaviour (i.e., normative expectation/ injunctive norm) could have a high empirical expectation – i.e., "Expectation" or a low empirical expectation - i.e., Exception. The framework presents four main paths, and each path will lead to three sub-paths to measure the effect on 1) MSI, 2) affect, and 3) subsequent behaviour. We first present the main four paths:

6.1 Path (a) - Conforming to the "Expectation"

First, we suggest that when the behaviour at stake is the "Expectation", conforming with it does not increase MSI nor increase moral credit (for instance standing in line is expected and does not lead to positive affect). According to the literature, conforming to descriptive norms results from heuristic and cognitive shortcuts (Lindstrom et al., 2018). The act of conforming is often based on

the automatic system rather than engaging in rational deliberation (Ariely, 2009). Also, when behaviour is descriptive and injunctive (expected to be performed and approved by others), individuals tend to conform to avoid negative consequences or social sanctions (Herrman, 1971). Therefore, we propose to test that conforming to the "Expectation" leads to no change in MSI (path a_1) and affect (path a_2).

6.2 Path (b) - Not Conforming to the "Expectation"

Second, not behaving according to the "Expectation" signals social sanction and dis-approval from others (Herrman, 1971). Social norms are externally regulated by external and internal expectations (informal and formal sanctions) (Bicchieri, 2006; Cialdini et al., 1991; Reno et al., 1993). Therefore, deviation or non-conformity with the social norm or "Expectation" is socially sanctioned and creates inner pressure to comply with moral obligations and internal sense of values (Schwartz, 1977; Biel et al., 2011). We suggest that sanctions for not conforming to social norms are more severe than the reward for conforming to social norms. Moreover, transgressions and violations of social norms implicate negative self-evaluation and experience of negative moral affect such as guilt and shame (Tangney, 1991; Tangney and Dearing, 2002; Tangney et al., 2007; Sheikh and Janoff-Bulman, 2009). Not all sanctions operate externally. Internal sanctions Internalized norms will lead people to follow norms due to negative affect like guilt or positive ones like a desire to participate in constructive social behaviour (Cooter 1996). Milgram's line-intrusion experiment showed that participants who broke the queue experienced physical symptoms due to increased anxiety from breaking the social norm (Milgram et al., 1986). We propose to test that not conforming to the "Expectation" leads to decreased MSI (path b_1) and negative affect (path b_2).

6.3 Path (c) - Engaging in the "Exception"

Third, engaging in the "Exception" is not based on automatic process or social conformity because the behaviour is not descriptive. Engaging in such behaviour could be driven by personal norms and anchored around oneself and self-expectations (Schwartz, 1977). People have internal behavioural regulation processes based on the anticipated degree of guilt and pride (Ryan and Deci, 2000). For instance, there is a collective awareness that driving an electric car is environmentally conscious and has a collective benefit shared as a public good (injunctive). However, people are not expected to exchange their vehicles for a new electric one and do not expect others to buy them (non-descriptive). Hence, if one decides to opt for an electric car, one

must be driven by either personal norms or a desire for a warm glow and social desirability effect (Andreoni, 1990). We propose to test that engaging in the "Exception" leads to higher MSI (path c_1) and positive affect (path c_2).

6.4 Path (d) - Not Engaging in the "Exception"

Fourth, we predict that not engaging in the "Exception" does not impact MSI or one's perceived moral credit balance. As there is no expectation to engage in this behaviour, moral self-regard is unlikely to be involved. This explains why many collective action problems are complex to solve. Therefore, we propose to test that not engaging in the "Exception" leads to no change in MSI (path d_1) and affect (path d_2).

Hypothesis 1a: MSI is higher when engaging in the "Exception" (path c1) compared to conforming to the "Expectation" (path a1).

Hypothesis 1b: Positive Affect are higher when engaging in the "Exception" (path c_2) compared to conforming to the "Expectation" (path a_2).

Hypothesis 1c: MSI is lower when not conforming to the "Expectation" (path b_1) compared to not engaging in the "Exception" (path d_1).

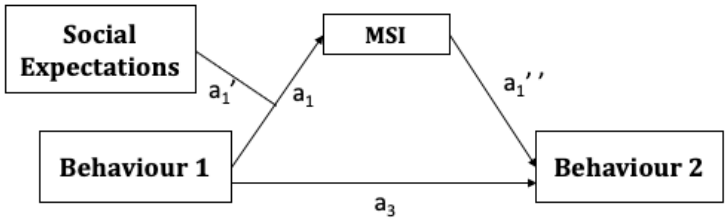
Hypothesis 1d: Positive Affect are lower when not conforming to the "Expectation" (path b_2) compared to not engaging in the "Exception" (path d_2).

The difference in experienced affect and MSI as a result of the intervention explains the effect of behaviour one on behaviour 2 (spillovers). The literature on behavioural spillovers (Dolan and Galizzi, 2015) presents positive and negative spillovers. Evidence for spillovers remains mixed (Nash et al., 2017; Galizzi and Whitmarsh, 2019). Many studies examined different moderator variables that try to explain when positive and negative spillovers occur (Thøgersen and Crompton, 2009; Truelove et al., 2014; Steinhorst et al., 2015). These include goals, values, self-identity, personal norms, skills and knowledge, and self-efficacy (Thøgersen, 2012). According to the norm activation model (Schwartz, 1977), personal norms are moral obligations. Also, according to the cognitive dissonance theory (Aronson, 1997), people behave consistently with their moral self-concept to maintain a positive emotional state. Thøgersen and Ölander (2003) found that when the pro-environmental personal norm is strong, it may promote spillover onto other behaviours. We therefore predict:

Hypothesis 2: Engaging in the "Exception" (path c3) leads to more positive spillovers (consistency) compared to engaging in the "Expectation" (path a3) (licensing), as long as the behaviour is not driven by social expectations.

The literature on behavioural spillovers, particularly moral licensing, presented the moral credit model with a hypothetical moral bank account where people accumulate credits when engaging in moral actions or offset immoral ones (Merritt et al., 2010; Miller and Effron, 2010). In this framework, we investigate moral licensing and test whether MSI mediates the effect of Behaviour 1 on Behaviour 2 (Figure 5) when MSI is moderated by social expectations (moderated mediation). In other words, how much does the increase in MSI due to social expectations when engaging in Behaviour 1 explain the negative spillover effect in Behaviour 2? Theoretically, this model focuses on one path where social expectations moderate the impact of Behaviour 1 on MSI and investigates whether moral licensing is more likely to occur due to the increase in MSI. Hence, if we feel more positive and moral after engaging in Behaviour 1 due to social expectations and norms, this increase in MSI will more likely lead us to transgress in Behaviour 2. We propose that when the change in MSI after engaging in Behaviour 1 is due to social expectations and norms rather than personal norms, then moral licensing is more likely to be experienced in Behaviour 2. The statistical diagram with the paths of the moderated mediation model is presented in Figure 5.

Figure 5: New statistical diagram of the moderated mediation model



Hypothesis 3: When social expectations moderate the effect of Behaviour 1 on MSI, MSI mediates the effect of behaviour 1 on behaviour 2. In other words, the negative spillover of behaviour 1 on behaviour 2 is indirectly explained by the increase in MSI (which is moderated by social expectations). Behaviour 1 (e.g., pro-environmental behaviour compared to a control condition or non-pro-environmental behaviour) leads to

an increase in MSI when people expect others to engage in the behaviour. In contrast, MSI, in turn, negatively predicts Behaviour 2.

7. Conclusion and Overview of the Present Research

The following research papers attempt to test the proposed hypotheses and model by employing four empirical studies. Study 1 explored the effect of expectation versus exception on moral self-image and affect and tested H1. Study 2 studied the different spillover effects and tested H2. It also investigated the underlying mechanism and tested a moderated mediated model related to H3. Study 3 replicates the moderated mediation model in another domain (pro-environmental domain and sustainable transport). Study 4 examines the suggested model on pro-environmental behaviours related to virtue signalling and pledging on possible behavioural spillovers. Table 5 summarises the main hypotheses, variables and statistical tests, and Table 6 presents the corresponding variables and conditions.

Table 5. Hypotheses, predictions, and statistical tests

Hypotheses	Dependent Variables	Explanation of Hypotheses	Statistical Test	Study
H1a	MSI: paths (a_1) and (c_1)	There is a significant difference between MSI in a_1 and c_1	t-test comparison	1,2
H1b	Affect: paths (a_2) and (c_2)	There is a significant difference between experienced affect in a_2 and c_2	t-test comparison	1,2
H1c	MSI: paths (b_1) and (d_1)	There is a significant difference between MSI in b_1 and d_1	t-test comparison	1,2
H1d	Affect: paths (b_2) and (d_2)	There is a significant difference between experienced affect in b_2 and d_2	t-test comparison	1,2
H2	Behaviour 2: paths (a_3) and (c_3)	There is a significant difference in Behaviour 2 across conditions.	Multiple regression	2,3,4
H3	MSI and behaviour paths (a_1) and (a_3)	Moderated mediation effect: SE moderate the effect of B1 on MSI, and MSI indirectly explains the effect of B1 on B2.	Moderated mediation model	2,3,4

Table 6. Conditions and key variables

Variable/ Condition	Study	Description
Conforming to "Expectation"	1,2	Ask participants to imagine conforming to the "Expectation."
Conforming to "Expectation"	1,2	Ask participants to imagine not conforming to the "Expectation."
Engaging in the "Exception"	1,2	Ask participants to imagine engaging in the "Exception."
Not Engaging in the "Exception"	1,2	Ask participants to imagine not engaging in the "Expectation."
Main dependent variables		

Affect	1,2, 3,4	Measure via 10 items (5 positive affect and 5 negative affect) answered on a 5-point Likert scale from 1 (very slightly to not at all) to 5 (Extremely).
Moral Self-Image (MSI)	1,2, 3,4	As per Jordan et al., 2015, the MSI scale is measured via 9 items answered on a 9-point Likert scale from 1 (much less *caring* than the person I want to be) to 9 (much more *caring* than the person I want to be).
Behaviour 2/ task 2	2, 3,4	Provide participants with an unpaid task to perform.
Moderators		
Social Expectations (SE)	1,2, 3,4	Average of the self-reported Empirical Expectation and Normative Expectation. Participants have a slider from 0 to 100

III. Paper 1

"Expectation" or "Exception":
Examining the Effect of Norms on Moral Self-Image and Affect

January 2022

Department of Psychological and Behavioural Science
London School of Economics and Political Science
Houghton Street, WC2A 2AE, London, UK
f.koaik@lse.ac.uk, d.krpan@lse.ac.uk and p.dolan@lse.ac.uk

1. Abstract

Does breaking the line in a context where queuing is the expectation bear similar self-perception and affect compared to breaking the line in a context where queueing is the exception? We present a theoretical framework that differentiates between behaviours that are the "Expectation" and others that are the "Exception". We tested this framework on a sample of participants from the UK by comparing standing in line (as the expectation) and getting an electric car (as the exception). The findings show that when participants wait in line (i.e., do conform to the expectation) and do not buy an electric car (i.e., do not engage in the exception), they feel neutral in terms of affect and their perceived moral self-image. However, when participants break the line (do not conform to the expectation), they think negatively about themselves and positively when buying an electric car (engage in the exception). This study shows that social expectations moderate the effect of engaging in positive behaviour on experienced affect and moral self-image.

Keywords: Social Norms, Affect, Moral Self-Image, Pro-Environmental Behaviour

2. Introduction

"Context matters" (Dolan, 2014). Does breaking the line in a context where queuing is the *expectation* bear similar self-perception and affect compared to breaking the line in a context where queueing is the *exception*? Also, does buying an electric car in a context where the latter is the social norm lead to similar self-evaluations and affect compared to purchasing an electric vehicle where the latter is the good exception? In this case, what matters in the context is social norms and expectations. How and why is it important?

When thinking of sustainable behavioural change, collective action problems appear in various policy domains (Ostrom, 2014), from citizens not paying taxes, public servants accepting bribes, households overconsuming energy and water, and citizens not being inclusive of gender, race, disability, and ethnicity. These challenges seem more manageable if everyone cooperates for social good. (Ostrom, 2014). Solving collective action problems and their underlying behavioural motives requires zooming in to understand the underlying impact of behaviour on moral affect. This helps zoom out and cascade behavioural change (leading to a snowballing effect to solve collective action problems).

One example of a collective action problem is environmental issues associated with a global carbon footprint (Hertwich and Peters, 2009). Several European countries started subsidising electric vehicles as an alternative fuel solution. The interest in electric cars has been growing in the face of climate change, the reduction of natural resources, and the increase in pollution in urban cities (Barth et al., 2016). Public debates on the personal costs and the benefits of purchasing an electric vehicle compared to a conventional vehicle are ongoing. However, looking beyond personal effects, people's attitudes towards opting for an electric vehicle are influenced by social context (Barth et al., 2016). Driving an electric car is generally approved by the public. However, it is still not widely adopted (Barth et al., 2016; Smith and Louis, 2009). Although the popularity of electric cars is increasing, the market ratio of electric vehicles is still low (Tu and Yang, 2019). It is more a good "Exception[3]".

[3] We define "**Exception**" to describe behaviours that i) have injunctive norms or normative expectations (they are approved of) and ii) are non-descriptive or do not have empirical expectations (i.e., we do not expect others to undertake them)

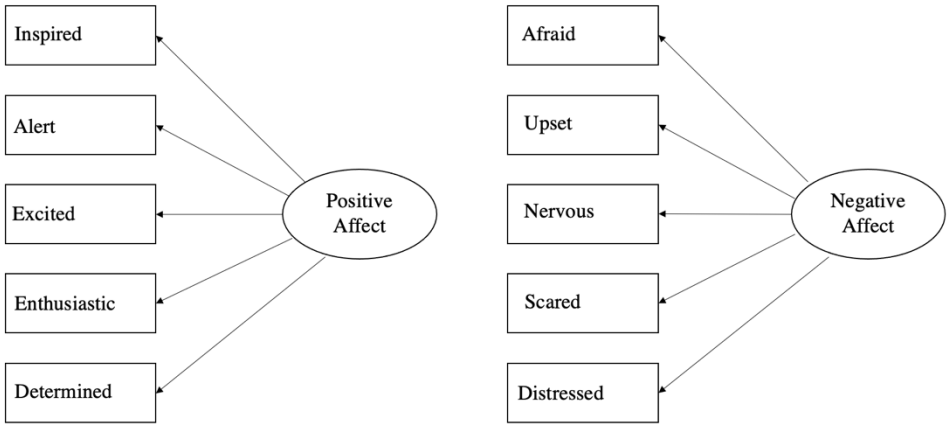
Another example of collective behaviour and coordinated practice that requires a substantive majority of individuals to engage in taking turns (MacCormick, 1998). “The phenomenon of lining up to wait is so commonplace that we rarely pause to reflect on it” (Fagundes, 2017). Even though no law governs it, people tend to automatically wait at the end of the line for their turn out of social and instinctive cooperation (Fagundes, 2017). “standing in line is normative” (MacCormick, 1998). It is more of a good “Expectation”.

The Effect of Moral Self-Image (MSI) and Affect

Affect

“Affect (the act of experiencing emotion) is a powerful force in decision-making. Emotional responses to words, images, and events can be rapid and automatic, so people can experience a behavioural reaction before they realise what they are reacting to” (Dolan et al., 2010). There is a strong argument in the literature that perceptions bring some affect of different types (Zajonc, 1980). Moreover, human behaviour is not only influenced by reactive and reflective responses but also affective reactions to anticipated consequences (Strack et al., 2016). In the pro-environment domain, empirical research explores how affect (both positive and negative affect) influence pro-environmental behaviour (Bissing-Olson et al., 2013; Ferguson and Branscombe, 2010; Bissing-Olson et al., 2016). More specifically, the literature focuses on the affect of pride and guilt since they guide moral and pro-social behaviour (Tangney et al., 2007; Tracy and Robins, 2007). People experience guilt and pride based on appraisals of their own behaviour and standards of right and wrong (Tangney et al., 2007; Tracy and Robins, 2007). They are more likely to feel proud after engaging in moral, positive, or valued behaviour and more likely to feel guilty and ashamed after engaging in immoral or negative behaviour. Mallett et al. (2013) showed that exposing people to their carbon footprint increased feelings of guilt. Other empirical experiments also measured the affective reaction of guilt and pride after exposing individuals to messages about their responsibility for environmental damage (Harth et al., 2013). This research explores how individuals react and feel after engaging in descriptive versus non-descriptive norms. The Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988) is one of the most widely used scales to measure affect at a specific point in time. The PANAS measures two mood dimensions: ten positive effects and the other ten negative effects. A shorter scale was created by Kercher (1992) and further validated by (Mackinnon et al., 1999) (Figure 6).

Figure 6. Short PANAS as a measure of positive and negative affect (Mackinnon et al., 1999)



Moral Self-Image

Moral Self-Image (MSI) captures a temporary perception of one's moral self that fluctuates based on recent feedback, social comparison, and other situational factors activated in one's working memory (Jordan et al., 2015; Monin and Jordan, 2009). Based on context, social identity, and situational factors, people evaluate their MSI and give themselves positive or negative labels (Kernis and Goldman, 2003). Unlike self-esteem and moral identity (which are more stable and less likely to be affected by temporal changes or actions), MSI resides in a "working self-concept" (Kernis and Johnson, 1990). Therefore, the literature discusses that MSI changes based on different events in the social world; it is also influenced by feedback from the environment. Jordan et al. (2015) developed an MSI scale as an exploratory process to measure MSI on the following traits: caring, compassionate, helpful, hard-working, friendly, fair, generous, honest, and kind. MSI is used as a mechanism to describe the malleable moral construct of individuals and helps explain individuals' moral inconsistencies (Monin and Miller, 2001; Zhong and Liljenquist, 2006; Sachdeva et al., 2009; Jordan et al., 2011).

The Role of Social Expectations

Social expectations are beliefs about other people's beliefs and behaviours (Bicchieri et al., 2014). Empirical expectations are beliefs about the behaviour of others (Bicchieri, 2006), also known as descriptive norms (Cialdini et al., 1990; Cialdini et al., 1991). Normative expectations are beliefs about other people's or second-order personal normative beliefs (Bicchieri, 2006), also known as

injunctive norms (Cialdini et al., 1990; Cialdini et al., 1991). There are two types of norms, and empirical research showed that each one has a distinct and independent impact on behaviour and intention by having separate sources of motivation (Cialdini et al., 1990; Kallgren et al., 2000; Manning, 2009; Reno et al., 1993; Rivis and Sheeran, 2003). Also, the two norms might not always be aligned (Smith and Louis, 2008).

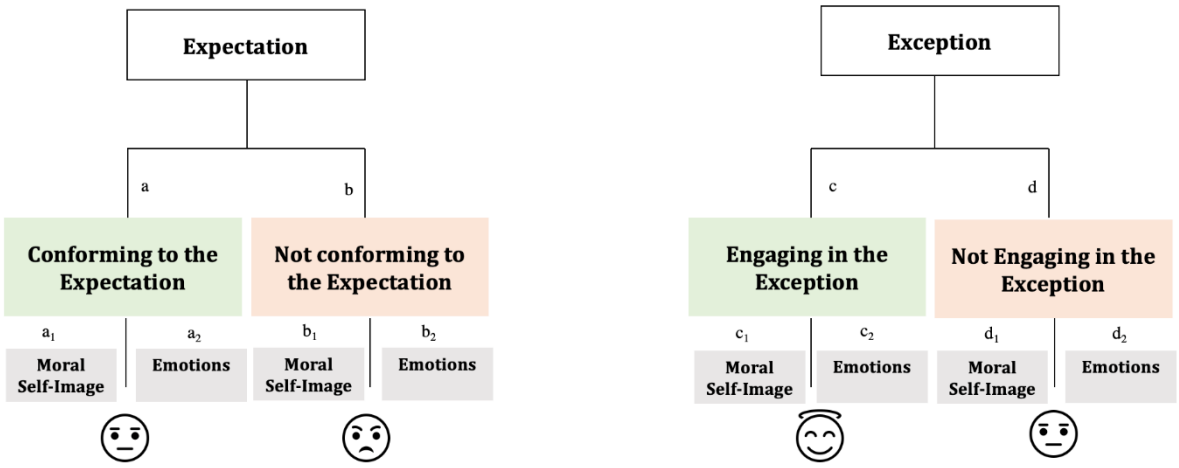
To date, little research has been done on how different social norms (descriptive or injunctive) influence moral self-appraisals and experienced affect and affect. Also, the question of how the conflict between injunctive and descriptive norms influences behaviour and emotion is overlooked (see Smith and Louis, 2008).

Present Research

First, we refer to the "**Expectation**" to describe behaviours that i) have injunctive norms or normative expectations and ii) have descriptive norms or empirical expectations (i.e., we expect others to undertake them, and they are approved). Second, we refer to the "**Exception**" to describe behaviours that i) have injunctive norms or normative expectations (they are approved of) and ii) are non-descriptive or do not have empirical expectations (i.e., we do not expect others to undertake them)

This paper runs an empirical experiment to test how (not) engaging in "Expectation" versus "Exception" influences moral self-image and moral affect (Figure 7). We predict that engaging in an injunctive norm (*Exception*) leads to more positive affect and moral self-image compared to engaging in an injunctive and descriptive norm (*Expectation*). Moreover, we predict that not engaging in injunctive and descriptive norms (*Expectation)* leads to more negative affect and moral self-image compared to engaging in injunctive norms (*Exception*).

Figure 7. Expectation versus Exception and Effect on MSI and Affect



3. Method

In this study, we employed an empirical method (2x2) to experimentally manipulate engaging versus not in "Expectation" or "Exception" behaviour and examined the impact on experienced affect and reported moral self-image. We aim to test the new suggested theoretical framework (see theoretical chapter – Figure 4). We selected the "Exception" behaviour to be "buying an electric car" and referred to it as a pro-environmental behaviour, and the "Expectation" behaviour to be "waiting in line" and referred to it as a rule of law behaviour. We predicted that the two behaviours have opposing social expectations in the UK. First, waiting in line is generally more descriptive than buying an electric car[4]. Second, both behaviours have a collective benefit, which means engaging in them increases the collective interest rather than just individual interest. Third, both behaviours are socially approved.

3.1 Participants and Procedure

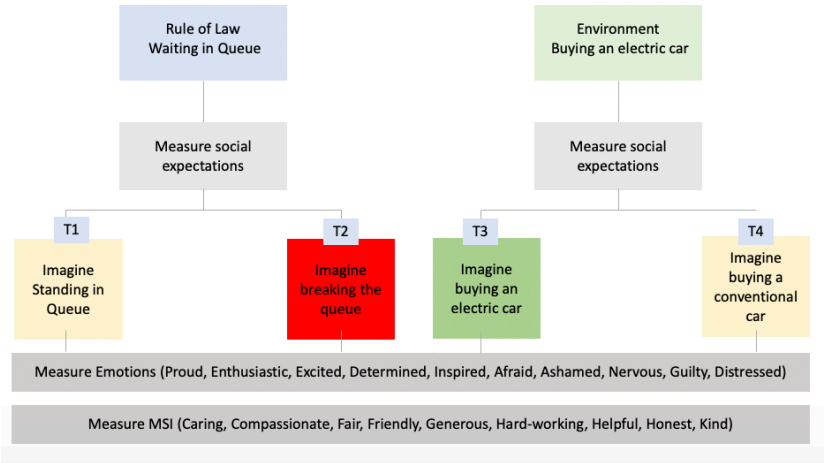
The study was designed on Qualtrics, and participants were recruited online on M-Turk. 415 participants, age (M=38.71, SD=12.33), 151 females and 264 males were recruited on M-Turk. The majority of participants in the study were from the USA (53.25%) and India (36.39%). Other

[4] At the time when the experiment was run in 2021, as buying an electric vehicle has been increasingly common in the UK in the past 2 years.

nationalities represented included Canada (0.96%), Brazil (2.65%), and several others making up the remaining 6.75%. This distribution reflects the diverse background of the sample population.

Participants completed the consent form, after which they answered the demographic questions. Next, they were randomly allocated to either the rule of law or the environmental condition. In the rule of law condition, they read a vignette scenario on the journey of grabbing a coffee on a Monday morning at a favourite coffee shop where a few other customers arrived simultaneously (representing the "Expectation"). The second group received a vignette scenario on buying a new car (representing the "Exception"). After that, the moderator was measured (see the *Measures* section below) to measure their social expectations of the given behaviour. Participants were asked (depending on the condition): 1) the likelihood of them buying an electric car/ or waiting in line for their turn, 2) their perceived empirical expectation (what they expect others to do), 3) their personal normative belief (what they think is the right thing to do), 4) their normative expectation (what they expect other to think is the right thing to do). Next, each group was randomised into two conditions, making them a total of four. The first condition asked participants to imagine standing in the queue. The second treatment asked them to imagine breaking the line. The third treatment asked participants to imagine opting for an electric car. The fourth treatment asked them to imagine opting for a conventional vehicle (Figure 8). After exposure to the corresponding manipulation, participants were asked questions assessing the dependent variables - affect and moral self-image - (see the *Measures* section below). The data for Study 1 were collected on August 15, 2021. This study complies with the ethics policy and procedures of the London School of Economics (LSE) and has been approved by its Research Ethics Committee.

Figure 8. Design of the experiment



3.2 Measures

Participants were randomised into four conditions; we refer to the four conditions (i.e., independent variables) as *Break* (1) (Breaking the line), *Wait* (2) (Waiting in line), *Electric* (3) (Buying an electric car), and *Conventional* (4) (Buying a conventional car).

Both outcome variables are continuous. The first dependent variable (*affect*) was measured via ten items answered on a 5-point Likert scale from 1 (very slightly or not at all) to 5 (Extremely), which required participants to indicate to what extent they felt the different affect when imagining engaging in the presented behaviour. The items include positive affect (proud, inspired, excited, enthusiastic, and determined) and negative affect (afraid, ashamed, nervous, guilty, and distressed)[5]. The average score reported five positive and five negative affect. The final variable is the difference between the average positive and average negative affect. The second dependent variable, Moral-Self Image (MSI) as per Jordan et al. (2015) scale, was measured via nine items (e.g., Compared to the *caring* person I want to be, I am:) answered on a 9-point Likert scale from 1 (Much less *caring* than the person I want to be) to 9 (Much more *caring* than the person I want to be) (more details on the scale is in Appendix A). We took the average reported score on each of the nine dimensions to indicate their moral self-regard.

The social expectation questions were:

1- **Self-reported behaviour:** First, they were asked about the likelihood of engaging in the behaviour (standing in line or opting for an electric car) on a 5-point Likert scale of 1 (Extremely unlikely) to 5 (Extremely Likely).

2- **Empirical Expectation (EE):** To get an understanding of the empirical expectations and quantify them, we asked participants the following:
"If we asked citizens in your social network, *Do you have an electric car?* In your view, what percentage answer YES and therefore have an electric car?"

[5] In the current study, we replaced three items ; alert, upset and scared from the short PANAS scale (Watson et al., (1988) with the following items proud, ashamed and guilty as they are relevant to the study. The items exhibited high inter-item correlations and significant overlap with other items. Correlation analyses revealed that these items demonstrated strong correlations (e.g., r=0.75, r=0.82). Furthermore, Cohen's d values were notably high (above 0.9), suggesting that these items contributed little unique variance to the overall measurement of affect.

To answer this question, participants were given a slider ranging from 0 to 100 (the percentage of citizens who believe that buying an electric car or *waiting in line* is good).

3- **Personal Normative Belief (PNB):** To understand whether participants approve or disapprove of waiting in line or choosing an electric car, we asked them about their opinion of engaging in the behaviour on a 5-point Likert scale of 1 (Extremely bad) to 5 (Extremely good).

4- **Normative Expectation (NE):** Using the same structure as EE, each participant is asked about their expectations of what others think is the right thing to do (when it comes to standing in line or choosing an electric car).

"If we asked citizens in your social network the following: *Do you think buying an electric car is the right thing to do?* In your view, what percentage of your network believes it is the right thing to do?

To answer this question, participants were given a slider ranging from 0 to 100 (the percentage of citizens who believe that buying an electric car or *waiting in line* is good).

3.3 Sample

Based on a-priori power analysis in G*Power 3.1 (Faul et al., 2007), a minimum total sample of N=400 (100 per condition) will be needed for a 95% probability of detecting a medium-size effect ($d = 0.5$) given a conventional level of significance ($\alpha = 0.05$).

3.4 Hypotheses

Hypothesis 1a: MSI when engaging in the "Exception" (path c_1) is higher compared to conforming to the "Expectation" (path a_1).

Hypothesis 1b: MSI, when not conforming to the "Expectation" (path b_1), is lower compared to not engaging in the "Exception" (path d_1).

Hypothesis 2a: Affect when engaging in the "Exception" (path c_2) are more positive compared to conforming to the "Expectation" (path a_2).

Hypothesis 2b: Affect when not conforming to the "Expectation" (path b_2) are lower compared to not engaging in the "Exception" (path d_2).

We also aim to run an exploratory analysis. First, we will analyse how Expectation (as a binary variable) moderates the effect of engaging versus not engaging in the behaviour on MSI and affect.

Second, we will run an interaction analysis on how social expectations (as a continuous variable) moderate the effect of engaging versus not engaging in the behaviour on MSI and affect.

4. Results

For data analysis, we use STATA/ SE 15.1.

After the social expectation questions, they were randomised into four conditions, "Queue" (92), "break the queue" (90), "Electric" (135) and "Conventional" (98). Table 7 below summarises each treatment condition's mean and standard deviation for the dependent variables MSI and affect.

Table 7. Mean (M) and Standard Deviation (SD) for the MSI and Affect

	MSI		Affect	
Condition (N=sample size)	M	SD	M	SD
Wait (92)	6.118	1.411	0.76	0.904
Break (90)	5.128	2.383	-0.515	1.214
Electric (135)	6.207	1.265	1.557	1.123
Conventional (98)	5.512	1.336	0.779	1.205

Note. DV refers to the dependent variable. Number in parentheses next to DVs indicate the possible range of values for each DV.

Manipulation check

To test whether there is a significant difference between the empirical expectations of the environmental condition and the rule of law. We performed an independent samples t-test (two-tailed). Participants in the *Environment* condition (M=25.94, SD=28.29) had lower empirical expectations compared to participants in the *Rule of law* condition (M=70.63; SD = 22.13) $t(415) = 17.52$, $p<0.01$, $d=1.73$, 95% CI [1.50, 1.95]. This indicates that people expect others to stand in the queue much more than buying an electric car.

Second, to test whether there is a difference in the normative expectation between the environment and the rule of law conditions, we performed an independent samples t-test (two-tailed). Participants in the *Environment* condition (M=57.69, SD=24.32) had lower normative expectations compared to participants in the Rule of Law condition (M=71.51, SD=22.08) $t(415) = 5.97$, $p<0.01$, $d=0.59$, 95% CI [0.32, 0.78]. This indicates that participants expect others to approve of standing in the queue more than buying an electric car.

Third, we also performed an independent samples t-test (two-tailed) to test whether there is a difference in the PNB between the environment and the rule of law conditions. Participants in the

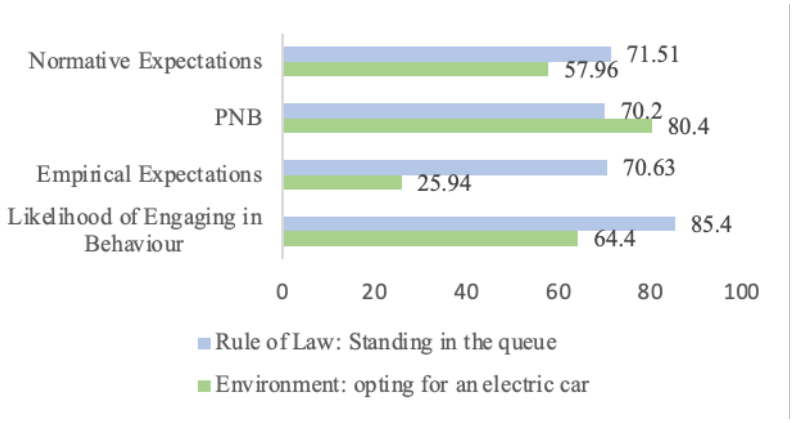
Environment condition (M=4.02, SD=0.88) had higher PNB compared to participants in the Rule of Law condition (M=3.51, SD=1.01) t (415) = -5.46, p<0.01, d=-.54, 95% CI [-0.73, -0.34]. This indicates that participants approve of buying an electric car more than standing in the queue. Comparing these results with the normative expectation, participants approve purchasing an electric vehicle more than waiting in line but do not expect others to believe the same.

Fourth, to test whether there is a difference in the likelihood of engaging in the behaviour between the environment and the rule of law condition, we performed an independent samples t-test (two-tailed). Participants in the Environment condition (M=3.22, SD=1.20) had a lower likelihood of engaging in the behaviour compared to the Rule of Law condition (M=4.27, SD=0.92) t (415) = 9.7, p<0.01, d=0.96, 96% CI [0.75, 1.16]. Although participants approve of the electric car, they are still more likely to stand in the queue, and the difference is highly significant.
These results indicate a significant difference in the social expectations between the two behaviours (Table 8 and Figure 9 below present the results). We can be confident that standing in the queue is the "Expectation" and opting for an electric car is the "Exception".

Table 8. Social expectations across the environment and the rule of law conditions

	Environment		**Rule of Law**	
	Opting for an electric car		*Standing in the queue*	
	Mean	**Standard Error**	**Mean**	**Standard Error**
Likelihood of Engaging in Behaviour	3.22	(1.20)	4.27	(0.92)
Empirical Expectations	25.94	(28.29)	70.63	(22.13)
PNB	4.02	(0.88)	3.51	(1.01)
Normative Expectations	57.96	(24.32)	71.51	(22.08)

Figure 9. Social expectations across the environment and the rule of law conditions



Hypotheses 1a and 1b

A one-way ANOVA indicates whether we have significant differences across the four groups for the first dependent variable, *MSI*. A statistically significant difference exists in the mean MSI F (415) = 10.25, p<0.01. Regression models testing the impact of the different conditions on MSI are presented in Table 9, whereas the means and 95% CIs for the variables are reported. Overall, the analyses showed that there is no significant difference between *Wait (M=6.11, SE=0.16)* versus *Electric (M=6.20, SE=0.13)* (b=0.08, SE=0.21, t=0.41, p=0.684), which is inconsistent with Hypothesis 1a (Table 9 Model 1). Also, there is no significant difference between *Break* (M=5.12, SE=0.17) and *Conventional (M=5.5*1, SE=0.16) (b=0.38, SE=0.23, t=1.63, p=0.105), which is also inconsistent with hypothesis 1b (Table 9 Model 2). Figure 10 presents the mean MSI across the four conditions; there are no significant differences between Wait and Electric and Break and Conventional.

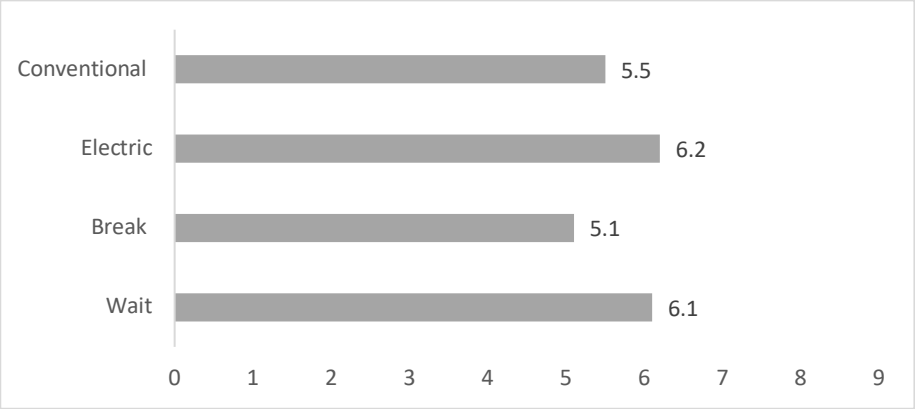
Table 9. The Effects of Wait (vs. Electric) and Break (vs. Conventional) on MSI

Variable	b	SE b	95% CI	T	p
DV: Self-reported Moral Self-Image (MSI)					
Model 1: Behaviour 1 - Wait (Baseline) vs. Electric					
Constant	6.118	0.168	[5.786, 6.449]	36.29	<0.001
Break	-0.989	0.239	[-1.461, -0.518]	-4.13	<0.001
Electric	0.089	0.218	[-0.340, 0.518]	0.41	0.684
Conventional	-0.605	0.234	[-1.067, -0.144]	-2.58	0.01
DV: Self-reported Moral Self-Image (MSI)					
Model 2: Behaviour 1 - Break (Baseline) vs. Conventional					
Constant	5.128	0.17	[4.793, 5.463]	30.09	<0.001
Wait	0.989	0.239	[0.518, 1.461]	4.13	<0.001

Electric	1.079	0.22	[0.646, 1.511]	4.9	<0.001
Conventional	0.384	0.236	[-0.079, 0.848]	1.63	0.105

Note. For models 1 and 2, $R^2 = 0.069$. Given that the study has 4 conditions, each regression model contains three dummy variables. However, key analyses testing the hypotheses are highlighted in grey. DV are dependent variable.

Figure 10. Average MSI in each condition



Hypotheses H2a and H2b

A one-way ANOVA indicates whether there are significant differences across the four groups for the first dependent variable, affect. There is a statistically significant difference in the mean emotion F (415) = 61.63, p<0.01. Regression models testing the impact of the different conditions on *Affect* are presented in Table 10, whereas the means and 95% CIs for the variables are reported. Overall, the analyses showed that there is a significant difference between *Wait (M=0.76, SE=0.904)* versus *Electric (M=1.557, SE=1.123)* (b=0.796, SE=0.162, t=5.26, p<0.001), as predicted by Hypothesis 2a (Table 10. Model 1). Therefore, participants who imagined buying an electric car experienced more positive affect compared to those who imagined waiting in line. Also, there is a significant difference between *Break* (M=-0.515, SE=1.214) and *Conventional (M=0.779*, SE=1.205) (b=1.295, SE=0.163, t=7.92, p<0.001), as predicted by hypothesis 2b (Table 10. Model 2). So, participants experienced more negative affect when they didn't stand in line compared to when they didn't buy an electric car. Figure 11 presents the mean MSI across the four conditions. Moreover, an additional interesting analysis shows no significant difference in the average affect between participants in the *Conventional* and the *Wait* conditions (0.18 +- 0.16, t =

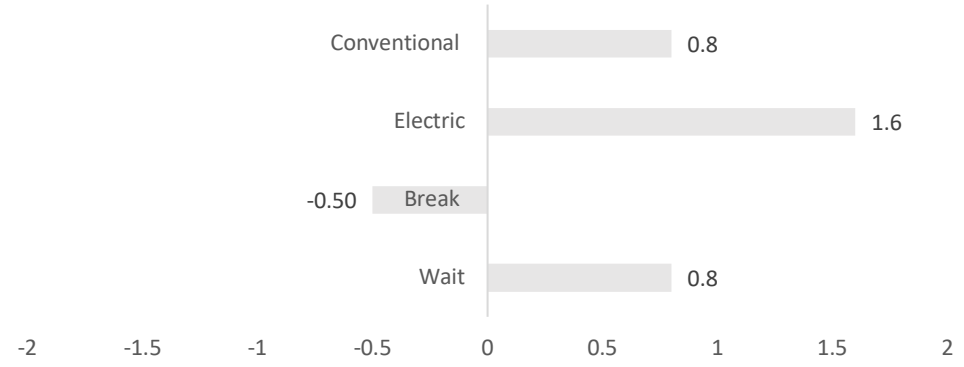
0.12, p>1.00). This indicates that conforming to the expectation and not engaging in the exception leads to no change in affect (as predicted in the theoretical framework).

Table 10. The Effects of Wait (vs. Electric) and Break (vs. Conventional) on Affect

Variable	b	SE b	95% CI	T	p
DV: Self-reported Affect					
Model 1: Behaviour 1 - Wait (Baseline) vs. Electric					
Constant	0.76	0.116	[0.531, 0.990]	6.51	<0.001
Break	-1.276	0.151	[-1.603, -0.949]	-7.68	<0.001
Electric	0.796	0.162	[0.498, 1.093]	5.26	<0.001
Conventional	0.018	0.166	[-0.301, 0.338]	0.12	0.908
Model 2: Behaviour 1 - Break (Baseline) vs. Conventional					
Constant	-0.515	0.118	[-0.747, -0.283]	-4.36	<0.001
Wait	1.276	0.166	[0.949, 1.603]	7.68	<0.001
Electric	2.072	0.152	[1.772, 2.372]	13.59	<0.001
Conventional	1.295	0.163	[0.973, 1.616]	7.92	<0.001

Note. For models 3 and 4, $R^2 = 0.3103$. Given that the study has 4 conditions, each regression model contains three dummy variables. However, key analyses testing the hypotheses are highlighted in grey. DV refers to the dependent variable.

Figure 11. Average affect in each condition



Exploratory Analyses – Interaction Effects

Next, we explore whether social expectation (SE) (continuous variable as the average of empirical and normative expectations) moderates the effect of each condition on MSI. Table 11 presents the regression with the interaction effects and the *Wait* condition as a baseline. SE positively moderates the impact of buying an electric car on MSI (b=3.936, SE=0.897, t=4.38 p<0.001). The higher participants expected others to opt for an electric vehicle and approve of it, the more moral they perceived themselves when imagining buying one.

Table 11. The Effects of Conditions x Social Expectation Interaction on MSI

Variable	**b**	**SE b**	**95% CI**	**t**	**p**
Model 1: DV = Moral Self-Image (MSI)					
Constant	6.438	0.488	[5.477, 7.399]	13.17	<0.001
Break	0.841	0.909	[0.276, 1.113]	0.92	0.356
Electric	-1.722	0.563	[-0.799, 0.124]	-3.06	0.002
Conventional	-2.063	0.586	[-0.380, 0.877]	-3.52	<0.001
SE	-0.476	0.688	[-1.83, 0.878]	-0.69	0.49
Break x SE	-2.388	1.213	[-4.77, -0.003]	-1.97	0.05
Electric x SE	3.936	0.897	[2.171, 5.701]	4.38	<0.001
Conventional x SE	3.381	0.991	[1.369, 5.267]	3.35	0.001

Note. For model 1, $R^2 = 0.1904$. Given that the study has 4 conditions, each regression model contains three dummy variables. However, key analyses testing the hypotheses are highlighted in grey. DV refers to the dependent variable.

5. Discussion

Overall, the findings of this study supported most of the predictions in the framework. Hypotheses 2a and 2b were confirmed, as we showed that behaviours with different social expectations lead to different self-reported affect. In other words, we showed that social expectations influence how we feel about our actions. Behaving in line with the "Exception" leads to positive affect compared to conforming to the "Expectation". Not conforming to the "Expectation" leads to more negative affect and moral self-regard compared to not engaging in the "Exception." However, the results did not confirm hypotheses 1a and 1b. MSI did not change between expectation and exception. These results indicate that standing in the queue gives people moral credit rather than no change in moral self-regard.

In this study, we showed that participants who imagined breaking the line experienced more negative affect compared to participants who imagined buying or not buying an electric car. At the same token, participants who imagined purchasing an electric vehicle experienced more positive affect than those who imagined waiting in line. The positive affect measured are alert, enthusiastic, determination, inspired and excited, and the negative affect are distress, upset, afraid, nervousness, and scared. **In this study, we explored a new dimension, where feelings of fear, nervousness, and distress aggravate more after transgression with a behaviour that is normative and socially expected.**

Moreover, the experiment showed that positive affect intensify when doing the exception or what is socially approved but not normative. For instance, if recycling is the norm in a neighbourhood, then people who recycle will not experience increased positive affect. However, if recycling is the exception in a neighbourhood, people who recycle will feel more positive affect, such as excitement and enthusiasm.

Another interesting insight is that participants in the *Wait* and the *Conventional* conditions experienced the same average affect. This is in line with the predicted theoretical model, and it presents relevant policy implications. **Doing the "Expectation" and not doing the "Exception" led to the same experienced affect**. Many environmental problems require simple behavioural solutions or actions to be adopted by individuals. Although, in many instances, people have the intention to adopt pro-environmental behaviours, these intentions fail to translate into actions, and this is due to various reasons (Steg and Vlek, 2009; Evans et al., 2013). Research showed that experienced affect induced in one context could affect individuals' decisions in different contexts (Andrade and Ariely, 2009). Therefore, experiencing positive affect when engaging in the exception can affect subsequent decisions, i.e., behavioural spillovers (Dolan and Galizzi, 2015, Galizzi and Whitmarsh, 2019) . Experiencing negative affect when transgressing and violating what is socially expected can influence decisions in different contexts. Also, in the context of collective action problems, not experiencing negative affect when not engaging in pro-environmental behaviour has implications.

6. Limitations

Some limitations are considered when interpreting the results, particularly when generalising the outcomes. First, the conditions elicited participants to imagine engaging or not engaging in the given behaviour; therefore, it is a hypothetical scenario. However, one point to raise is that the chosen behaviours are relatable, and participants can easily imagine the given scenario. Another limitation is self-reported affect and MSI; again, participants could act out of social desirability when reporting their affect (especially those related to buying an electric car). This behaviour is attractive and recommended; they might report higher to increase social desirability. In future studies, we aim to test the hypotheses with actual behaviours that participants either decide to engage in.

7. Conclusion

So far, no research looked at how social expectations and norms influence how individuals perceive their moral self-regard and how they feel about themselves when engaging or not engaging in a behaviour. In this study, we ran a simple experimental design to test whether social expectations influence the effect of conforming versus not on MSI and affect. The results were significant for the affect experienced. They verified the predictions that conforming to the "Expectation" leads to more positive affect and MSI compared to engaging in the "Exception", and not conforming to the "Expectation" leads to more negative affect and MSI compared to not engaging in the "Exception." Nevertheless, MSI did not differ between the Expectation and the Exception. One interpretation of these results could be that pro-environmental behaviours (electric cars, in this case) do not portray morality, whereas taking turns is closely related to morality. Still, having an electric vehicle brings a lot of positive affect, including enthusiasm, determination, and inspiration. In addition, social expectations moderated the effect of buying an electric car on MSI. This explains that despite purchasing an electric car is still not a descriptive norm, but social expectations positively moderate the effect of engaging in this behaviour on moral self-image.
Further research is needed to show the theoretical framework and understand the implications of this interaction between social expectations and MSI/ affect on behavioural spillovers. Whilst this experiment showed that social expectations lead to a significant difference in experienced affect, future research will build on these insights to better understand how these differences translate into different behavioural spillovers. The results contribute to the ongoing debate in the literature on when and why people experience different behavioural spillovers, i.e., consistency versus balancing.

8. References

Barth, M., Jugert, P., & Fritsche, I. (2016). Still underdetected – Social norms and collective efficacy predict the acceptance of electric vehicles in Germany. Transportation Research Part F: Traffic Psychology and Behaviour, 37, 64–77. https://doi.org/10.1016/j.trf.2015.11.011

Bicchieri, C., Lindemans, J. W., & Jiang, T. (2014). A structured approach to a diagnostic of collective practices. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.01418

Bissing-Olson, M. J., Fielding, K. S., & Iyer, A. (2016). Experiences of pride, not guilt, predict pro-environmental behaviour when pro-environmental descriptive norms are more positive. Journal of Environmental Psychology, 45, 145–153. https://doi.org/10.1016/j.jenvp.2016.01.001

Bissing-Olson, M. J., Iyer, A., Fielding, K. S., & Zacher, H. (2013). Relationships between daily affect and pro-environmental behavior at work: The moderating role of pro-environmental attitude: DAILY AFFECT AND PRO-ENVIRONMENTAL BEHAVIOR. Journal of Organizational Behavior, 34(2), 156–175. https://doi.org/10.1002/job.1788

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. In M. P. Zanna (Ed.), Advances in Experimental Social Psychology (Vol. 24, pp. 201–234). Academic Press. https://doi.org/10.1016/S0065-2601(08)60330-5

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. Journal of Personality and Social Psychology, 58(6), 1015. https://doi.org/10.1037/0022-3514.58.6.1015

Dolan, P. (2014). Happiness by Design: Finding Pleasure and Purpose in Everyday Life. Penguin UK.

Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). MINDSPACE: Influencing behaviour for public policy [Monograph]. Institute of Government. http://www.instituteforgovernment.org.uk/publications/

Evans, B. E., Greaves-Lord, K., Euser, A. S., Tulen, J. H. M., Franken, I. H. A., & Huizink, A. C. (2013). Determinants of physiological and perceived physiological stress reactivity in children and adolescents. PLoS ONE, 8(4). https://doi.org/10.1371/journal.pone.0061724

Fagundes, D. (2017). The Social Norms of Waiting in Line. Law & Social Inquiry, 42(4), 1179–1207. https://doi.org/10.1111/lsi.12256

Ferguson, M. A., & Branscombe, N. R. (2010). Collective guilt mediates the effect of beliefs about global warming on willingness to engage in mitigation behaviour. Journal of Environmental Psychology, 30(2), 135–142. https://doi.org/10.1016/j.jenvp.2009.11.010

Galizzi, M. M., & Whitmarsh, L. (2019). How to Measure Behavioral Spillovers: A Methodological Review and Checklist. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/article/10.3389/fpsyg.2019.00342

Harth, N. S., Leach, C. W., & Kessler, T. (2013). Guilt, anger, and pride about in-group environmental behaviour: Different affect predict distinct intentions. Journal of Environmental Psychology, 34, 18–26. https://doi.org/10.1016/j.jenvp.2012.12.005

Hertwich, E. G., & Peters, G. P. (2009). Carbon Footprint of Nations: A Global, Trade-Linked Analysis. Environmental Science & Technology, 43(16), 6414–6420. https://doi.org/10.1021/es803496a

Jordan, A. H., Monin, B., Dweck, C. S., Lovett, B. J., John, O. P., & Gross, J. J. (2011). Misery Has More Company Than People Think: Underestimating the Prevalence of Others' Negative Affect. Personality and Social Psychology Bulletin, 37(1), 120–135. https://doi.org/10.1177/0146167210390822

Jordan, J., Leliveld, M. C., & Tenbrunsel, A. E. (2015). The Moral Self-Image Scale: Measuring and Understanding the Malleability of the Moral Self. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.01878

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the Moral Self: The Effects of Recalling Past Moral Actions on Future Moral Behavior. Personality and Social Psychology Bulletin, 37(5), 701–713. https://doi.org/10.1177/0146167211400208

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A Focus Theory of Normative Conduct: When Norms Do and Do not Affect Behavior. Personality and Social Psychology Bulletin, 26(8), 1002–1012. https://doi.org/10.1177/01461672002610009

Kercher, K. (1992). Assessing Subjective Well-Being in the Old-Old: The PANAS as a Measure of Orthogonal Dimensions of Positive and Negative Affect. Research on Aging, 14(2), 131–168. https://doi.org/10.1177/0164027592142001

Kernis, M. H., & Goldman, B. M. (2003). Stability and variability in self-concept and self-esteem. In Handbook of self and identity (pp. 106–127). The Guilford Press.

Kernis, M. H., & Johnson, E. K. (1990). Current and typical self-appraisals: Differential responsiveness to evaluative feedback and implications for affect. Journal of Research in Personality, 24(2), 241–257. https://doi.org/10.1016/0092-6566(90)90019-3

Kroll, J., & Egan, E. (2004). Psychiatry, Moral Worry, and the Moral Affect. Journal of Psychiatric Practice, 10(6), 352–360. https://doi.org/10.1097/00131746-200411000-00003

MacCormick, N. (1998). Norms, Institutions, and Institutional Facts. Law and Philosophy, 17(3), 301–345. https://doi.org/10.2307/3504883

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the Positive and Negative Affect Schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. Personality and Individual Differences, 27(3), 405–416. https://doi.org/10.1016/S0191-8869(98)00251-7

Mallett, R. K., Melchiori, K. J., & Strickroth, T. (2013). Self-Confrontation via a Carbon Footprint Calculator Increases Guilt and Support for a Proenvironmental Group. Ecopsychology, 5(1), 9–16. https://doi.org/10.1089/eco.2012.0067

Monin, B., & Jordan, A. H. (2009). The dynamic moral self: A social psychological perspective. In Personality, identity, and character: Explorations in moral psychology (pp. 341–354). Cambridge University Press. https://doi.org/10.1017/CBO9780511627125.016

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. Journal of Personality and Social Psychology, 81(1), 33–43. https://doi.org/10.1037/0022-3514.81.1.33

Nash, N., & Whitmarsh, L. (2023). *One thing leads to another? Pro-environmental behavioural spillover* (pp. 63–77). https://doi.org/10.4337/9781800882133.00011

Ostrom, E., & Ostrom, V. (2014). Choice, Rules and Collective Action: The Ostrom's on the Study of Institutions and Governance. ECPR Press.

Reno, R., Cialdini, R., & Kallgren, C. (1993). The Transsituational Influence of Social Norms. Journal of Personality and Social Psychology, 64, 104–112. https://doi.org/10.1037/0022-3514.64.1.104

Rivis, A., & Sheeran, P. (2003). Descriptive Norms as an Additional Predictor in the Theory of Planned Behaviour: A Meta-Analysis. Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues, 22(3), 218–233. https://doi.org/10.1007/s12144-003-1018-2

Sachdeva, S., Iliev, R., & Medin, D. L. (n.d.). PSYCHOLOGICAL SCIENCE Research Article Sinning Saints and Saintly Sinners The Paradox of Moral Self-Regulation.

Smith, J. R., & Louis, W. R. (2008). Group norms and the attitude-behaviour relationship. Social and Personality Psychology Compass, 3(1), 19–35. https://doi.org/10.1111/j.1751-9004.2008.00161.x

Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. Journal of Environmental Psychology, 29(3), 309–317. https://doi.org/10.1016/j.jenvp.2008.10.004

Strack, F., Pauli, P., & Weyers, P. (2016). Editorial: Emotion and Behavior. Frontiers in Psychology, 7. https://www.frontiersin.org/article/10.3389/fpsyg.2016.00313

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral Affect and Moral Behavior. Annual Review of Psychology, 58, 345–372. https://doi.org/10.1146/annurev.psych.56.091103.070145

Tracy, J. L., & Robins, R. W. (2007). The psychological structure of pride: A tale of two facets. Journal of Personality and Social Psychology, 92(3), 506–525. https://doi.org/10.1037/0022-3514.92.3.506

Tu, J.-C., & Yang, C. (2019). Key Factors Influencing Consumers' Purchase of Electric Vehicles. Sustainability, 11(14), 3863. https://doi.org/10.3390/su11143863

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief positive and negative affect measures: The PANAS scales. Journal of Personality and Social Psychology, 54(6), 1063–1070. https://doi.org/10.1037//0022-3514.54.6.1063

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, 35(2), 151–175. https://doi.org/10.1037/0003-066X.35.2.151

Zhong, C.-B., & Liljenquist, K. (2006). Washing Away Your Sins: Threatened Morality and Physical Cleansing. Science, 313(5792), 1451–1452. https://doi.org/10.1126/science.1130726

IV. Paper 2

Feeling licensed after COVID-19 Rapid Test? Moderated Mediation of Social Expectations on MSI and Behavioural Spillovers

March 2022

Department of Psychological and Behavioural Science
London School of Economics and Political Science
Houghton Street, WC2A 2AE, London, UK
f.koaik@lse.ac.uk, d.krpan@lse.ac.uk and p.dolan@lse.ac.uk

1. Abstract

"I took the rapid test and did my part!"

One of the counterintuitive concepts in behavioural science is moral licensing. It describes when individuals' past good deeds justify their subsequent negative behaviour. In this study, we examine the effect of social norms and expectations on the likelihood of experiencing moral licensing after taking the COVID-19 rapid test, standing in a queue, and purchasing an electric car. More specifically, it explores whether social expectations affect moral self-image after taking a rapid test and how moral self-image explains the likelihood of donating to charity after taking a rapid test. We conducted an online experiment with 1,089 randomly allocated participants across six conditions. The findings explained a positive and significant moderated mediation model where social expectations moderated the effect of taking the rapid test ("behaviour 1") on moral self-image, and the increase in moral self-image indirectly explains the likelihood of donation to the UK Cancer Research Centre ("behaviour 2"). The results contribute to the literature on moral licensing and negative behavioural spillovers. It also offers practical implications to policymakers when designing policies and programs with the aim of instilling behavioural change; spillover effects are important but quite complex to capture and predict. This empirical study provides a new perspective on how social norms could lead to negative behavioural spillover effects, specifically moral licensing.

Keywords: Moral Licensing, Behavioural Spillovers, Social Expectations, Moral Self-Image

2. Introduction

On 18 May 2020, the United Kingdom (UK) government required citizens with COVID-19 symptoms such as cough, high temperature, or loss of taste or smell to get a Polymerase Chain Reaction (PCR) test[6]. Since April 2021, the UK National Health System (NHS) provided free access to rapid Lateral Flow Tests (LFT) that could be used at home for quick results. Rapid LFT is intended to detect COVID-19 in people at home with immediate results. The tests are over 80% effective at detecting infected people more likely to transmit the virus to others[7] (Smith et al., 2021). The government recommends that asymptomatic adults use rapid tests twice a week (every 3-4 days)[8] to avoid the transmission of the virus (Department of Health and Social Care, 2021). If the results are positive, a PCR test should be followed up. More recently, on January 11, 2022, the UK government announced that LFT is sufficient to detect infection in COVID-19. People who get positive results are required to self-isolate immediately. "I am grateful to the public and all of our critical workers who continue to test regularly and self-isolate, when necessary, along with other practical and important public health behaviours, as this is the most effective way of stopping the spread of the virus and keeping our friends, families and communities safe" (Chief Executive of UK Health Security Agency, Dr Jenny Harries, 2021).

The results of rapid tests (whether positive or negative) should be reported on a government website[9]; however, most results need to be reported (National Audit Office, 2021). In fact, up until May 23, 2021, around 691 million rapid LFTs were distributed in England. However, only 14% were registered as used. The NHS Trace and Test are undergoing behavioural research to understand the reasons behind the low reporting of the results (National Audit Office, 2021). In this case, public integrity and transparency are key factors. Taking a rapid test twice a week to prevent infection before going to crowded public places would reduce the spread of the virus. However, low reporting and uptake of the test indicate that people need to take the test more frequently or be made aware of the need and benefit of reporting the results. In either case, the

[6] Department of Health and Social Care, 2020
[7] https://www.nhs.uk/conditions/coronavirus-covid-19/testing/regular-rapid-coronavirus-tests-if-you-do-not-have-symptoms/
[8] https://www.gov.uk/government/news/new-campaign-urges-public-to-get-tested-twice-a-week
[9] https://www.gov.uk/report-covid19-result

behaviour of taking a rapid LFT is approved (the right thing to do) but not very normative or socially expected.

With the increasing and dynamic policies communicated by governments during the pandemic, Krpan and Dolan (2022) investigated how commanding messages impact compliance with COVID-19 behavioural recommendations. The research also explored the potential "spillover" and "spillunder" effects of the messages (Krpan and Dolan, 2022; Dolan & Galizzi, 2015; Krpan et al., 2021). When it comes to taking the LFT, there has not been any empirical investigation of whether social norms influence affect and moral appraisals after taking or not taking the test and the effect of social expectations on behavioural spillovers.

"No behaviour sits in a vacuum" (Dolan and Galizzi, 2015). Behavioural spillovers (Dolan and Galizzi, 2015) and spillunders (Krpan et al., 2021) try to look at all the ripples' effects of behavioural change interventions. The "mapping of these ripples is nowadays one of the most exciting pursuits in psychological research" (Kahneman, 2011, p.53). In fact, one of the most counterintuitive behavioural spillovers is moral licensing (Meritt et al., 2010; Blanken et al., 2015; Dolan and Galizzi, 2015). Permitting negative spillover, which describes engaging in a prior good deed, gives a license to perform an immoral or negative behaviour afterwards. We propose to explore how the feedback people receive from their social world shapes their views of their own moral selves and how these moral self-perceptions influence subsequent behaviours.

This paper aims to look at how social expectations influence behavioural spillovers through moral self-image. More specifically, it explores whether social expectations affect moral self-image after taking a rapid LFT, followed by whether moral self-image explains the likelihood of donating to charity after taking a rapid test. The aim is to contribute to the literature on moral licensing and negative behavioural spillovers by exploring how social expectations and moral self-regard could explain the possible spillover effect of engaging in behaviour 1 on behaviour 2.

This paper runs an empirical experiment to test whether 1) social expectations moderate the effect of engaging versus not engaging in behaviour on moral self-image, and 2) moral self-image mediates the effect of engaging versus not engaging in behaviour 1 on behaviour 2.[10]

[10] This paper will test predictions 1, 2 and 3 that are presented in the theoretical chapter (pages 19-21).

3. Method

The study supports answering the following research questions:

1) How do social expectations of a behaviour influence individuals' moral self-image and affect?
2) How do moral self-image and affect influence subsequent behaviour (behavioural spillover)?

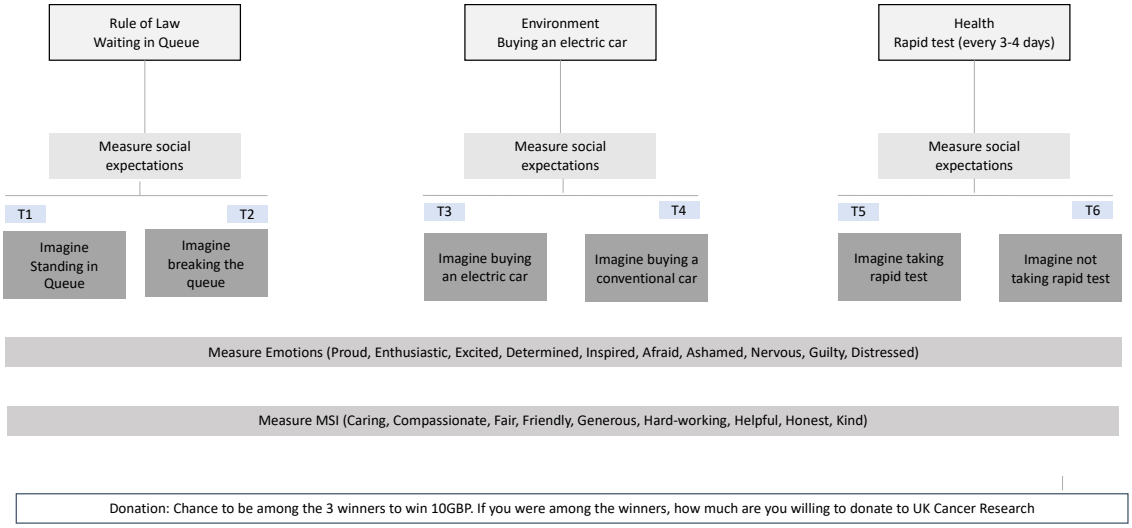
3.1 Participants and Procedure

The study was designed on Qualtrics, and participants were recruited from the UK via Prolific Academic. After completing the consent form, participants answered general demographic questions. Next, they were randomised into three conditions: 1) the rule of law (standing in the queue), 2) the environment (opting for an electric car), or 3) the rapid test (getting tested every 3-4 days). In the rule of law condition, they read a vignette scenario on the journey of grabbing a coffee on a Monday morning at a favourite coffee shop where a few other customers arrived at the same time. The second group received a vignette scenario on buying a new electric car. The third group received a vignette scenario on the UK government's recommendation to get a rapid LFT every 3-4 days before accessing closed public areas. Next, social expectations of the given behaviour were measured. Participants were asked about 1) the likelihood of them engaging in the behaviour, 2) their perceived empirical expectation (what they expect others to do), 3) their personal normative belief (what they think is the right thing to do), 4) their normative expectation (what they expect other to think is the right thing to do). Next, each group was randomised again into two conditions, making them a total of 6. The first condition asked participants to imagine standing in the queue (T1). The second condition asked them to imagine breaking the line (T2). The third condition asked participants to imagine opting for an electric car (T3). The fourth condition asked them to imagine opting for a conventional vehicle (T4). The fifth condition asked them to imagine getting LFT every 3-4 days (T5). The sixth condition asked them to imagine not getting LFT every 3-4 days (T6). After exposure to the manipulation, participants were asked questions assessing the dependent variables - affect and moral self-image - (see the *Measures* section below). Before leaving the experiment, we informed them about the chance for three participants to win a GBP 10 bonus. Then, they were asked how much they would be willing to donate (from 0 to 10 GBP) to UK Cancer Research[11].

[11] https://www.cancerresearchuk.org/

The three winners were randomly selected, and donations were requested on their behalf. The data for Study 2 were collected on October 25, 2021. This study complies with the ethics policy and procedures of the London School of Economics (LSE) and has been approved by its Research Ethics Committee. Figure 12 presents the design of the study and the six conditions.

Figure 12. Design of Study 2



3.2 Measures

The six independent variables are *Break* (1) (imagine breaking the line), *Wait* (2) (imagine waiting in line), *Electric* (3) (imagine buying an electric car), *Conventional* (4) (imagine buying a conventional vehicle, *Test* (5) (imagine getting LFT every 3-4 days), *Notest* (6) (imagine not getting LFT every 3-4 days).

Three continuous outcome variables were measured. The first dependent variable was affect measured via ten items answered on a 5-point Likert scale from 1 (very slightly or not at all) to 5 (Extremely), which required participants to report different affect when imagining engaging in the presented behaviour. The average score was reported on the five positive (inspired, enthusiastic, determined, proud and excited) and five negative affect (distressed, guilty, ashamed, nervous, and afraid)[12]. The final variable results from the difference between

[12] In the current study, we replaced three items ; alert, upset and scared from the short PANAS scale (Watson et al., (1988) with the following items proud, ashamed and guilty as they are relevant to the study. The items exhibited high inter-item correlations and significant overlap with other items. Correlation analyses revealed that these items demonstrated strong

the average positive and average negative affect. The second dependent variable, Moral-Self Image (MSI) as per Jordan et al. (2015) scale was measured via nine items answered on a 9-point Likert scale from 1 (Much less *caring* than the person I want to be) to 9 (Much more *caring* than the person I want to be). The average reported score on each of the nine dimensions indicated their MSI. The third dependent variable is the average amount of donations entered by participants in GBP. The social expectations questions were measured and used as moderators.

3.3 Sample

Based on a-priori power analysis in G*Power 3.1 (Faul et al., 2007), a minimum total sample of N= 619 will be needed for a 95% probability of detecting a medium-size effect (d = 0.2), assuming that numerator df is 10 (this comprises four hypotheses tests and six exploratory analyses), and that the number of groups is six as the study has six conditions. The total sample size was 1,089, thus indicating that the study was highly powered to detect medium effects, which is the norm endorsed by the field (Giner-Sorolla, 2016).

3.4 Hypotheses

This study will test the following hypotheses.

Hypothesis 1: The average MSI and affect differ across conditions; participants under positive conditions (*Wait*, *Electric*, and *Test*) experience lower MSI and affect than those under negative conditions (*Break*, *Conventional*, and *Notest*).

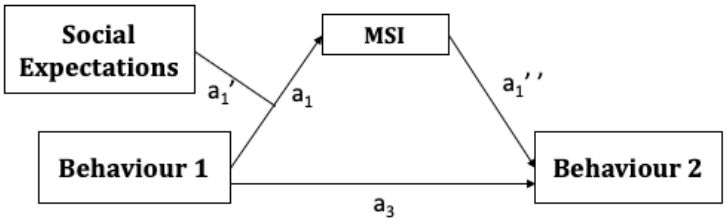
Hypothesis 2: The average donations significantly differ across conditions; participants under positive conditions (*Wait*, *Electric*, and *Test*) donate less than those under negative conditions (*Break*, *Conventional*, and *Notest*).

Hypothesis 3a: SE positively moderates the effect of engaging versus not in Behaviour 1 on MSI.; i.e., the higher the SE, the higher the MSI after engaging in Behaviour 1 (that is, the effect of Behaviour 1 on MSI is stronger for people who have higher SE).

Hypothesis 3b: When MSI is positively influenced (moderated) by SE, MSI negatively predicts Behaviour 2 (i.e., donations), which corresponds to moral licensing (i.e., a negative spillover effect). (Figure 13)

correlations (e.g., r=0.75, r=0.82). Furthermore, Cohen's d values were notably high (above 0.9), suggesting that these items contributed little unique variance to the overall measurement of affect.

Figure 13. Statistical diagram of the moderated mediation model



4. Results

For data analysis, we use STATA/ SE 15.1.

A total of 1,089 individuals participated in the experiment in exchange for 7.83 GBP per hour. Participants were recruited on Prolific Academic with three eligibility criteria: 1) country of residence in the United Kingdom, a minimum approval rate of 95%, and a minimum number of submissions of 100. The study was conducted on October 25, 2021.

Each participant was randomly assigned to one of the three conditions: Queue (n=363), Environment (n=362), and Rapid test (n=364). After answering the social expectations questions, each group was further randomised into two different conditions (total of 6 conditions): "*Wait"* (n=180), "*Break*" (n=183), "*Electric*" (n=182), "*Conventional*" (n=180), "*Test*" (n=183), "*Notest*" (n=181). In our sample of 1,089 participants, age (M=37.37, SD=12.42), 781 (71.72%) female and 308 (28.28%) male. Table 12 below summarises each treatment condition's mean and standard deviation for the dependent variables MSI and affect.

Table 12. Mean and Standard Errors (SE) (in parentheses) of dependent variables (Affect and MSI) by condition

		Wait	Break	Electric	Conventional	Rapid Test	No Rapid Test
Emotion	**Mean**	0.86	-2.08	1.50	0.87	0.36	-0.89
	SE	(0.06)	(0.08)	(0.08)	(0.10)	(0.08)	(0.09)
MSI	**Mean**	5.29	3.18	5.21	4.92	5.17	3.89
	SE	(0.07)	(0.14)	(0.06)	(0.07)	(0.07)	(0.11)

Manipulation checks

We run manipulation checks to test whether a significant difference exists between social expectations across conditions. There is a significant difference between social expectations (SE) in terms of the three conditions: environment, rule of law, and rapid test. Based on multiple regression output, waiting in the queue has 43% higher social expectations compared to buying an electric car (SE=0.011, t=36.91, 95% CI [0.41, 0.46]). Whereas the difference between the environment and the rapid test is smaller but still very significant, taking the rapid test has 38.5% higher social expectations compared to buying an electric car (SE=0.011, t=6.43, 95% CI [0.05, 0.09]). This explains that people expect others to stand in the queue and approve of it more than buying an electric car. Also, they expect others to take a rapid test and approve of it more than buying an electric vehicle. These results indicate a significant difference in social expectations across conditions (Table 13).

Table 13. Social expectations across the environment, rule of law, and rapid test conditions

	Rule of Law Standing in the queue		**Environment** Opting for an electric car		**Rapid Test** Taking Lateral Flow Test for COVI-19	
	Mean	**Standard Error**	**Mean**	**Standard Error**	**Mean**	**Standard Error**
Likelihood of Engaging in Behaviour (0 to 5)	4.57	(0.77)	3.37	(1.15)	3.57	(1.33)
Empirical Expectations (0-100)	81.17	(16.50)	15.72	(14.34)	35.90	(21.37)
PNB (0-5)	3.70	(1.00)	4.02	(0.79)	3.92	(1.01)
Normative Expectations (0-100)	83.76	(17.54)	61.36	(22.79)	56.48	(22.41)

Hypothesis 1

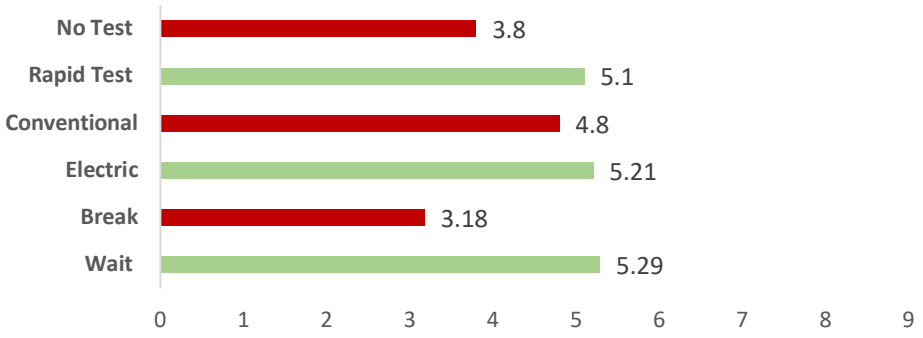
First, a one-way ANOVA indicated a significant difference across the six groups for the first dependent variable, MSI (1,089) = 81.91 p<0.0001. Then, a Bonferroni posthoc test revealed that MSI was significantly higher in the *Conventional* condition compared to the *Break* condition (b= 1.73, SE= 0.13), t = 12.75, p<0.001). So, as predicted, participants experienced more negative MSI when they imagined not standing in line compared to not buying an electric car.

Also, MSI was significantly higher in the *Notest* condition compared to the *Break* condition (b= 0.30 SE= 0.13, t = 5.20, p<0.001). As predicted, participants experienced more negative MSI when they imagined breaking the line compared to not taking the rapid LFT.

Additionally, MSI was significantly lower in the *Notest* condition compared to the *Conventional* condition (b=1.02, SE=0.13, t = -7.54, p<0.001). So, participants experienced more negative MSI when they imagined not taking LFT than not buying an electric car.

However, there was no significant difference in MSI between participants under the *Electric* condition compared to those under the *Wait* condition, between the *Electric* condition compared to those under the *Test* condition, and between the Wait condition compared to those under the *Test* condition. So, MSI differed significantly among "negative" conditions but did not differ among "positive" conditions (Figure 14).

Figure 14. Average MSI across conditions

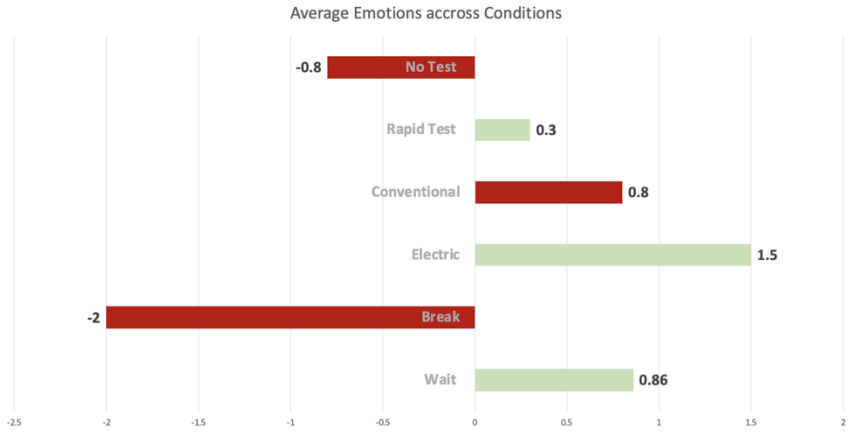


Second, a one-way ANOVA indicated a highly significant difference across the six groups for the second dependent variable, *Affect* (1,089) = 241.46 p<0.0001. Then, a Bonferroni post hoc test revealed statistically significant differences in experienced affect between all groups as predicted. More specifically, affect were significantly more positive in the *Electric* condition compared to the *Wait* condition (b= 0.64, SE= 0.12, t=5.23, p<0.000). So, participants

experienced more positive affect when they imagined buying an electric car compared to standing in line.

Also, affect were significantly more positive in the *Electric* condition compared to the *Test* condition (b= 1.14, SE= 0.12, t=9.38, p<0.000). So, participants experienced more positive affect when they imagined buying an electric car compared to taking the rapid test every 3-4 days. Additionally, affect were significantly more positive in the *Wait* condition compared to the *Test* condition (b= 0.5, SE= 0.12, t=4.12, p<0.001). So, participants experienced more positive affect when they imagined waiting in line compared to taking the test (Figure 15).

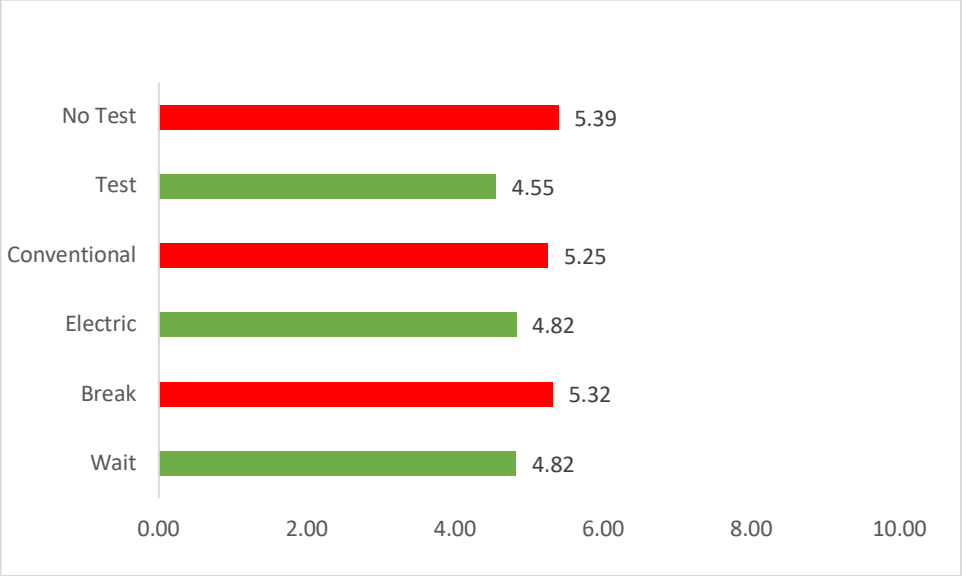
Figure 15. Average emotion across conditions



Hypothesis 2

In this part, we will look at the third independent variable: the number of donations from 1 to 10 GBP. Figure 16 shows the average donations across conditions. Overall, there was a significant difference between positive and negative conditions; positive conditions (*Wait*, *Electric* and *Test*) donated less than negative conditions (*Break*, *Conventional* and *Notest*) (b= –0.58, SE= 0.18, t= -3.11 and p<0.002), as predicted in H2. This indicates that the likelihood of the negative spillover effect was higher under positive conditions compared to negative conditions. When participants were asked to imagine engaging in the behaviour, they were more likely to donate than participants who imagined not engaging in the behaviour.

Figure 16. Donations in conditions



Zooming into the difference between conditions, the difference in average donations between *Break* and *Test* was significant at a 5% significance level. Specifically, breaking the line led to 0.766 more donations compared to getting tested (SE= 0.327, t=2.34 and p<0.019). Also, there was a significant difference between the *Test* and *Conventional*; participants who imagined buying a conventional car donated by 0.702 more than those who imagined taking LFT (SE= 0.328, t= 2.14, p<0.033). There was also a significant difference between *Test* and *Notest*. Participants who imagined not taking the test (*Notest*) donated more than those who imagined taking the LFT (T*est*) by 0.842 (SE= 0.328, t=2.57, p<0.01) (Table 14).

Table 14. The Effects of Behaviour 1 (six conditions) on Donation

Variable	b	SE b	95% CI	T	p
DV: Donations (GBP)					
Behaviour 1 - Test (Baseline)					
Wait	0.271	0.328	[-0.372, 0.916]	0.83	0.408
Break	0.766	0.327	[0.124, 1.408]	2.34	0.019

Electric	0.274	0.327	[-0.368, 0.917]	0.84	0.402
Conventional	0.702	0.328	[0.057, 1.347]	2.14	0.033
Notest	0.842	0.328	[0.198, 1.485]	2.57	0.01
Constant	4.55	0.231	[4.096, 5.004]	19.67	<0.001

Note. $R^2 = 0.0098$. DV refers to the dependent variable.

Exploratory Analyses

Moreover, to test what is causing the difference in donations, we computed simple linear regression models between donations and MSI and between donations and affect. There is a significant negative effect between MSI and donations at a 5% significance level. The higher the MSI, the lower the donations by –0.14, SE= 0.06, t=-2.30, p<0.021. However, affect did not affect donations as much; with higher affect, donations decreased by –0.05, SE= 0.56, t= -0.95, p>0.341. **This explains that MSI influenced donations much more than affect.**

In addition, to understand which condition experienced a higher likelihood of moral licensing, we ran a multiple regression model to compare the average donations among positive conditions (*Wait*, *Electric* and *Test*), participants in the *Test* condition donated the least (b= -0.77, SE= 0.26, t= -2.88, p<0.004) compared to the baseline conditions (*Break*, *Conventional* and *Notest*). Also, to understand which condition experienced a higher likelihood of moral cleansing, we ran a multiple regression model to compare the average donations among negative conditions (*Break*, *Conventional*, and *Notest*). Participants in the *Notest* condition donated the most (b= 0.66, SE= 0.26, t=2.46, p<0.014) compared to the baseline conditions (*Wait*, *Electric* and *Test*).

Discussion Hypothesis 2

Participants in the positive conditions donated less than those in the negative conditions, indicating permitting spillover effects (Dolan and Galizzi, 2015; Sachdeva et al., 2009; Zhong and Liljenquist, 2006). In other words, participants who were asked to imagine engaging in the behaviour (whether it is waiting in line, buying an electric car, or taking the rapid test) felt morally licensed and donated less compared to those who were asked to imagine not engaging in the respective behaviours. This explains moral licensinf effect. Moreover,

participants who imagined not engaging in the behaviours (breaking the line, not buying an electric car, or not getting an LFT test) donated more. This explains purging spillovers (Dolan and Galizzi, 2015), where people strive to maintain a positive self-image, so their self-regulation makes them engage in corrective actions after transgressions (Zhong and Liljenquist, 2006).

Overall, among the positive conditions, participants in the *Test* condition donated the least. In other words, those who imagined taking the rapid test donated significantly less than other participants, so they experienced permitting spillovers the most. They felt more morally licensed not to donate compared to participants who imagined buying an electric car or waiting in line.

To understand what influences donations, we found a significant negative effect between the average donations and MSI. Affect did not have any impact on donations, but MSI did. When MSI increases by one unit, the average donations decrease by 0.14. This relationship between MSI and donations explains permitting and purging spillovers. According to the Moral Credit Model (Meritt et al., 2010; Miller and Effron, 2010; Monin and Miller, 2001), the engagement in an initial positive behaviour gives moral credit, which offsets a subsequent negative behaviour (in this case, imagining taking the test increase moral credit and made participants feel licensed not donating much). Also, the engagement in an initial negative behaviour reduced the moral credit and made participants feel the need to donate more (moral cleansing) (Zhong and Liljenquist, 2006).

Another key finding is that *Test* and *Notest* conditions have the lowest and highest donations, respectively, compared to the other conditions. The rest of the analysis will focus on *Test* and *Notest* conditions to better understand the interaction effect (moderation and mediation) of social expectations and MSI on the effect of behaviour 1 on behaviour 2.

As an exploratory analysis, we ran a moderation analysis of the likelihood of engaging in the behaviour of taking the test on the reported MSI and affect. The effect of taking the test on affect is stronger with higher likelihood of engaging in the behaviour. However, this realtionshios does nto hold for the MSI.

Hypothesis 3a

To better understand the mechanism behind the change in MSI across conditions, we computed interaction analysis with Social Expectations (SE)[13] as the moderator (it ranges from 0-1). SE represent how much participants expect others to engage in the given behaviour and approve of it. The aim is to determine whether the change in SE moderates how different conditions influence MSI. With the *Notest* condition as a baseline, SE positively moderates the effect of *Wait* on MSI (b=2.495, SE= 0.814, t=3.06, p=0.002). SE also moderated the effect of *Electric* on MSI (b=2.763, SE 0.83, t=3.33, p=0.001). Moreover, SE positively moderates the effect of the *Test* on MSI (b= 2.932, SE= 0.743, t=3.95, p<0.001). The higher participants expected others to take the wait in a queue, buy an electric car and rapid test and approve of them, the more moral they perceive themselves when they imagine engaging in the behaviour (Table 15).

Table 15. The Effects of Test x Social Expectation Interaction on MSI Variable (*Notest* as baseline)

Variable	b	SE b	95% CI	t	p
DV = Moral Self-Image (MSI)					
Constant	4.77	0.26	[4.258, 5.282]	18.28	<0.001
Wait	-0.034	0.594	[-1.201, 1.132]	-0.06	0.953
Break	-1.442	0.609	[-2.638, -0.245]	-2.36	0.018
Electric	0.089	0.374	[-0.644, 0.823]	0.24	0.812
Conventional	0.352	0.375	[-0.384, 1.090]	0.94	0.348
Test	-0.091	0.368	[-0.813, 0.631]	-0.25	0.804
SE	-1.821	0.505	[-2.813, -0.829]	-3.6	<0.001
Wait x SE	2.495	0.814	[0.897, 4.093]	3.06	0.002
Break x SE	1.652	0.829	[0.026, 3.279]	1.99	0.046
Electric x SE	2.763	0.83	[1.134, 4.393]	3.33	0.001
Conventional x SE	1.312	0.82	[-0.297, 2.923]	1.6	0.11
Test x SE	2.932	0.743	[1.474, 4.391]	3.95	<0.001

Note. R^2 = 0.2883. Highlighted in grey are the significant moderation effect at 0.25% significance level.

[13] Social Expectations (SE) is a continuous variable computed from the average of EE and NE (Empirical and Normative Expectations)

With *Test* as a baseline, SE moderates the effect of *Notest* on MSI (b=-2.932, SE= 0.743, t= -3.95, p<0.001) (Table 16).

Table 16. The Effects of Test x Social Expectation Interaction on MSI Variable (T*est* as baseline)

Variable	b	SE b	95% CI	t	p
DV = Moral Self-Image (MSI)					
Constant	4.679	0.259	[4.169, 5.189]	18.01	<0.001
Wait	0.056	0.594	[-1.201, 1.132]	0.09	0.924
Break	-1.35	0.609	[-2.546, -0.154]	-2.22	0.027
Electric	0.18	0.373	[-0.552, 0.912]	0.48	0.629
Conventional	0.444	0.375	[-0.291, 1.179]	1.18	0.237
Notest	0.091	0.368	[-0.631, 0.813]	0.25	0.804
SE	1.111	0.545	[0.041, 2.181]	2.04	0.042
Wait x SE	-0.437	0.839	[-2.084, 1.210]	-0.52	0.603
Break x SE	-1.279	0.853	[-2.955, 0.395]	-1.5	0.134
Electric x SE	-0.168	0.855	[-1.846, 1.509]	-0.2	0.843
Conventional x SE	-1.619	0.845	[-3.279, 0.039]	-1.92	0.056
Notest x SE	-2.932	0.743	[-4.391, -1.474]	-3.95	<0.001

Note. R^2 = 0.2883. Highlighted in grey are the significant moderation effect at 0.25% significance level.

Discussion Hypothesis 3a

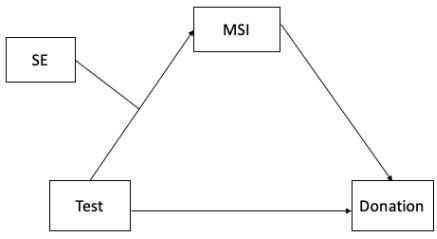
The interaction results were significant for the *Test* conditions at the FDR significance level. SE positively and significantly moderate the effect of getting tested on MSI. Expecting others to get tested and approve of it increases moral self-regard. When SE is low, individuals do not feel moral about taking a rapid test. This indicates that the behaviour of getting a rapid lateral flow test is interdependent (Bicchieri, 2006). These moderation analyses suggest that getting an electric car always leads to positive emotional outcomes, and breaking the line leads to negative outcomes. **However,** the rapid LFT is not independent and depends on social expectations.

In H2, we showed that MSI negatively influences donations. In H3a, we showed that SE moderated the effect of getting tested and not getting tested on MSI. In the next part, we put all these relationships in a moderated mediated model to further study how SE moderates the effect of getting tested (B1) on MSI and how the latter mediates the relationship between getting tested and donations (B2).

Hypothesis 3b

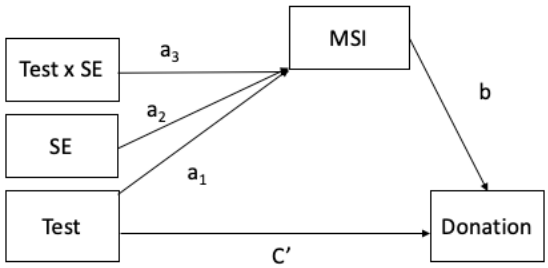
H3b tests a moderated mediation model to better understand all the underlying mechanisms between SE and MSI on the effect of getting rapid lateral flow COVID-19 test on donations. Based on Hayes (2018) Process Model 7 on *Conditional Process Modelling*, we ran a first stage *moderated mediation* (Baron and Kenny, 1986; Zhao et al., 2010; Preacher et al., 2007) as the focus is to estimate the indirect effect of the product of the independent variable (Test) and the moderator (SE) on the dependent variable (Donation) through the mediator (MSI). The model is represented in Figure 17 in a conceptual diagram; the indirect (or mediated) effect of getting tested on Donations via MSI (the mediator) is assumed to be contingent on the level of SE (how much the behaviour of getting tested is socially expected).

Figure 17. Conceptual Diagram



The statistical diagram (Figure 18) operationalises the conceptual model as two equations. The first equation (M) involves estimating regression parameters from a model where the proposed mediator (MSI) is regressed into the independent variable - Test (X), proposed moderator - SE (W), and the product of the independent moderator variables - Test x SE (XW). In other words, the first model is consistent with a standard moderated multiple regression.
The second equation (Y) involves the dependent variable—donation (Y)—regressed into the proposed mediator (MSI) and the independent variable (Test). In other words, the second model is consistent with a standard mediation regression.

Figure 18: Statistical Diagram



Equation (1)

$M = i_M + a_1*test + a_2*SE + a_{3*}$ (Test*SE)

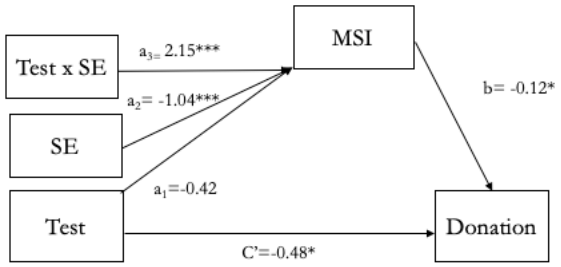
Equation (2)

$Y= i_y + c'*Test + b*MSI$

For Equation (1) and as presented in H3a, there is a positive and statistically significant effect of the interaction term (moderator) Test*SE on MSI. In other words, there is evidence that the effect of getting tested on MSI is conditioned on SE (b=2.15, SE=0.65, t=3.30, p<0.001, 95% CI [0.87, 3.44]). So, the higher the SE, the higher the MSI when imagining taking the rapid test.

For Equation (2), the mediator MSI has a negative effect on the relationship between getting tested and donations. This effect is significant (b= -0.12, SE=0.06, t=-1.96, p<0.05, 95% CI [-0.25, 0.0000]. However, the direct effect of testing on donation was negative and not significant (b=-0.48, SE=0.25, t=-1.90, p<0.058, 95% CI [-.992, 0.016]). Figure 19 presents the conceptual diagram with the results.

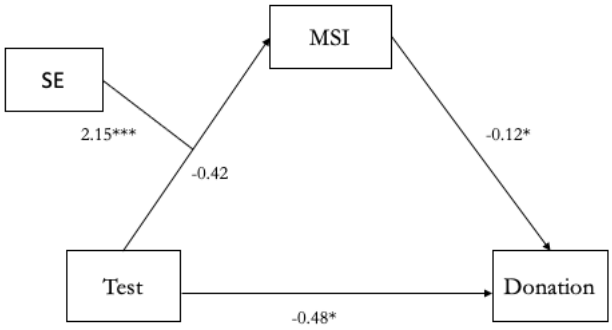
Figure 19. Conceptual Diagram: results of the moderated mediation model



In the model, we specifically test whether the indirect effect of Behaviour 1 (i.e., Test) on Donation through SE moderates MSI. To examine and reject the null hypothesis that SE has no moderation effect on the mediation of MSI between Test and Donation, we look at the *Index of Moderation Mediation* (IMM) in the STATA final output. IMM quantifies the degree to which the conditional indirect effect is a linear function of the moderator (Hayes, 2015). IMM is "a_3b_1", indicating the change in the conditional indirect effect (a_3) per unit increase of the moderating variable. IMM = 2.15* (-0.12) = - 0.26

Figure 20 presents the statistical diagram with the results.

Figure 20. Statistical Diagram: results of the moderated mediation model



Based on 10000 bootstrap confidence interval testing, IMM is statistically significant (-0.26, Bootstrap error=-0.003, 95% CI [-0.5819, -0.0099]). The negative effect of IMM indicates that as we move from lower SE to higher SE, the indirect effect becomes increasingly negative. In other words, the higher the SE, the stronger the negative indirect impact of engaging verus not engaging in Behaviour 1 on donations through MSI. **This explains that when getting the rapid test, social expectations moderate the negative indirect effect of MSI on donation.**

Now that we have provided evidence of a significant negative IMM (i.e., we have showed that there is a significant moderated mediation effect), we will look more closely at the Conditional Indirect Effect (CIE).

To quantify this indirect effect, Hayes (2015) express it in the following equation (3):

Conditional Indirect Effect (CIE) = $(a_1b_1 + a_3b_1W)$

Using the values from our regression outputs, the function for the conditional indirect effect is f(W)= (-0.42+2.15*SE) *-0.12. The model provides a different effect for various levels of SE. The model tests three levels of CIE (CIElow for -1sd, CIEmed for mean, and CIEhigh for +1sd). Table 17 highlights the three levels, the output of the Bootstrap results, and the 95% confidence interval.

Table 17. Moderated Mediation: Bootstrap Results

	SE level	Observed Coefficient	Bootstrap error	[95% Intervals]	Confidence	Significance

CIElow (-1sd)	0.30	-0.029	0.020	-0.079	0.003	Not significant
CIEmed (mean)	0.55	-0.096	0.051	-0.201	-0.004	Significant
CIEhigh (+1sd)	0.80	-0.163	0.087	-0.348	-0.006	Significant

The results show that there is a significant conditional indirect effect of SE at SE levels higher than the mean (i.e., higher than 0.55).

5. Discussion

As predicted in the framework, there is a significant difference between reported MSI across negative conditions only. Participants who imagined breaking the line reported the lowest MSI (3.18), followed by those who imagined not taking the test (3.8). Participants who imagined not buying an electric car reported the highest MSI among the negative conditions (4.8). This difference is completely absent among positive groups (i.e., *Wait*, *Electric* and *Test*). It seems that moral self-regard didn't differ when they imagined waiting in line, buying an electric car, or taking the test (unlike what we predicted in the framework). It could be that UK citizens identify standing in the queue with morality. So, there is a strong association between waiting in line and positive moral self-regard.

Regarding the second behaviour, the average donations made by participants, there is a significant difference between positive and negative conditions. Participants who imagined waiting in line, buying an electric car, and taking the rapid test donated less than those who imagined not engaging in these behaviours. This indicates the manifestation of the moral licensing effect. Looking at the difference among the positive conditions, participants who imagined taking the rapid test donated the least.

This led us to investigate all these relationships between Test, SE, MSI and donations. We present a moderated mediation model that helps link all these causal relationships in one model. The positive and significant results showed that social expectations explain when people feel licensed versus consistent with their previous actions. In this experiment, social expectations significantly moderated the effect of getting tested on MSI, and then MSI mediated the effect of getting tested on donations.

Theoretical Implication on Behavioural Spillovers

The objective of this study is to show that there is a higher likelihood of moral licensing when social expectations positively moderate the indirect effect of MSI on the influence of B1on B2. In this experiment, we zoomed into the behaviour of getting rapid tests, and we were able to present a significant *Moderated Mediation Model*. We found that when social expectations are high, MSI is more likely to increase as a result of engaging versus not engaging in Behaviour 1 (i.e., participants felt more moral about getting tested when they expected others to get tested and approve of it), and when MSI increased, donations decreased significantly. Therefore, the indirect mediation effect of MSI on B1 to B2 is moderated by social expectations. These findings are in line with our predictions. The results also contribute to the literature on moral licensing versus moral consistency. There is a huge debate on when people stay consistent with their past behaviours and when they feel licensed to transgress (Blanken et al., 2015). This model shows that when social expectations moderate the effect of B1 on MSI, it is more likely to lead to moral licensing rather than consistency. For instance, if we are driven by social expectations, we are more likely to experience negative behavioural spillovers. Future research will examine how this moderated mediation model could be applied in different settings, such as pro-environmental behaviour and virtue signalling. Also, it would be interesting to look at whether consistency and positive spillovers are more likely to happen when personal norms influence MSI rather than social expectations. If B1 increases MSI with higher levels of personal normative beliefs, would this increase in MSI indirectly influence positive spillovers in B2?

Practical Policy Implications

How do these findings translate into practical and public policy applications? “No behaviour sits in a vacuum” (Dolan and Galizzi, 2015). When policymakers design policies and programs to instil behavioural change, spillover effects are important but quite complex to capture and predict. This empirical study provides a new perspective on when people are more likely to experience moral licensing. Also, it shows that social expectations influence how people feel about their actions and behaviours. For instance, people who experience a positive appraisal of their morality after wearing a mask only when they expect others to approve of and wear it are more likely to experience moral licensing. Since social expectations influence how people feel about their moral selves, they are more likely to transgress and act in the opposite direction. However, when behaviour is an exception, and not many people engage in it, there is wide

approval that it is the right thing to do. Acting on it will always lead to a positive moral self-image and is less likely to lead to moral licensing.

6. Limitation

The first limitation of the present study is that behaviour one consisted of a hypothetical behaviour where participants were asked to imagine engaging or not engaging in the given behaviour. However, one point to raise is that the chosen behaviours are relatable, and participants can easily imagine the given scenario. The second limitation is self-reported affect and MSI; again, participants could act out of social desirability when reporting their affect (especially those related to buying an electric car). Knowing this behaviour is attractive and recommended; they might report higher to increase social desirability. Future research will then need to probe the mechanism using actual behaviour that participants get to perform.

7. Conclusion

The present research makes several main contributions. Most importantly, it looks at the effect of social expectations on behavioural spillovers, particularly moral licensing. It also offers a detailed examination of the mechanisms and underlying interactions behind the moral licensing effect. Whereas previous research in the area of behavioural spillovers, particularly consistency versus balancing, focused on moderators (Blanken et al., 2015) and did not investigate the mediation role of moral self-regard (Jordan et al., 2015), our focus was on exploring the role of MSI and social expectations. In line with the moral credit model (Merritt et al., 2010), we showed that MSI indirectly explains the effect of behaviour 1 on behaviour 2 when social expectations moderate the effect of behaviour 1 on MSI. In other words, when we engage in behaviour 1 and social expectations lead to a more positive moral self-image, we are more likely to transgress in behaviour 2 and experience moral licensing. For example, taking the rapid test and feeling good about this behaviour because it is socially expected, we are more likely to feel licensed to transgress in immoral action afterwards. It would be interesting to explore in future research when personal norms (rather than social expectations) moderate the effect of behaviour 1 on MSI, are we more likely to stay consistent in future actions? Also, future research could replicate these results in different domains and settings to better understand when people experience moral licensing versus consistency.

8. References

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological), 57(1), 289–300.

Bicchieri, C. (2006). The grammar of society: The nature and dynamics of social norms (pp. xvi, 260). Cambridge University Press.

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A Meta-Analytic Review of Moral Licensing. Personality and Social Psychology Bulletin, 41(4), 540–558. https://doi.org/10.1177/0146167215572134

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. Journal of Economic Psychology, 47, 1–16. https://doi.org/10.1016/j.joep.2014.12.003

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Jordan, J., Leliveld, M. C., & Tenbrunsel, A. E. (2015). The Moral Self-Image Scale: Measuring and Understanding the Malleability of the Moral Self. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.01878

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1–22. https://doi.org/10.1080/00273171.2014.962683

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, *85*(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100

Krpan, D., Galizzi, M. M., & Dolan, P. (2021). Looking at Spillovers in the Mirror: Making a Case for "Behavioral Spillunders". Frontiers in Psychology, 10, 1142. https://doi.org/10.3389/fpsyg.2019.01142

Krpan, D., & Dolan, P. (2022). You Must Stay at Home! The Impact of Commands on Behaviors During COVID-19. Social Psychological and Personality Science, 13(1), 333–346. https://doi.org/10.1177/19485506211005582

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral Self-Licensing: When Being Good Frees Us to Be Bad. Social and Personality Psychology Compass, 4(5), 344–357. https://doi.org/10.1111/j.1751-9004.2010.00263.x

Miller, D. T., & Effron, D. A. (2010). Psychological License. In Advances in Experimental Social Psychology (Vol. 43, pp. 115–155). Elsevier. https://doi.org/10.1016/S0065-2601(10)43003-8

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. Journal of Personality and Social Psychology, 81(1), 33–43. https://doi.org/10.1037/0022-3514.81.1.33

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. Multivariate Behavioral Research, 42(1), 185–227. https://doi.org/10.1080/00273170701341316

Sachdeva, S., Iliev, R., & Medin, D. L. (2009.). PSYCHOLOGICAL SCIENCE Research Article Sinning Saints and Saintly Sinners The Paradox of Moral Self-Regulation.

Zhao, X., Lynch Jr., J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. Journal of Consumer Research, 37(2), 197–206. https://doi.org/10.1086/651257

Zhong, C.-B., & Liljenquist, K. (2006). Washing Away Your Sins: Threatened Morality and Physical Cleansing. Science, 313(5792), 1451–1452. https://doi.org/10.1126/science.1130726

V. Paper 3

Train or plane?

Moral Licensing after pro-environmental Behaviour – A Moderated Mediation Model

May 2023

Department of Psychological and Behavioural Science
London School of Economics and Political Science
Houghton Street, WC2A 2AE, London, UK
f.koaik@lse.ac.uk, d.krpan@lse.ac.uk and p.dolan@lse.ac.uk

Train or plane?

Moral Licensing after Pro-Environmental Behaviour - A Moderated Mediation Model

1. Abstract

Tackling climate change requires fundamental lifestyle changes to reduce emissions, such as adopting low-emission transport. However, the focus has been on behavioural interventions while less attention is given to the possible ripple effects of one behaviour on other actions. "Behavioural spillover is an enticing concept that targets behavioural interrelationships in ways that may catalyse broader lifestyle changes, spreading from one behaviour to another, bringing about greater impacts" (Nash and Whitmarsh, 2023; Hampton and Whitmarsh, 2023). This paper employs an empirical approach to explore the pro-environmental behavioural spillovers, specifically moral licensing, when choosing sustainable transport. To date, mixed studies showed inconsistent causal processes to explain behavioural spillovers. To better understand when moral licensing is more likely to occur, we explore how the feedback people receive from their social world shapes their views of their moral selves and how these moral self-perceptions influence subsequent behaviours. We test a moderated mediation model examining how social expectations influence behavioural spillovers through moral self-image. We conducted an online experiment on 600 participants in the UK and tested the effect of taking the train compared to the plane from London to Amsterdam. The findings present a significant moderated mediation model where social expectations positively moderated the effect of taking the train on moral self-image, and the increase in moral self-image indirectly explains the decrease in donating for CO2 offsetting compared to taking the plane. The results contribute to the literature on moral licensing and offer a detailed examination of the mechanisms and underlying interactions behind the moral licensing effect.

Keywords: Pro-Environmental Behaviour, Moral Licensing, Behavioural Spillovers, Social Expectations, Moral Self-Image

2. Introduction

Pro-environmental behavioural spillover has gained momentum in social sciences to move beyond simple behavioural interventions to realise sustainable lifestyle change (Nash et al., 2017). Insights from pro-environmental spillover effects and the magnitude of these effects are important when designing environmental policies geared to promote sustainable behaviours, instil attitudinal changes, and increase awareness and compliance with pro-environmental policies (Thøgersen and Noblet, 2012). In addition to ensuring positive spillover effects, policymakers are looking to reduce the occurrence of negative spillover effects, such as moral licensing, that are counterproductive. Even though people are increasingly aware of climate change and its profound consequences, little is reflected in their travel behaviour (Cohen and Higham, 2010; McDonald et al., 2015; Henle, 2022). To the best of our knowledge, there is little empirical evidence on behavioural spillovers in the pro-environmental domain, specifically related to alternative sustainable transport. Some of the evidence is related to reported intentions and self-reported measures instead of actual behavioural change (Xu et al., 2018) or, in other cases, from correlational analysis rather than causality (Thøgersen, 2012). Therefore, there is a need for more experimental research that looks at the underlying mechanisms to explain different behavioural spillover effects and capture them better when people are likely to stay consistent with their initial pro-environmental behaviours. In this paper, we investigate the spillover effect in the domain of air travel behaviour, specifically related to taking the train versus the plane, to reduce emissions. We empirically measure the effect of social norms (which has been neglected in the literature on behavioural spillovers (Nash et al., 2017) on moral self-image when engaging in the pro-environmental behaviour of taking the plane. Reflective of the big interest in behavioural spillovers in the social sciences literature and the growing research on climate action, this study builds on this work and extends the theory by testing a new moderated mediation model to understand the underlying mechanisms behind moral licensing. It tests social norms as a moderator and moral self-image as a mediator to explain the likelihood of moral licensing when taking sustainable transport.

Mode of Transportation: Train or Plane

Sustainable mobility contributes to social and economic welfare by meeting the needs of society by moving freely and accessing services without leading to environmental damage and resource depletion (Whitmarsh et al., 2007). One of the most significant contributors to CO2 emissions from the transport sector is aviation, where CO2 emissions from passenger volumes

are expected to result in 22% of global CO2 emissions by 2050 (Cames et al., 2015). “The air transport industry is growing faster than we are currently producing and introducing technological and operational advances which reduce the environmental impact at source” (Commission of the European Communities, 2001: 1). Moreover, air transport has three main environmental effects related to air pollution, noise, and climate change (or global warming). Two main solution pathways could help reduce these effects: technological advances or behavioural change to induce the adoption of low-carbon transport modes (Dällenbach, 2020). Research showed that substituting the mode of transportation from trains to planes reduces the negative environmental impact of air transport (Hoikkala and Magnussen, 2019; Dällenbach, 2020). High-speed trains minimise carbon dioxide emissions (Whitelegg et al., 2001). The same trip leads to an average of 80-90% less emissions from the train compared to the plane (Dällenbach, 2020). When choosing alternative travel modes, travel time is the most important criterion for taking the plane over other alternative modes of travel (Cascetta et al., 2011; European Commission, 2010; Behrens and Pels, 2012). Other studies showed that in addition to travel time, the lower costs, the frequency of connections and the punctuality of trains play a crucial role (Bieger and Laesser, 2001; Behrens and Pels, 2012). Bieger and Laesser (2004) highlight that the underlying cost, as well as the sustainability of the behaviour, influence the preference to choose the train. The shift towards a more sustainable mode of transport is a complex policy issue with various possible solutions, ranging from technological and innovative solutions that can curb the environmental burden of aviation to behavioural change of citizens. The behaviour of citizens represents a key area as the changes to the psychological drivers are as critical as infrastructure changes (Hunecke et al., 2010). Many studies look into the drivers behind substituting flying behaviour to achieve global climate change goals (Albalate et al., 2015), with a growing literature on the importance of changing the behaviour of citizens and the social norms behind sustainable transportation (Möser and Bamberg, 2008; Lanzini and Khan, 2016). However, no study tested the spillover effects of sustainable transport, i.e., when does engaging in this positive behaviour increase the likelihood of engaging in additional pro-environmental behaviours?

Pro-Environmental Behavioural Spillovers

In social psychology, according to the Theory of Planned Behaviour (TPB) (Azjen, 1991), behaviours are driven by the perceived consequences of a specific action, the perceived social norms, and the perceived control over the situation. Similarly, based on the Value Belief Norm

(VBN) theory (Stern, 2000), personal norms (i.e., internalised rules and beliefs to behave in a certain way) drive behaviours to preserve the value of specific concepts or objects. For instance, if an individual values the environment and the latter is threatened, personal norms intrinsically drive the behaviour to act sustainably. In addition, research showed the importance of social norms (Cialdini et al., 1990; Cialdini and Goldstein, 2004) in adopting pro-environmental behaviours.

From reasoned theories to "non-reasoned" theories (Nash et al., 2019), behavioural spillover research draws on the fact that engaging in one behaviour can sometimes influence engaging in another related behaviour. "Behavioural spillover research concerns the possibility of voluntary, wider lifestyle shifts beyond piecemeal behaviour change" (Nash et al., 2019). Behavioural spillovers (Dolan and Galizzi, 2015) and spillunders (Krpan et al., 2021) try to look at all the ripples of the effects of behavioural change interventions. These ripple effects are captured in various disciplines related to psychology, sociology, and economics (Austin et al., 2011). Regarding pro-environmental behaviour, spillover effects could take different forms, from positive to negative effects (Truelove et al., 2014; Dolan and Galizzi, 2015; Nash et al., 2017; Nilsson et al., 2017).

Positive behavioural spillovers occur when a targeted behavioural intervention leads to engagement in one pro-environmental behaviour, increasing the likelihood of engaging in other positive behaviours (Thøgersen, 2004; Thøgersen, 2014). Such positive pathways stem from a general desire for consistency with past behaviours (Bem, 1972), pro-environmental self-identity (Lauren et al., 2018), or increased environmental support and awareness (Carrico et al., 2018). For instance, a study in Wales examined positive spillovers after introducing single-use plastic bag charges, increasing people's environmental self-perceptions (Poortinga et al., 2013). Many studies also examined the positive effect of increasing environmental self-identity and a "greener" sense of self that increases the likelihood of acting consistently with past behaviours (Lauren et al., 2018).

Negative behavioural spillovers explain that phenomenon when a targeted behavioural intervention limits engagement in subsequent untargeted actions (Blanken et al., 2015). Moral licensing is one of the most counterintuitive negative behavioural spillovers (Meritt et al., 2010; Miller and Effron, 2010; Blanken et al., 2015; Dolan and Galizzi, 2015). It describes the idea that engaging in a prior good deed gives a license to perform an immoral or negative behaviour afterwards. Individuals accumulate hypothetical moral credits in a metaphorical moral bank account and possibly use them to offset or license subsequent negative behaviour (Miller and

Effron, 2010). This allows people to maintain a positive moral self-image (Nisan and Horenczyk, 1990; Jordan et al., 2011). When it comes to the pro-environmental domain, moral licensing predicts subsequent immoral behaviours after engaging in a first pro-environmental action that boosts positive self-perception, inflates self-image, and positive pro-environmental identity (Kouchaki, 2011; Cornelissen et al., 2013).

Aims of the paper

To better understand when moral licensing is more likely to occur in the pro-environmental domain, we propose to explore how the feedback people receive from their social world shapes their views of their moral selves and how these moral self-perceptions influence subsequent behaviours. Social expectations are subjective assumptions about what people do and what they approve of (Bicchieri, 2006, 2014). The latter could take two forms: normative expectations describing what others approve of and empirical expectations describing what others do. In addition, moral self-image measures how (im)moral people perceive themselves at a particular point in time; it is malleable and influenced by contextual, social, and situational factors (Jordan et al., 2015). Hence, this paper examines how social expectations influence behavioural spillovers through moral self-image. More specifically, it explores whether social expectations affect moral self-image after taking the train ("behaviour 1") and how this moral self-image subsequently explains the likelihood of paying for CO2 offsetting ("behaviour 2"). The aim is to contribute to the literature on moral licensing and negative behavioural spillovers by exploring how social expectations and moral self-regard influence the effect of engaging in behaviour 1 on behaviour 2. This research is novel as it explores social norms/ expectations and moral self-regard in the context of spillovers and moral licensing (social expectations and moral self-regard have not yet been investigated in the context of spillovers and moral licensing).

This paper runs an empirical experiment to test whether:

1) social expectations moderate the effect of (not) engaging in behaviour on moral self-image,
2) moral self-image mediates the effect of (not) engaging in behaviour 1 (taking the train instead of the plane) on behaviour 2 (deciding how much to offset for CO2 emissions).[14]

[14] This paper will test predictions 1, 2 and 3 that are presented in the theoretical chapter (pages 19-21).

3. Method

The study contributes to understanding how social expectations of climate change behaviours influence individuals' moral self-image and affect. Additionally, it explains the underlying psychological mechanisms by revealing the role of moral self-image and affect in predicting subsequent behaviour related to CO2 offsetting (i.e., the manifestation of possible behavioural spillover).

3.1 Participants and Procedure

The study was designed on Qualtrics, and participants were recruited from the UK via Prolific Academic. After completing the consent form, participants answered general demographic questions. Next, social expectations of climate change behaviour were measured. Participants were asked about 1) the likelihood of them engaging in climate change behaviour, 2) their perceived empirical expectation (what they expect others to do), 3) their personal normative belief (what they think is the right thing to do), 4) their normative expectation (what they expect other to think is the right thing to do) (Bicchieri, 2014).

Next, they were randomly assigned to two conditions: 1) imagine taking the train from London to Amsterdam and 2) imagine taking the plane from London to Amsterdam. In the first group, participants were asked to imagine taking the train from London to Amsterdam, in which case they would be producing 14 Kg in CO2 emissions instead of 124 Kg by plane. In the second group, participants were asked to imagine taking the plane instead, producing approximately 124 Kg of CO2 emissions, compared to almost 14 Kg of CO2 emissions by train. After exposure to the manipulation, participants were asked questions assessing the dependent variables - affect and moral self-image - (see the *Measures* section below). Finally, just before the experiment ended, participants were informed about the chance to win a 10 GBP bonus. Then, they were asked how much they would be willing to buy CO2 certificates to offset their CO2 footprint (Carbon Emission Task (Berger and Wyss, 2021)). Therefore, they had to report their willingness to pay (from 0 to 10 GBP) for CO2 offsetting. The winners were randomly selected, and donations were requested on their behalf. This study complies with the ethics policy and procedures of the London School of Economics (LSE) and has been approved by its Research Ethics Committee.

3.2 Measures

The study was a between-subject design where participants were allocated to one of the two conditions: (1) imagine taking the train from London to Amsterdam and (2) imagine taking the plane from London to Amsterdam. The study involved one independent variable (Behaviour 1) consisting of two levels: *Train* (1) (imagine taking the train from London to Amsterdam) and *Plane* (0) (imagine taking the plane from London to Amsterdam)

Three continuous outcome variables were measured. The first dependent variable was affect measured via ten items answered on a 5-point Likert scale from 1 (very slightly or not at all) to 5 (Extremely), which required participants to report different affect when imagining engaging in the presented behaviour (Thompson, 2007; Watson et al., 1988). The average score was reported on the five positive (inspired, enthusiastic, determined, proud and excited) and five negative affect (distressed, guilty, ashamed, nervous, and afraid)[15]. The final variable results from the difference between the average positive and average negative affect. The second dependent variable, Moral-Self Image (MSI) as per Jordan et al. (2015) scale was measured via nine items (i.e., caring, compassionate, helpful, hard-working, friendly, fair, generous, honest, and kind) answered on a 9-point Likert scale from 1 (Much less *caring* than the person I want to be) to 9 (Much more *caring* than the person I want to be). The average of the reported scores on each of the nine dimensions was used to indicate their MSI. MSI was used as a mediator in the analysis. The third dependent variable is the Carbon Emission Task (CET) (Berger and Wyss, 2021) where they need to choose how much to donate for Co2 certificates to offset their CO2 emissions (from 0-10 GBP). They are given the chance to win 10 GBP by entering a raffle at the end of the study. The social expectations questions were measured and used as a moderator.

3.3 Sample Size

[15] In the current study, we replaced three items ; alert, upset and scared from the short PANAS scale (Watson et al., (1988) with the following items proud, ashamed and guilty as they are relevant to the study. The items exhibited high inter-item correlations and significant overlap with other items. Correlation analyses revealed that these items demonstrated strong correlations (e.g., r=0.75, r=0.82). Furthermore, Cohen's d values were notably high (above 0.9), suggesting that these items contributed little unique variance to the overall measurement of affect.

Based on a-priori power analysis in G*Power 3.1 (Faul et al., 2007), a minimum total sample of N= 600 (300 per condition) will be needed for a 95% probability of detecting a medium-size effect (d = 0.2) given a level of significance (α = 0.05).

3.4 Hypotheses

This study will test the following hypotheses:

H1: Participants in the train condition report higher MSI and more positive affect than those in the plane condition.

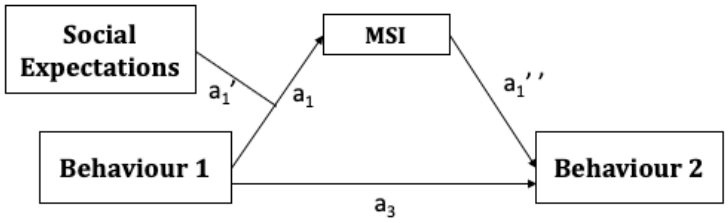
H2: Participants in the train condition are less likely to donate CO2 certificates compared to those in the plane condition.

H3a: SE moderates the effect of B1 (train versus plane) on MSI.

H3b: MSI predicts the likelihood of donating for CO2 certificates.

H4: There is a significant moderated mediation with Behaviour 1 as the independent variable, SE as a moderator, MSI as a mediator, and Behaviour 2 as the dependent variable. That is, the influence of engaging versus not engaging in Behaviour 1 (taking the train versus plane) on MSI is moderated by SE (i.e., this influence is stronger for people who score higher on SE), and MSI, in turn, negatively predicts Behaviour 2 (donating CO2 certificates). Moderated mediation model of SE and MSI that explains the moral licensing effect. (Figure 21)

Figure 21. Statistical diagram of the moderated mediation model



4. Results

For data analysis, we use STATA/ SE 17. 605 individuals participated in the experiment in exchange for 5 GBP per hour. Participants were recruited on Prolific Academic with three eligibility criteria: 1) country of residence in the United Kingdom, a minimum approval rate of 95%, and a minimum number of submissions of 100. The study was conducted on December 27, 2022. Each participant was randomly assigned to one of the two conditions: Train (n=306)

or plane (n=299). In our sample of 605 participants, age (M=43.8, SD=13.9), 361 were female, and 244 were male. Table 18 below summarises each treatment condition's mean and standard deviation for the dependent variables MSI and affect.

Table 18. Mean and Standard Errors (SE) (in parentheses) of dependent variables (Affect and MSI) by condition

		Train	**Plane**
Emotion	**Mean**	2.17	2.04
	SE	(0.03)	(0.03)
MSI	**Mean**	5.35	4.51
	SE	(0.05)	(0.07)

Social Expectations

Social expectations around the behaviour of taking the train instead of a plane to save on CO2 emissions and combat climate change present a big difference between what is personally approved and what is expected to be approved by others. In other words, on a scale from 0-100, the average reported personal normative belief is 80.36. hence, most participants believe that taking the train instead of the plane is the right thing to do. However, they believe that 36.81% of other people in their social network take the train instead of the plane, and they believe that only 57.52% approve of this behaviour. In addition, participants are 67.9% likely to engage in the mentioned behaviour. These figures show that there is a pluralistic ignorance when it comes to people's perception of what others do or approve of doing when it comes to climate change behaviours. Although most approve of pro-environmental behaviour, they do not expect others to do the same or even engage in it (Table 19).

Table 19. Social expectations on the climate change behaviour

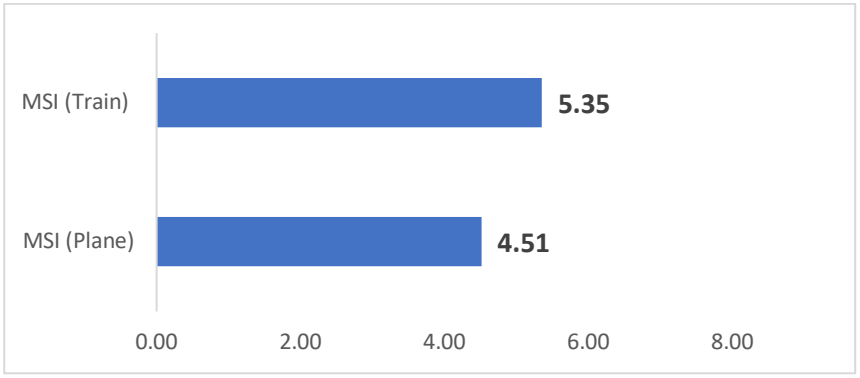
	Climate Change
	Taking a train rather than a plane to save on CO2 emissions
Likelihood of Engaging in Behaviour (0-100)	67.9
	(0.98)
Empirical Expectations (0-100)	36.81
	(0.94)

PNB	80.36
(0-100)	(1.00)
Normative Expectations	57.52
(0-100)	(0.98)

Hypothesis 1

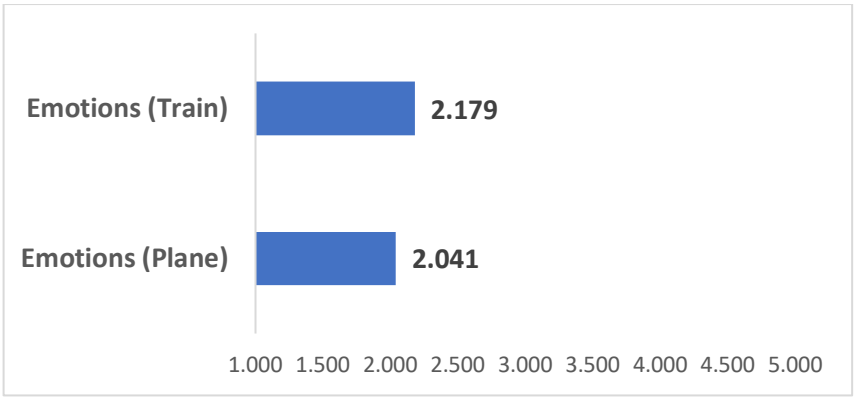
To test if there is a significant difference in MSI across the two conditions, an independent samples t-test (two-tailed) was performed. Participants in the *Train* condition (M=5.35, SD=0.05) had higher MSI compared to participants in the *Plane* condition (M=4.51, SD=0.07) t (605) = -8.64, p<0.01, d=-.83, 95% CI [-1,02, -0.64]. This indicates that participants experienced much higher MSI when imagining taking the train instead of the plane to save on CO2 emissions (Figure 22).

Figure 22. Average MSI across conditions



Also, an independent samples t-test (two-tailed) was performed to compare affect across the two conditions. Participants in the *Train* condition (M=2.17, SD=0.03) had more positive affect compared to participants in the *Plane* condition (M=2.04, SD=0.03) t (605) = -2.93, p<0.003, d=-.13, 95% CI [-0.23, -0.04]. This indicates that participants experienced slightly higher average affect when imagining taking the train instead of the plane to save on CO2 emissions.

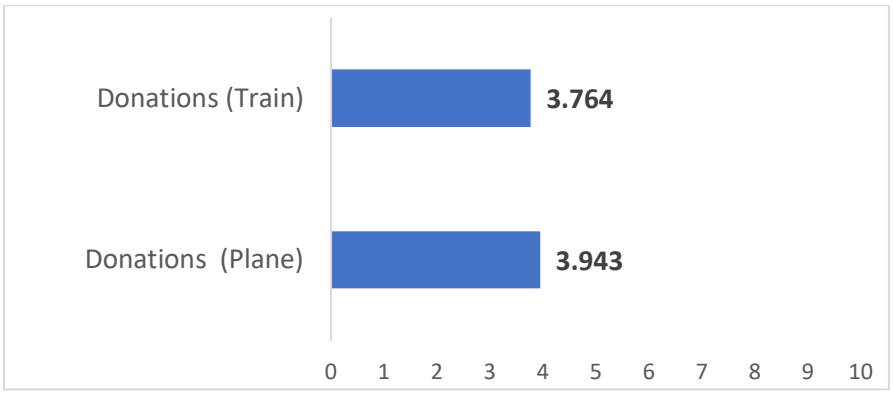
Figure 23. Average emotion across conditions



Hypothesis 2

In this part, we look at the independent variable: the number of donations for CO2 offsetting from 1 to 10 GBP. Figure 24 shows the average donations across the two conditions. An independent samples t-test (two-tailed) shows that participants in the *Train* condition (M=3.76, SD=0.18) had lower donations for CO2 offsetting on average compared to participants in the *Plane* condition (M=3.94, SD=0.18); however, this difference was not statistically significant, t (605) = 0.69, p=0.45, 95% CI [-0.32, 0.68].

Figure 24. Donations for CO2 Offsetting in each condition



Hypothesis 3a

To better understand the underlying mechanism, we computed an interaction analysis with Social Expectations (SE) as the moderator (it ranges from 0-100). SE represent how much participants expect others to engage in the given behaviour and approve of it. The aim is to

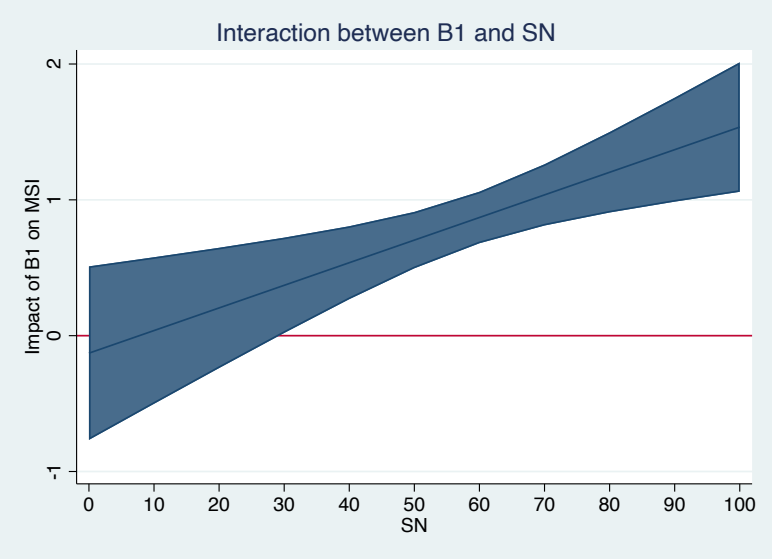
look at whether the change in SE moderates how different conditions influenced MSI. With the *Plane* condition as a baseline, SE significantly moderates the effect of *Train* on MSI (b=0.016, SE= 0.005, t=3.12, p=0.002). The higher the participants expected others to take the train, the more moral they perceived themselves when they imagined engaging in the behaviour (Table 20).

Table 20. The Effects of Behaviour 1 x Social Expectation Interaction on MSI Variable (*Plane* as baseline)

Variable	b	SE b	95% CI	t	p
DV = Moral Self-Image (MSI)					
Constant	5.26	0.23	[4.8099, 5.720]	22.72	<0.001
Train	-0.12	0.32	[0-.76733, 0.51003]	-0.40	0.693
SE	-0.012	0.003	[-.02043, -.00544]	-3.39	0.001
Train x SE	0.016	0.005	[4.8099, 5.720]	3.12	0.002

Note. $R^2 = 0.1187$. Highlighted in grey are the significant moderation effect at 1% significance level.

Figure 25. Interaction Analysis between Behaviour 1, Social Norms and MSI



Hypothesis 3b

To test whether MSI predicts the likelihood of donating for CO2 certificates. We computed simple linear regression models between donations and MSI. There is a significant negative effect between MSI and donations at a 5% significance level. The higher the MSI, the lower the donations by –0.26, SE= 0.10, t=-2.62, p<0.009. We also computed a mediation analysis with a significant indirect effect of MSI (M= -0.22; [-0.40, -0.05]). This shows that MSI indirectly explains the effect of taking the train versus the plane on the second behaviour (CO2 offsetting).

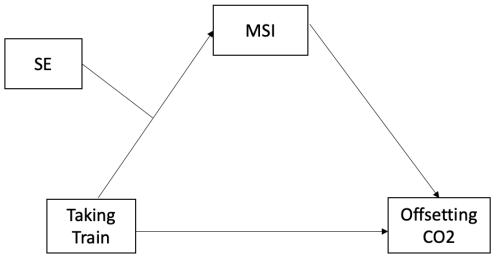
In H3a, we showed that SE moderated the effect of using a train on MSI. In H3b, we showed that MSI predicted the donations made for CO2 offsetting. In the next part, we put all these relationships in a moderated mediated model to further study how SE moderates the effect of getting the train (B1) on MSI and how the latter mediates the relationship between getting the train and donations (B2).

Hypothesis 4

H4 tests a moderated mediation model to understand better the underlying mechanisms between SE and MSI regarding the effect of using the train instead of the plane on donations. Based on Hayes's (2018) Process Model 7 on *Conditional Process Modelling*, we ran a first stage *moderated mediation* (Baron and Kenny, 1986; Zhao et al., 2010; Preacher et al., 2007) as the focus is to estimate the indirect effect of the product of the independent variable (Test) and the moderator (SE) on the dependent variable (Donation) through the mediator (MSI). The

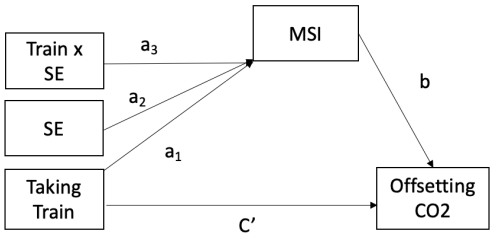
model is represented in Figure 26 in a conceptual diagram. The indirect (or mediated) effect of getting tested on Donations via MSI (the mediator) is assumed to be contingent on the level of SE (how much the pro-environmental behaviour is socially expected).

Figure 26. Conceptual Diagram



The statistical diagram (Figure 27) operationalises the conceptual model as two equations. The first equation (M; see next page) involves estimating regression parameters from a model where the proposed mediator (MSI) is regressed into the independent variable – Behaviour 1 (B1) (X), proposed moderator - SE (W), and the product of the independent moderator variables – B1 x SE (XW). In other words, the first equation is consistent with a standard moderated multiple regression. The second equation (Y; see next page) involves the dependent variable - Donation (Y), regressed into the proposed mediator (MSI) and the independent variable (B1). In other words, the second model is consistent with a standard mediation regression.

Figure 27. Statistical Diagram



Equation (1)

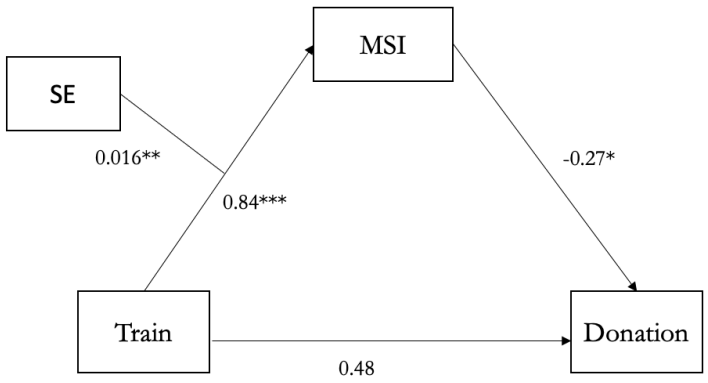
$M = i_M + a_1*B1 + a_2*SE + a_{3*} (B1*SE)$

Equation (2)

$Y= i_y + c'*B1 + b*MSI$

For Equation (1) and as presented in H3a, there is a positive and statistically significant effect of the interaction term (moderator) B1*SE on MSI. In other words, there is evidence that the effect of using the train on MSI is slightly conditioned on SE (b=0.016, SE=0.005, t=3.12, p<0.002, 95% CI [0.006, 0.027]). So, the higher the SE, the stronger the positive influence of taking the train (versus the plane) on MSI when taking the train instead of the plane. For Equation (2), there is a negative effect of the mediator MSI on the relationship between taking the train and donations. The effect is significant at the 5% significance level (b= -0.27, SE=0.10, t=-2.53; p<0.012, 95% CI [-0.48, 0.06]. Also, the direct effect of taking the train on donation is not significant (b=0.048, SE=0.26, t=0.18, p<0.858, 95% CI [-.482, 0.578]). Figure 28 presents the conceptual diagram with the results.

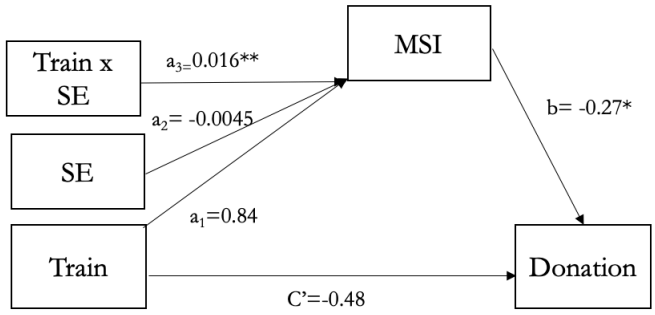
Figure 28. Conceptual Diagram: Results of the moderated mediation model



In the model, we are specifically testing whether the indirect effect of B1 (i.e., taking train versus plane) on B2 (Donation) through MSI is moderated by SE. To examine and reject the null hypothesis that SE has no moderation effect on the mediation of MSI between B1 and

Donation, we look at the *Index of Moderation Mediation* (IMM) in the STATA final output. IMM quantifies the degree to which the conditional indirect effect is a linear function of the moderator (Hayes, 2015). IMM is "a_3b_1", indicating the change in the moderating variable's conditional indirect effect (a3) per unit increase.
IMM = 0.016* (-0.27) = - 0.0044. Figure 29 presents the statistical diagram with the results.

Figure 29. Statistical Diagram: results of the moderated mediation model



Based on 10000 bootstrap confidence interval testing, IMM is statistically significant (-0.0044, Bootstrap error=-0.002, 95% CI [-0.009, -0.000]. The negative effect of IMM indicates that as we move from lower SE to higher SE when thinking of taking the train instead of the plane, the indirect effect becomes increasingly negative. In other words, the higher the SE when considering taking the train leads to higher MSI and lower donations in CO2 offsetting. **This provides evidence that social expectations moderate the negative impact of taking the train (versus a plane) through MSI's donation.**
Now that we have shown a significant negative IMM (i.e., a significant moderated mediation effect), we will look more closely at the Conditional Indirect Effect (CIE). To quantify this indirect effect, Hayes (2015) express it in the following equation (3):

Conditional Indirect Effect (CIE) = ($a_1b_1 + a_3b_1W$)
Using the values from our regression outputs, the function for the conditional indirect effect is f(W)= (0.84+0.016*SE) *-0.27. The model provides a different effect for various levels of SE. The model tests three levels of CIE (CIElow for -1sd, CIEmed for mean, and CIEhigh for +1sd). Table 21 highlights the three levels of the output of the Bootstrap results and the 95% confidence interval.

Table 21. Moderated Mediation: Bootstrap Results

	SE level	Observed Coefficient	Bootstrap error	[95% Confidence Intervals]		Significance
CIElow (-1sd)	Low	-0.145	0.068	-0.290	-0.030	Significant
CIEmed (mean)	Medium	-0.235	0.097	-0.432	-0.047	Significant
CIEhigh (+1sd)	High	-0.302	0.127	-0.553	-0.058	Significant

5. Discussion

As predicted in the framework, there is a significant difference between the reported MSIs of the two conditions. Participants who imagined taking the plane reported a lower MSI (4.51) than those who imagined taking the train (5.35). Self-reported affect are also different but not as significant as MSI (2.17 for the train group compared to 2.04 for the plane group). So, it seems there is a strong association between pro-environmental behaviour and positive moral self-regard.

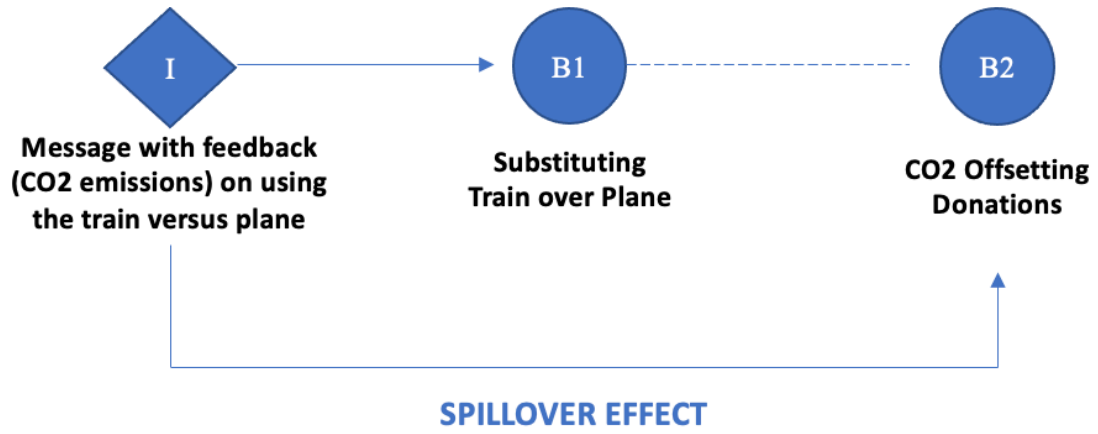
When it comes to the second behaviour, which is the average donations made by participants for CO2 offsetting, there is no significant difference between the two conditions. However, MSI significantly moderates the relationship between taking the train versus the plane on the donations made for CO2 offsetting. The latter indicates the effect of moral self-regard on the spillover effect and highlights the role that moral self-image plays in participants' subsequent decisions. It further presents the Moral Credit Model (Merritt et al. 2010, Miller and Effron 2010), in which people accumulate moral credits in a hypothetical moral bank account. Moral credits are used to either engage in moral actions or offset immoral actions (Merritt et al. 2010, Miller and Effron 2010).

In addition, the moderated mediation model tested the relationships between social expectations, MSI and the two behaviours. The positive and significant results showed that social expectations explain when people feel licensed versus consistent with their previous actions. In this experiment, social expectations significantly moderated the effect of taking the train on MSI, and then MSI mediated the effect of taking the train on CO2 offsetting.

Implications on Behavioural Spillovers

The present research contributes to the literature's behavioural spillovers and moral licensing discussion (Nash et al., 2017, 2019; Gholamzadehmir et al., 2019; Truelove et al., 2014). In this experiment, we tested the behavioural spillovers pro-environmental behaviour related to transport substitution to reduce emissions and present a significant *Moderated Mediation Model*. The experimental design fits the paradigm of behavioural spillovers (Figure 30) and presents one of the few quantitative and empirical evidence of the underlying psychological mechanism of the moral licensing effect. It presents moral self-image as one mediator that indirectly explains the moral licensing effect. In addition, it presents social norms as one of the moderators that influence engaging in pro-environmental behaviour on MSI.

Figure 30. The paradigm of Behavioural Spillovers



The study showed a higher likelihood of moral licensing when social expectations positively moderate the indirect effect of MSI on the influence of behaviour 1 on behaviour 2. We found that when social expectations were high, MSI increased (i.e., participants felt significantly

more positive about their moral self-image when taking the train, especially when they expected others to do the same and approve of it). When MSI increased, donations decreased significantly compared to the other conditions. Therefore, the indirect mediation effect of MSI on Behaviour 1 to Behaviour 2 is moderated by social expectations. These findings are in line with our predictions. It contributes to the debate in the literature on when moral licensing is more likely to occur, i.e., when people stay consistent with their past behaviours and when they feel licensed to transgress (Blanken et al., 2015). This model presents that when social expectations moderate the effect of B1 on MSI, it is more likely to lead to moral licensing rather than consistency. For instance, if we are driven by social expectations, we are more likely to experience negative behavioural spillovers. Future research will look at how this moderated mediation model could be applied in different settings, such as online virtue signalling.

Policy Implications

How do these findings translate into practical and public policy applications? This research contributes to designing policies and programs to instil long-lasting behavioural change (Nash et al., 2017 and 2019), as spillover effects are important but quite complex to capture and predict (Truelove et al., 2014; Galizzi and Whitmarsh, 2019). More specifically, it shows that social expectations influence how people feel and perceive their pro-environmental actions. As policymakers and behavioural science practitioners have been using social norms messaging to induce behavioural change, this empirical study posits questions on when social norms drive a pro-environmental behaviour, which could lead to a negative spillover effect. People are more likely to experience positive moral self-regard when behaving according to social expectations and this increase in moral credit gives them license to transgress in subsequent behaviour. Therefore, such research is needed to capture positive or negative spillover effects and understand the moderators and mediators that explain these spillover effects. Only by doing so can policymakers and practitioners devise policies and interventions that account for these effects and maximise the net environmental impacts (Nash et al., 2017).

6. Limitation

The first limitation of the present study is that the first behaviour (about imagining taking the train versus the plane) consisted of a hypothetical behaviour where participants were asked to imagine engaging or not engaging in the given behaviour. However, one point to raise is that

the chosen behaviour is very relatable, and participants can easily imagine the given scenario. Although we measure the second behaviour as actual behaviour performance (actual donation for CO2 offsetting), the second limitation is the self-reported affect and MSI; again, participants could act out of social desirability when reporting their affect. Knowing this behaviour is attractive and recommended; they might report higher to increase social desirability. A third limitation is that the sample involved in this study was recruited through Prolific Academia from the UK. More research needs to be conducted to understand similar behavioural effects in non-WEIRD countries and examine the influence of cultural and normative differences. Future research will then need to probe the mechanism using actual behaviour that participants get to perform. Also, future research should seek to conduct behavioural spillover experiments in the field and on a larger sample size to resemble the general public closely.

7. Conclusion

The present research makes several main contributions and adds to the current literature that seeks to identify the conditions under which moral licensing in the pro-environmental domain occurs, specifically behaviours related to sustainable transport. Most importantly, it tests a novel moderated mediation model in the context of pro-environmental behaviours. It looks at the effect of social norms on behavioural spillovers, particularly moral licensing. It also offers a detailed examination of the mechanisms and underlying interactions behind the moral licensing effect. Understanding how social norms, messaging, and communication might backfire and lead to unintended consequences is crucial for policymakers and practitioners when devising related policies. Whereas previous research in the area of behavioural spillovers, particularly consistency versus balancing, focused on moderators (Blanken et al., 2015) and did not investigate the mediation role of moral self-regard (Jordan et al., 2015), our focus was on exploring the role of MSI and social expectations. In line with the moral credit model (Merritt et al., 2010), we showed that MSI indirectly explains the effect of behaviour 1 on behaviour 2 when social expectations moderate the effect of behaviour 1 on MSI. In other

words, when we engage in behaviour 1 and social expectations lead to a more positive moral self-image, we are more likely to transgress in behaviour 2 and experience moral licensing. It would be interesting to explore in future research the difference between acting in public versus private in the initial behaviour on subsequent behavioural spillovers. Also, future research could replicate these results in different domains and settings to better understand when people experience moral licensing versus consistency.

8. References

Ajzen, I. (1991). The theory of planned behaviour. *Theories of Cognitive Self-Regulation*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Albalate, D., Bel, G., & Fageda, X. (2015). Competition and cooperation between high-speed rail and air transportation services in Europe. *Journal of Transport Geography*, *42*, 166–174. https://doi.org/10.1016/j.jtrangeo.2014.07.003

Austin, A., Cox, J., Barnett, J., & Thomas, C. (2011). *Exploring catalyst behaviours: Full Report: A report to the Department for Environment, Food and Rural Affairs*.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Behrens, C., & Pels, E. (2012). Intermodal competition in the London–Paris passenger market: High-Speed Rail and air transport. *Journal of Urban Economics*, *71*(3), 278–288. https://doi.org/10.1016/j.jue.2011.12.005

Bem, D. J. (1972). Self-Perception Theory11Development of self-perception theory was supported primarily by a grant from the National Science Foundation (GS 1452) awarded to the author during his tenure at Carnegie-Mellon University. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 6, pp. 1–62). Academic Press. https://doi.org/10.1016/S0065-2601(08)60024-6

Berger, S., & Wyss, A. M. (2021). Measuring pro-environmental behavior using the carbon emission task. *Journal of Environmental Psychology*, *75*, 101613. https://doi.org/10.1016/j.jenvp.2021.101613

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms* (pp. xvi, 260). Cambridge University Press.

Bicchieri, C. (2016). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Bieger, T., & Laesser, C. (2001). The role of the railway with regard to mode choice in medium range travel. *Tourism Review*, *56*(1/2), 33–39. https://doi.org/10.1108/eb058354

Bieger, T., & Laesser, C. (2004). Information Sources for Travel Decisions: Toward a Source Process Model. *University of St.Gallen*, *42*. https://doi.org/10.1177/0047287504263030

Blanken, I., Van De Ven, N., & Zeelenberg, M. (2015). A Meta-Analytic Review of Moral Licensing. *Personality and Social Psychology Bulletin*, *41*(4), 540–558. https://doi.org/10.1177/0146167215572134

Cames, M., Graichen, J., Siemons, A., & Cook, V. (2015). *Emission Reduction Targets for International Aviation and Shipping*.

Carrico, A., Raimi, K., Truelove, H., & Eby, B. (2017). Putting Your Money Where Your Mouth Is: An Experimental Test of Pro-Environmental Spillover From Reducing Meat Consumption to Monetary Donations. *Environment and Behavior*, *50*. https://doi.org/10.1177/0013916517713067

Cascetta, E., Papola, A., Pagliara, F., & Marzano, V. (2011). Analysis of mobility impacts of the high speed Rome–Naples rail link using withinday dynamic mode service choice models. *Journal of Transport Geography*, *19*(4), 635–643. https://doi.org/10.1016/j.jtrangeo.2010.07.001

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 24, pp. 201–234). Academic Press. https://doi.org/10.1016/S0065-2601(08)60330-5

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015. https://doi.org/10.1037/0022-3514.58.6.1015

Cialdini, R., & Goldstein, N. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, *55*, 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Cohen, S., Higham, J., & Cavaliere, C. (2011). Binge flying: Behavioural addiction and climate change. *Annals of Tourism Research*, *38*, 1070–1089. https://doi.org/10.1016/j.annals.2011.01.013

Cornelissen, G., Dewitte, S., & Warlop, L. (2011). Are Social Value Orientations Expressed Automatically? Decision Making in the Dictator Game. *Personality and Social Psychology Bulletin*, *37*(8), 1080–1090. https://doi.org/10.1177/0146167211405996

Dällenbach, N. (2020). Low-carbon travel mode choices: The role of time perceptions and familiarity. *Transportation Research Part D: Transport and Environment*, *86*, 102378. https://doi.org/10.1016/j.trd.2020.102378

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Galizzi, M. M., & Whitmarsh, L. (2019). How to Measure Behavioral Spillovers: A Methodological Review and Checklist. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/article/10.3389/fpsyg.2019.00342

Gholamzadehmir, M., Sparks, P., & Farsides, T. (2019). Moral licensing, moral cleansing and pro-environmental behaviour: The moderating role of pro-environmental attitudes. *Journal of Environmental Psychology*, *65*, 101334. https://doi.org/10.1016/j.jenvp.2019.101334

Gleave, S. D. (2006). Air and rail competition and complementarity. *Case Study Report Prepared for European Commission DG Energy and Transport*, 1–149.

Haggar, P., Whitmarsh, L., & Nash, N. (2023). A Drop in the Ocean? Fostering Water-Saving Behavior and Spillover Through Information Provision and Feedback. *Environment and Behavior*, *55*(6–7), 520–548. https://doi.org/10.1177/00139165231201371

Hampton, S., & Whitmarsh, L. (2023). Choices for climate action: A review of the multiple roles individuals play. *One Earth*, *6*, 1157–1172. https://doi.org/10.1016/j.oneear.2023.08.006

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1–22. https://doi.org/10.1080/00273171.2014.962683

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, *85*(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100

Henle, M. (2022). *Pro-environmental behavior and motivation to offset flight emissions among flight passengers* [Thesis: Master, ZHAW Zürcher Hochschule für Angewandte Wissenschaften]. https://doi.org/10.21256/zhaw-26324

Hunecke, M., Haustein, S., Böhler, S., & Grischkat, S. (2010). Attitude-based target groups to reduce the ecological impact of daily mobility behavior. *Environment and Behavior*, *42*(1), 3–43.

Jordan, A. H., Monin, B., Dweck, C. S., Lovett, B. J., John, O. P., & Gross, J. J. (2011). Misery Has More Company Than People Think: Underestimating the Prevalence of Others' Negative Affect. *Personality and Social Psychology Bulletin*, *37*(1), 120–135. https://doi.org/10.1177/0146167210390822

Jordan, J., Leliveld, M. C., & Tenbrunsel, A. E. (2015). The Moral Self-Image Scale: Measuring and Understanding the Malleability of the Moral Self. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01878

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the Moral Self: The Effects of Recalling Past Moral Actions on Future Moral Behavior. *Personality and Social Psychology Bulletin*, *37*(5), 701–713. https://doi.org/10.1177/0146167211400208

Kouchaki, M. (2011). Vicarious Moral Licensing: The Influence of Others' Past Moral Actions on Moral Behavior. *Journal of Personality and Social Psychology*, *101*, 702–715. https://doi.org/10.1037/a0024552

Krpan, D., Galizzi, M. M., & Dolan, P. (2019). Looking at Spillovers in the Mirror: Making a Case for "Behavioral Spillunders". *Frontiers in Psychology*, *10*, 1142. https://doi.org/10.3389/fpsyg.2019.01142

Lanzini, P., & Khan, S. A. (2017). Shedding light on the psychological and behavioral determinants of travel mode choice: A meta-analysis. *Transportation Research Part F: Traffic Psychology and Behaviour*, *48*, 13–27. https://doi.org/10.1016/j.trf.2017.04.020

López, R. A., Sánchez, T. R., & Aicart, L. M. (2021). Attitudes towards the environment related to travel and personal mobility. *XIV Conference on Transport Engineering, CIT2021*, *58*, 201–208. https://doi.org/10.1016/j.trpro.2021.11.028

McDonald, R. I., Chai, H. Y., & Newell, B. R. (2015). Personal experience and the 'psychological distance' of climate change: An integrative review. *Journal of Environmental Psychology*, *44*, 109–118. https://doi.org/10.1016/j.jenvp.2015.10.003

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral Self-Licensing: When Being Good Frees Us to Be Bad. *Social and Personality Psychology Compass*, *4*(5), 344–357. https://doi.org/10.1111/j.1751-9004.2010.00263.x

Miller, D. T., & Effron, D. A. (2010). Psychological License. In *Advances in Experimental Social Psychology* (Vol. 43, pp. 115–155). Elsevier. https://doi.org/10.1016/S0065-2601(10)43003-8

Möser, G., & Bamberg, S. (2008). The effectiveness of soft transport policy measures: A critical assessment and meta-analysis of empirical evidence. *Journal of Environmental Psychology*, *28*, 10–26. https://doi.org/10.1016/j.jenvp.2007.09.001

Nash, N., & Whitmarsh, L. (2023). *One thing leads to another? Pro-environmental behavioural spillover* (pp. 63–77). https://doi.org/10.4337/9781800882133.00011

Nash, N., Whitmarsh, L., Capstick, S., Hargreaves, T., Poortinga, W., Thomas, G., Sautkina, E., & Xenias, D. (2017). Climate-relevant behavioral spillover and the potential contribution of social practice theory. *WIREs Climate Change*, *8*(6). https://doi.org/10.1002/wcc.481

Nash, N., Whitmarsh, L., Capstick, S., Thøgersen, J., Gouveia, V., de Carvalho Rodrigues Araújo, R., Harder, M. K., Wang, X., & Liu, Y. (2019). Reflecting on Behavioral Spillover in Context: How Do Behavioral Motivations and Awareness Catalyze Other Environmentally Responsible Actions in Brazil, China, and Denmark? *Frontiers in Psychology*, *10*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.00788

Nilsson, M., Chisholm, E., Griggs, D., Howden-Chapman, P., McCollum, D., Messerli, P., Neumann, B., Stevance, A.-S., Visbeck, M., & Stafford Smith, M. (2018). Mapping interactions between the sustainable development goals: Lessons learned and ways forward. *Sustainability Science*, *13*, 1–15. https://doi.org/10.1007/s11625-018-0604-z

Nisan, M., & Horenczyk, G. (1990). Moral balance: The effect of prior behaviour on decision in moral conflict. *British Journal of Social Psychology*, *29*(1), 29–42. https://doi.org/10.1111/j.2044-8309.1990.tb00884.x

Poortinga, W., Whitmarsh, L., & Suffolk, C. (2013). The introduction of a single-use carrier bag charge in Wales: Attitude change and behavioural spillover effects. *Journal of Environmental Psychology*, *36*, 240–247. https://doi.org/10.1016/j.jenvp.2013.09.001

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behavioral Research*, *42*(1), 185–227. https://doi.org/10.1080/00273170701341316

Stern, P. (2000). Toward a Coherent Theory of Environmentally Significant Behavior. *Journal of Social Issues*, *56*, 407–424. https://doi.org/10.1111/0022-4537.00175

Thøgersen, J. (1999). Spillover processes in the development of a sustainable consumption pattern. *Journal of Economic Psychology*, *20*(1), 53–81. https://doi.org/10.1016/S0167-4870(98)00043-9

Thøgersen, J. (2004). A Cognitive Dissonance Interpretation of Consistencies and Inconsistencies in Environmentally Responsible Behavior. *Journal of Environmental Psychology - J ENVIRON PSYCHOL*, *24*, 93–103. https://doi.org/10.1016/S0272-4944(03)00039-2

Thøgersen, J. (2006). Norms for environmentally responsible behaviour: An extended taxonomy. *Journal of Environmental Psychology*, *26*(4), 247–261. https://doi.org/10.1016/j.jenvp.2006.09.004

Thøgersen, J., & Noblet, C. (2012). Does Green Consumerism Increase the Acceptance of Wind Power? *Energy Policy*, *51*. https://doi.org/10.1016/j.enpol.2012.09.044

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., & Vandenbergh, M. P. (2014). Positive and negative spillover of pro-environmental behaviour: An integrative review and theoretical framework. *Global Environmental Change*, *29*, 127–138. https://doi.org/10.1016/j.gloenvcha.2014.09.004

Van der Werff, E., Steg, L., & Keizer, K. (2014). I Am What I Am, by Looking Past the Present: The Influence of Biospheric Values and Past Behavior on Environmental Self-Identity. *Environment and Behavior*, *46*(5), 626–657. https://doi.org/10.1177/0013916512475209

Wallace, E., Buil, I., & Chernatony, L. (2020). 'Consuming Good' on Social Media: What Can Conspicuous Virtue Signalling on Facebook Tell Us About Prosocial and Unethical Intentions? *Journal of Business Ethics*, *162*. https://doi.org/10.1007/s10551-018-3999-7

"Whitelegg, J., "Haq, G., "Cambridge, H., & "Vallack, H. (2010). Towards a Zero Carbon Vision for UK Transport. In *[]*. https://www.osti.gov/etdeweb/servlets/purl/992702

Whitmarsh, L., Haxeltine, A., & Wietschel, M. (n.d.). *Sustainable transport visions: Expert and non-expert stakeholder perspectives on sustainable transport*.

Xu, L., Zhang, X., & Ling, M. (2018). Pro-environmental spillover under environmental appeals and monetary incentives: Evidence from an intervention study on household waste separation. *Journal of Environmental Psychology*, *60*, 27–33. https://doi.org/10.1016/j.jenvp.2018.10.003

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

VI. Paper 4

Virtue Signaling in Public or Private?

Behavioural Spillovers in Climate Change Action

January 2024

Department of Psychological and Behavioural Science
London School of Economics and Political Science
Houghton Street, WC2A 2AE, London, UK
f.koaik@lse.ac.uk, d.krpan@lse.ac.uk and p.dolan@lse.ac.uk

Virtue Signaling in Public or Private?

Behavioural Spillovers in Climate Change Action

1. Abstract

Understanding behavioural spillovers has gained momentum in the literature; however, the evidence must be more comprehensive on the psychological pathways underlying these ripple effects. The current study investigates whether engaging in pro-environmental behaviour in public (i.e., posting on social media) or private (i.e., pledging) results in spillover to carbon offsetting behaviour. It also tests normative beliefs and social expectations as moderators between engaging in the first behaviour and moral self-image. Then, it examines whether moral self-image is a mediator in the spillover process. The paper presents results from an empirical online experiment on participants (N=900) from the UK who were randomly assigned to either the public group (i.e., requested to like and share a climate change post on social media), private group (i.e., asked to pledge in private to recycle to combat climate change); and control group (i.e., asked to read a post on climate change with no action). Findings showed different spillover effects in each treatment and, most importantly, that moral self-image is a strong mediator that explains these effects and quantifies the moral credit model (Meritt et al., 2011). Participants in the private condition experienced consistent and positive behavioural spillovers compared to the other two conditions, and this spillover effect was mediated by increased moral self-image. At the same time, participants in the public condition experienced lower moral self-image (social expectations moderated this relationship), which indirectly explained the increase in carbon offsetting in the subsequent behaviour. Future research could explore additional moderators and mediators, test actual behaviours in the field, and look at the role of self-identity in increasing positive spillover effects.

Keywords: Pro-Environmental Behaviour, Virtue Signalling, Pledging, Behavioural Spillovers, Social Norms, Moral Self-Image

2. Introduction

"…**Liking a page on Facebook is the least we can do, but also the most we can do**". (Seth Meyers, Weekend Update, Saturday Night Live, September 22, 2012).

We live in times where social media is the key denominator of people's exposure, and social causes and developmental challenges are fundamental. For instance, when a climate change crisis needs actual behavioural change to make an impact, simply sharing or liking an online post is not considered a substantial or meaningful contribution. As it becomes easier for individuals to signal their virtues to their social network, it is crucial to ensure that people still take pro-social behaviours that are sustainable and impactful. Virtue signalling is a phenomenon that describes the willingness of people to engage in costless and simple acts that project their kind and virtuous support for a social cause (Noland, 2019). ***Does virtue signalling facilitate an easy and costless way for people to feel less willing to engage in actual pro-social actions?***

In fact, public pro-sociality posits a paradox (Kraft-Todd et al., 2023); performing pro-social behaviour in public is required to influence others. However, it could also position the individual's intention as mere virtue signalling. This is specifically relevant to early adopters of specific social causes or pro-environmental behaviours (for example, buying an electric car or eating a vegan diet…). Can the positive moral worth gained after publicly adopting these behaviours influence subsequent environmental actions?

By looking at virtue signalling from a behavioural lens, one aspect that raises questions is the possible behavioural spillovers (Dolan and Galizzi, 2015) and spillunders (Krpan et al., 2021) that encompass all the ripple effects of behavioural interventions. Positive behavioural spillovers occur when people are more likely to engage in more meaningful behaviours once they engage in a particular behaviour (Aronson, 1968; Bem, 1972; Festinger, 1957; Heider, 1958). One example is the Foot-In-The-Door (FITD) theory (Festinger, 1957), which states that individuals are more likely to stay consistent with their initial actions. More specifically, the FITD is relevant to virtue signalling. Many marketing and advertising campaigns assume that engaging people in an initial act of token support for a social cause will lead them to support more impactful and costly behaviours.

In contrast, negative spillover effects refer to people acting inconsistently with their previous actions. ***Moral licensing is*** one of the most counterintuitive types of negative behavioural spillovers (Meritt et al., 2010; Blanken et al., 2015; Dolan and Galizzi, 2015). It is a permitting negative spillover that describes how engaging in a prior good deed gives a license to perform an immoral or negative behaviour afterwards. The latter is very relevant to virtue signalling, where initial costless action (signalling support for a pro-social cause to one's social network) could give hypothetical moral credit to people and make it permissible for them not to engage in subsequent pro-social actions.

As this is a growing area of research, a limited range of psychological pathways and models where behavioural spillovers occur have been studied. From a theoretical perspective, a better understanding of when people are more likely to stay consistent with initial behaviour is needed. From a policy and practical perspective, capturing behavioural spillovers supports policymakers and practitioners in applying sustainable and long-lasting pro-environmental interventions to tackle climate change. This will ensure that program benefits are maximised (i.e., positive spillovers) and unintended negative consequences are eliminated (i.e., negative spillover effects such as moral licensing). Hence, the present empirical study examines pro-environmental behavioural spillovers in the domain of climate change action and tests whether the public observability of the initial behaviour leads to different subsequent behaviours. More specifically, it looks at engaging in virtue signalling compared to pledging in private on carbon offsetting. It examines the role of social expectations and moral self-image as a moderator and a mediator, respectively.

Climate Change Collective Action Problem

Behavioural change is crucial in tackling climate change (Whitmarsh and Lorenzoni, 2010). Often, it is narrowly conceived as individual-level action; however, it requires actions across different roles, such as individual households, members of the community, and employees at organisations, among others. In addition to household actions (such as recycling, sorting, and energy conservation), individuals can participate in publicly observable behaviour that signal their commitment to climate change cause to their social network. These behaviours could include participation in policy formulation through crowdsourcing, collective action and political action through protests and voting to show support for the social cause. Moreover, one of the common publicly observable behaviours entails engaging in climate change conversations and interacting with others on social platforms to increase awareness and

collection action (Withmarsh et al., 2021). In addition, according to the Integrated Framework for Encouraging Pro-Environmental Behaviour (IFEP) people may be motivated to engage in pro-environmental behaviour for hedonic reasons (e.g., because it is enjoyable), for gain reasons (e.g., because it saves money), or for normative reasons (e.g., because they think protecting the environment is the right thing to do) (Steg et al., 2014). Yet, as suggested above, many (but not all, as we will explain later) pro-environmental actions involve a conflict between normative goals on the one hand, and hedonic and gain goals, on the other hand (e.g., Lindenberg and Steg, 2007, Nordlund and Garvill, 2003, Samuelson, 1990, Steg et al., 2005, Steg and Nordlund, 2012).

Virtue Signalling

Recently, the topic of virtue signalling has been debated. Some argue that advertising and engaging with the public is a positive strategy to promote engagement and support for a social cause. On the other hand, critics perceive this phenomenon as being driven by pure self-desirability, impression management, and laziness (Bal et al., 2013; White & Peloza, 2009). Moreover, it might lead to negative spillover effects, specifically in the form of moral licensing. "Virtue Signaling" has a negative connotation as it refers to being self-centred and a desire for self-desirability and observability rather than pure altruism.

Another related phenomenon in the literature is Slacktivism, or "Social Media Activism" (Noland, 2019), which refers to performing a costless act that displays support for a social cause that requires minimal effort on social media. Social media has made this phenomenon more conspicuous. Projecting one's charitable behaviours or support to a social cause on Facebook by simply liking, sharing or posting a virtual post or image is very salient and reflects positively on one's self-identity and values. It is usually accompanied by a lack of eagerness to put more effort into making a meaningful impact and change (Noland, 2019).

The literature differentiates between token support and meaningful support (Kristofferson et al., 2014). Token support is when people engage in an act that shows support towards a social cause without investing time and effort, thus leading to no impact or behavioural change. However, meaningful support involves necessary actions where resources and effort are invested to achieve social change and positive impact. Kristofferson et al. (2014) looked at the mediating role of the social observability of the initial act (i.e., whether the support is made in public or private). When an initial behaviour is made in public, it is more likely to lead to moral licensing and disengagement from more meaningful acts. However, when an initial act is made in private, it will more likely lead to consistent behaviour and positive spillover.

Moreover, Noland (2019) looked into how the Theory of Planned Behaviour (TBD) (Ajzen, 1991, 2012) predicts slacktivism engagement (behavioural attitudes, subjective norm, perceived behavioural control, behavioural intention). A qualitative study showed that behavioural attitudes and intention are the main determinants of slacktivism engagement. In addition, Wallace et al. (2018) extended the work related to "Conspicuous Donation Behaviour (CDB) (Grace and Griffins, 2009) to Conspicuous Virtue Signaling (CVS). It examined whether engaging in CVS on Facebook predicts self-esteem, intention to donate offline, and intention to purchase counterfeit luxury products (engaging in unethical acts). However, it was all based on qualitative methodology and did not reflect actual behaviours.

Behavioural Spillovers

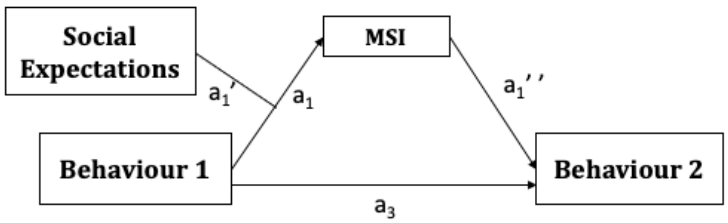
"No behaviour sits in a vacuum, and one behaviour can greatly affect what happens next" (Dolan and Galizzi, 2015). Dolan and Galizzi (2015) developed a conceptual framework with a tripartition of behavioural spillovers: promoting, permitting, and purging. Decades of research in social psychology argue that individuals have a solid drive to remain consistent with their beliefs, intentions, attitudes, and past behaviours (Gawronski and Strack, 2012). This drive towards "consistency," i.e., the tendency to prefer behaving consistently with one's prior beliefs and behaviours (i.e., positive behavioural spillovers), is debated with empirical evidence that a reverse phenomenon is at play (Monin and Miller, 2001). The phenomenon refers to moral balancing or negative behavioural spillovers; behaving in one direction leads to subsequent opposite behaviour. To inform the empirical debate on the striking contradiction between moral balancing and moral consistency, emerging research tests different moderators to explain when previous behaviours lead to
consistency versus balancing.

Although there is a debate in the literature on whether this similar phenomenon leads to positive or negative effects (Bal et al.,2013; White & Peloza, 2009), there is little empirical research to measure any spillover effect after engaging in publicly observable behaviour (i.e., virtue signalling) compared to private (e.g., pledging) in the pro-environmental domain. Does engaging in virtue signalling increase or decrease the likelihood of subsequently engaging in pro-environmental behaviour? What are the underlying and mediating behavioural phenomena at stake?

Social Expectations and Moral Self-Image: A new spillover pathway

This study will examine the roles of social expectations and moral self-image as a moderator and a mediator, respectively. The literature on behavioural spillovers, particularly moral licensing, presented the moral credit model with a hypothetical moral bank account where people accumulate credits when engaging in moral actions or offset immoral ones (Merritt et al., 2010; Miller and Effron, 2010). In this framework, we investigate moral licensing and test whether MSI mediates the effect of Behaviour 1 on Behaviour 2 (Figure 31) when MSI is moderated by social expectations (moderated mediation). In other words, how much does the increase in MSI due to social expectations when engaging in Behaviour 1 explain the negative spillover effect in Behaviour 2? Theoretically, this model focuses on one path where social expectations moderate the impact of Behaviour 1 on MSI and investigates whether moral licensing is more likely to occur due to the increase in MSI. Hence, if we feel more positive and moral after engaging in Behaviour 1 due to social expectations and norms, this increase in MSI will more likely lead us to transgress in Behaviour 2. We propose that when the change in MSI after engaging in Behaviour 1 is due to social expectations and norms rather than personal norms, then moral licensing is more likely to be experienced in Behaviour 2. The statistical diagram with the paths of the moderated mediation model is presented in Figure 31.

Figure 31. Statistical diagram of the moderated mediation model



This paper runs an empirical experiment to test whether 1) virtue signalling leads to moral licensing more than privately pledging pro-environmental commitment, and 2) social norms and moral self-regard explain the spillover effect.

Study Objectives

The study supports answering the following research questions:

1) Are people more likely to feel morally licensed when engaging in pro-environmental behaviour in public (observable on social media) compared to private (personal pledge)?
2) How does moral self-image influence subsequent behaviour?

The first objective is to test whether the public observability of pro-environmental behaviour influences behavioural spillovers differently. Second, we examine whether pledging or committing in private would increase consistency and a positive spillover effect will occur. Third, we look at how social expectations influence how participants perceive their moral self-regard after engaging in pro-environmental behaviour in public and private. Finally, we investigate moral self-image as a possible mediator of the spillover process. We predict that moral self-image indirectly explains the relationship between engaging in pro-environmental behaviour and subsequent behaviour.

3. Method

The study aims to answer the following research questions:

- Does signalling to the public our pro-environmental commitment (through liking or sharing a social media post) lead to a higher likelihood of moral licensing compared to not signalling?
- Is there a difference between signalling to the public or privately pledging a pro-environmental action on behavioural spillovers?
- Do Social Expectations moderate the effect of engaging in a pro-environmental action in public on Moral Self-Image (MSI)?
- Does MSI mediate the negative effect of engaging in the first behaviour on the second one (CO2 Offsetting)?

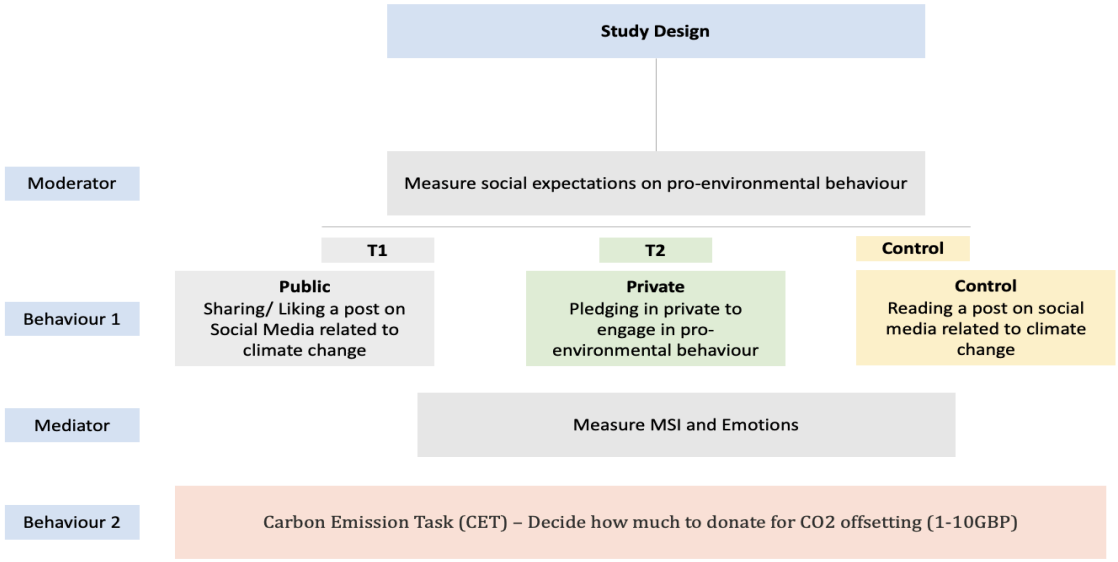
4.1 Participants and Procedure

The study was designed on Qualtrics, and participants were recruited from the UK via Prolific Academic. First, after completing the consent form, participants answered questions related to social expectations of climate change behaviour. Participants were asked about 1) the likelihood of them engaging in climate change behaviour, 2) their perceived empirical expectation (what they expect others to do), 3) their personal normative belief (what they think

is the right thing to do), 4) their normative expectation (what they expect other to think is the right thing to do) (Bicchieri, 2014).

Second, respondents were randomly allocated to a condition in which they were asked to engage in pro-environmental behaviour either (T1) in public through sharing/ liking a social media post related to #climatechange or (T2) in private through pledging/ committing to engage in pro-environmental behaviour, or (C) reading a passage on pro-environmental action online. Third, we measured respondents' moral self-image and affect once participants completed the first task. This was used as a mediator. Fourth, respondents were informed of a chance to enter a random draw and win 10 GBP. After that, they were presented with a task related to CO2 carbon offsetting. They were then asked to decide how much they would be willing to forego from the bonus amount to donate for CO2 offsetting. The amount they chose was used to buy CO2 certificates. They can choose from 0-10 GBP. Finally, we measured a range of covariates (social-demographic characteristics etc.). This study complies with the ethics policy and procedures of the London School of Economics (LSE) and has been approved by its Research Ethics Committee. Figure 32 presents the design of the study and the three conditions.

Figure 32. Study 4 Design



4.2 Measures

The three levels of the independent variable are *Public* (1), sharing/ liking a social media post related to #climatechange and *Private* (2), pledging/ committing to engage in pro-environmental behaviour to combat climate change, and *Control* (3), reading a passage on pro-environmental action online. Three continuous outcome variables were measured. The first dependent variable was affect measured via ten items answered on a 5-point Likert scale from 1 (very slightly or not at all) to 5 (Extremely), which required participants to report different affect when imagining engaging in the presented behaviour. The average score was reported on the five positive (inspired, enthusiastic, determined, proud and excited) and five negative affect (distressed, guilty, ashamed, nervous, and afraid)[16]. The final variable results from the difference between the average positive and average negative affect. The second dependent variable was Moral-Self Image (MSI)) as per Jordan et al. (2015) scale was measured via nine items answered on a 9-point Likert scale from 1 (1 (Much less *caring* than the person I want to be) to 9 (Much more *caring* than the person I want to be). The average of the reported scores on each of the nine dimensions was used to indicate their MSI. The third dependent variable is the average amount of donations entered by participants in GBP. The social expectations questions were measured and used as moderators.

4.3 Sample

Based on a-priori power analysis in G*Power 3.1 (Faul, Erdfelder, Lang, and Buchner, 2007), a minimum total sample of N= 900 (300 per condition) will be needed for a 95% probability of detecting a medium-size effect (d = 0.2) given a level of significance ($\alpha = 0.05$).

4.4 Hypotheses

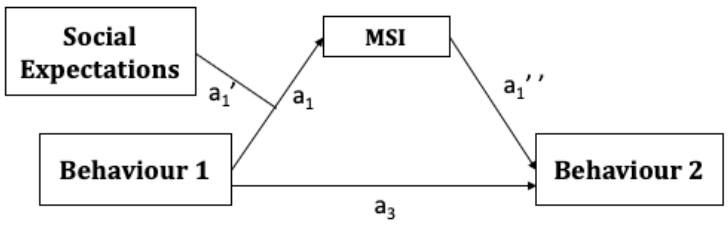
This study will test the following hypotheses:

H1: MSI between (T1), (T2) and (C) are significantly different —T1 and T2 will experience higher MSI compared to C.

[16] In the current study, we replaced three items ; alert, upset and scared from the short PANAS scale (Watson et al., (1988) with the following items proud, ashamed and guilty as they are relevant to the study. The items exhibited high inter-item correlations and significant overlap with other items. Correlation analyses revealed that these items demonstrated strong correlations (e.g., r=0.75, r=0.82). Furthermore, Cohen's d values were notably high (above 0.9), suggesting that these items contributed little unique variance to the overall measurement of affect.

H2: Average donations between (T1), (T2) and (C) are significantly different—T1 will donate less than C and T2 (compensation), and T2 will donate more than C and T1 (consistency).
H3: MSI negatively (positively)predicts the likelihood of donating for CO2 offsetting in T1 (T2) respectively.
H4: There is a significant moderated mediation with T1 (*sharing a post on social* media *compared to the other two groups)* as the independent variable, SE as a moderator, MSI as a mediator, and Behaviour 2 as the dependent variable. That is, the influence of engaging in T1 (*sharing a post on social* media *compared to the other two groups)* on MSI is moderated by SE (i.e., this influence is more significant for people who score higher on SE), and MSI, in turn, negatively predicts Behaviour 2 (donating CO2 certificates).
H5: There is a significant moderated mediation with T2 (*privately pledging compared to the other two groups)* as the independent variable, SE as a moderator, MSI as a mediator, and Behaviour 2 as the dependent variable. That is, the influence of engaging in T2 *privately pledging compared to the other two groups)* on MSI is moderated by SE (i.e., this influence is stronger for people who score higher on SE), and MSI, in turn, positively predicts Behaviour 2 (donating CO2 certificates).

Figure 33. Statistical diagram of the moderated mediation model



4. Results

Pilot Results

To ensure that participants would share or like the social media post on climate change, we ran a small pilot targeting 100 participants. In this pilot, participants were requested to share or like a post with #climatechange. Then, they were asked to report whether they liked or shared a post and answer an open-ended question to describe the post they found. The objective of the pilot is to test whether this method will lead people to use social media and respond to the task.

The results were positive: 70% of the sample confirmed liking or resharing a post and described the posts they read.

Study Results

905 individuals participated in a between-subject online experiment in exchange for 5 GBP per hour. Participants were recruited on Prolific Academic with three eligibility criteria: 1) country of residence in the United Kingdom, a minimum approval rate of 95%, and a minimum number of submissions of 100. The study was conducted on April 29, 2023. For data analysis, we use STATA/ SE 17.

Each participant was randomly assigned to one of the three conditions: T1 (n=298), T2 (n=303), or C(n=304). Of our sample of 905 participants, 502 were female, and 403 were male. Table 22 below summarises each treatment condition's mean and standard deviation for the dependent variables MSI and affect.

Table 22. Mean and Standard Errors (SE) (in parentheses) of dependent variables (Affect and MSI) by condition

		T1	T2	Control
Affect	**Mean**	2.28	2.34	2.50
	SE	(0.03)	(0.03)	(0.03)
MSI	**Mean**	5.08	5.31	5.07
	SE	(0.04)	(0.06)	(0.05)

Social Expectations

Social expectations of the behaviour of taking action to combat climate change present a big difference between what is personally approved and what is expected to be approved by others. In other words, on a scale from 0-100, the average reported personal normative belief is 85.61. Hence, most participants believe that acting towards reducing climate change and protecting the environment is the right thing to do. However, they believe that 46.13% of other people in their social network take pro-environmental actions, but they believe that 70% approve of this behaviour. This shows that there is general consent that pro-environmental behaviours to combat climate change are approved among their social network. However, there are lower expectations that other citizens would engage in these actions. In other words, although most approve of pro-environmental behaviour and believe others approve, they do not expect other citizens to take concrete actions (Table 23).

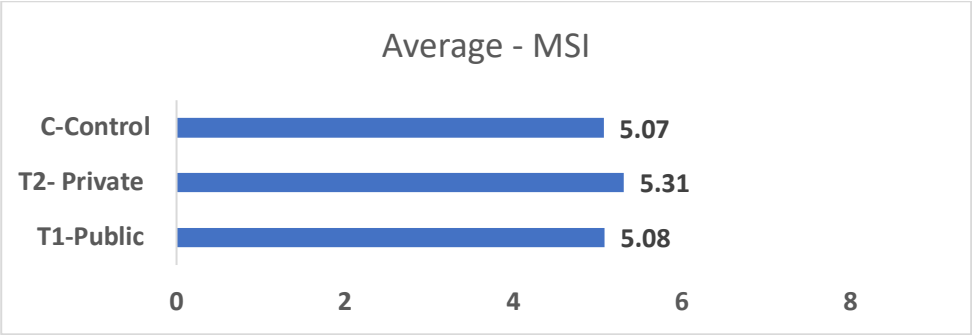
Table 23. Social expectations on the climate change behaviour

	Climate Change Taking pro-environmental behaviours to combat climate change
Empirical Expectations (0-100)	46.13 (0.84)
Personal Normative Belief (0-100)	85.61 (0.62)
Normative Expectations (0-100)	70.00 (0.69)

Hypothesis 1

A one-way ANOVA was performed to test if there was a significant difference in MSI across the three conditions. A statistically significant difference in MSI ($F = 5.61$, $p=0.0038$) at least between T2 (private condition) and the control group was detected.

Figure 34. Average MSI across conditions



Following Anova, we conducted pairwise comparisons of means for the variable MSI across three categorical factors: T1, T2, and C. A significant difference was observed between T2 (private) and the other 2 conditions with a mean difference of 0.2388 (95% CI: [0.0988, 0.3788], $p < 0.001$). This indicates that the mean MSI under T2 is significantly higher than T1 and C, suggesting that the presence of this condition positively impacts MSI. No significant difference was found between the groups for T1 with a mean difference of -0.1111 (95% CI: [-0.2523, 0.0302]). This interval includes zero, indicating that T1 does not have a statistically significant effect on MSI. Similarly, for the control condition, no significant difference was observed with a mean difference of -0.1284 (95% CI: [-0.2689, 0.0121]).

These findings indicate that while T2 significantly affects MSI, neither T1 nor C shows a statistically significant impact. Therefore, MSI significantly increased when participants

engaged in a private behaviour (where they are asked to pledge) compared to the control and public conditions (table 24).

Table 24. Pairwise Comparisons of Means for MSI

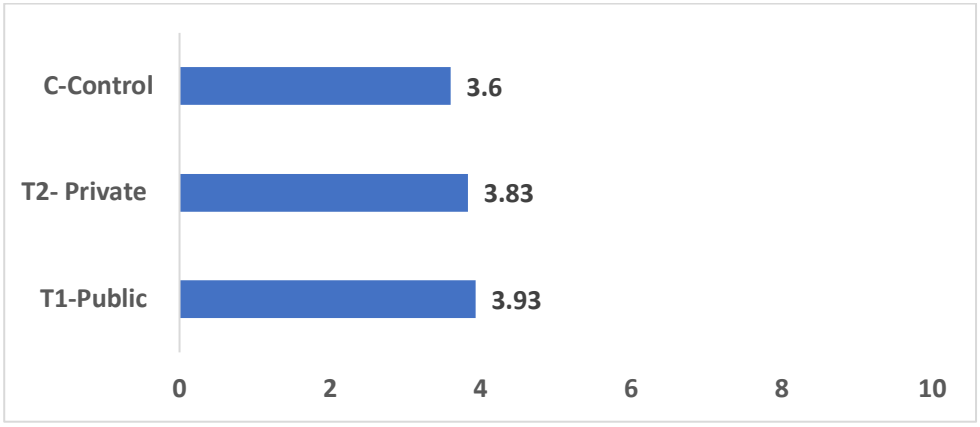
Conditions	Comparison	Mean Difference	Std. Error	95% Confidence Interval	p-value
T2- Private	1 vs 0	0.2388	0.0713	[0.0988, 0.3788]	< 0.001
T1 - Public	1 vs 0	-0.1111	0.0720	[-0.2523, 0.0302]	0.120
Control	1 vs 0	-0.1284	0.0716	[-0.2689, 0.0121]	0.060

Hypothesis 2

In this part, we look at the independent variable: the number of donations from 1 to 10 GBP. Figure 35 shows the average donations across the three conditions. An independent Anova test shows no significant difference between the *Public* condition (M=3.93), *Private* condition (M=3.83) and control (M=3.6) F (902) = 1.09, p=0.33.

Overall, there was no significant difference between the conditions. However, it shows that some participants faced negative spillovers (e.g., moral licensing), and others faced positive spillovers (e.g., foot in the door). This indicates the importance of further exploring the role of social expectations on the likelihood of different behavioural spillovers.

Figure 35. Average CO2 offsetting in conditions



Hypothesis 3

Moreover, to test what leads to the difference in donations, we computed simple linear regression models between donations and MSI. There is a significant effect between MSI and donations at a 5% significance level. The higher the MSI, the higher the donations by –0.035, SE= 0.011, t=2.98, p<0.003.

Hypothesis 4 – Moderated Mediation T1-Public, SE, MSI and Donations

Before computing the moderated mediation model, we ran a moderation analysis in the first place to see if moderation exists. We computed an interaction analysis with Social Expectations (SE)[17] as the moderator (it ranges from 0-100). SE represent how much participants expect others to engage in the given behaviour and approve of it. SE significantly moderates the effect of the *Public* condition*, i.e., sharing a climate change post on social media* on MSI (b=-0.0078, SE= 0.0036, t=-2.17, p=0.03). The higher the participants expected others to take and approve of pro-environmental behaviours to combat climate change, the less morally they perceived themselves when they shared a social media post on climate change (Table 25 and Figure 36).

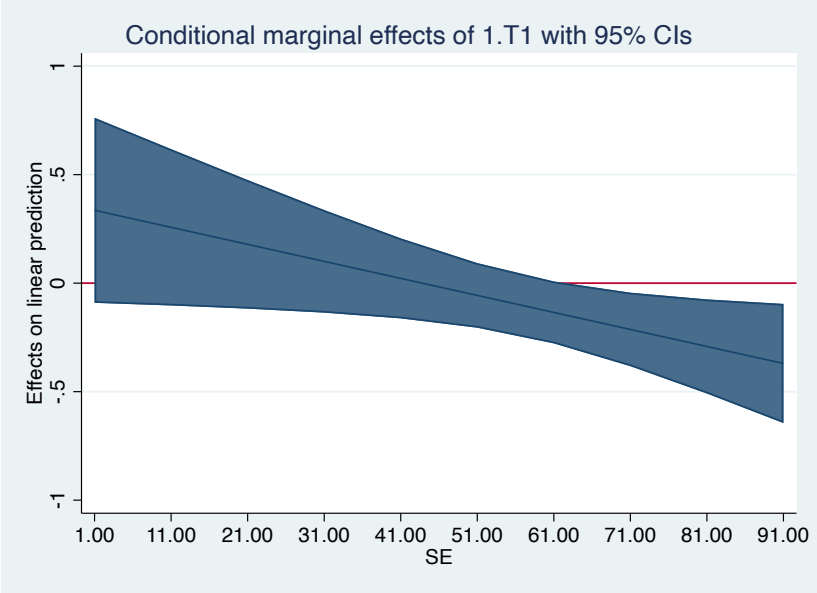
Table 25. The Effects of T1-Public x Social Expectations Interaction on MSI Variable (T2 and C baseline)

Variable	b	SE b	95% CI	t	p
DV = Moral Self-Image (MSI)					
Constant	4.04	0.12	[4.7088, 5.1844]	40.83	0
T1-Public	0.34	0.22	[-0.0895, 0.7767]	1.56	0.12
SE	0.004	0.0019	[0.0004,0.0081]	2.19	0.029
T1-Public x SE	-0.0078	0.0036	[-.01493, 0.0007]	-2.17	0.03

Note. R2 = 0.0094. Highlighted in grey are the significant effect at 5% significance level.

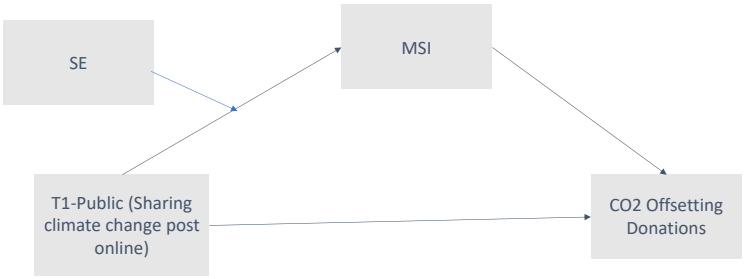
Figure 36: Conditional Marginal Effects of the Moderation Analysis

[17] Social Expectations (SE) is a continuous variable computed from the average of EE and NE (Empirical and Normative Expectations)



Based on Hayes (2018) Process Model 7 on *Conditional Process Modelling*, we ran a first stage *moderated mediation* (Baron and Kenny, 1986; Zhao et al., 2010; Preacher et al., 2007) as the focus is to estimate the indirect effect of the product of the independent variable (T1-Public) and the moderator (SE) on the dependent variable (CO2 Offsetting-Donation) through the mediator (MSI). The model is represented in Figure 37 in a conceptual diagram; the indirect (or mediated) of sharing a climate change post on social media on the amount of CO2 offsetting (i.e., donations) via MSI (the mediator) is assumed to be contingent on the level of SE (how much the pro-environmental behaviour is socially expected).

Figure 37. Conceptual Diagram

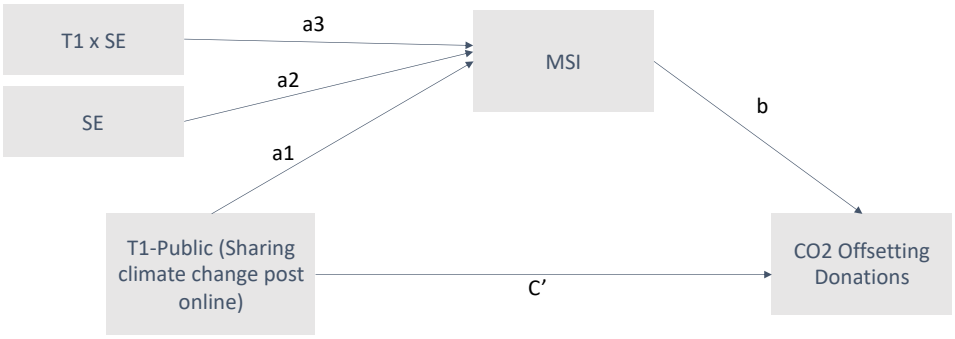


The statistical diagram (Figure 38) operationalises the conceptual model as two equations. The first equation (M, see next page) involves estimating regression parameters from a model where the proposed mediator (MSI) is regressed into the independent variable (X) – T1 Public (i.e.,

sharing a social media post on climate change), proposed moderator – SE (W), and the product of the independent moderator variables – T1 x SE (XW). In other words, the first model is consistent with a standard moderated multiple regression.

The second equation (Y, see below) involves the dependent variable – CO2 Offsetting Donation (Y) regressed into the proposed mediator (MSI) and the independent variable (X). In other words, the second model is consistent with a standard mediation regression.

Figure 38: Statistical Diagram



Equation (1)

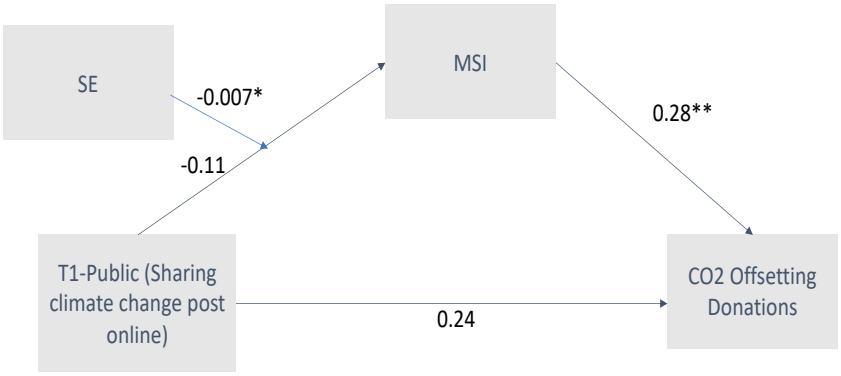
$$M = i_M + a_1*B1 + a_2*SE + a_{3*} (B1*SE)$$

Equation (2)

$$Y= i_y + c'*B1 + b*MSI$$

For Equation (1) and as presented in H3, there is a negative and statistically significant effect of the interaction term (moderator) B1*SE on MSI. In other words, there is evidence that the effect of sharing a climate change post on social media on MSI is conditioned on SE (b=-0.0078, SE=0.0036, t=-2.17, p<0.03, 95% CI [-0.014, 0.000]). So, as SE increases, the influence of sharing a public climate change post online (versus private and control) on MSI becomes more negative. For Equation (2), there is a positive effect of the mediator MSI on the relationship between sharing a climate change post and CO2 offsetting donations. This effect is significant at the 1% significance level (b= 0.28, SE=0.09, t=3.04; p<0.002, 95% CI [0.10; 0.46]. Moreover, the direct effect of sharing a climate change post on social media on CO2 offsetting donation is not significant (b=0.24, SE=0.20, t=1.24, p<0.216, 95% CI [-0.145; 0.644]). Figure 39 presents the conceptual diagram with the results.

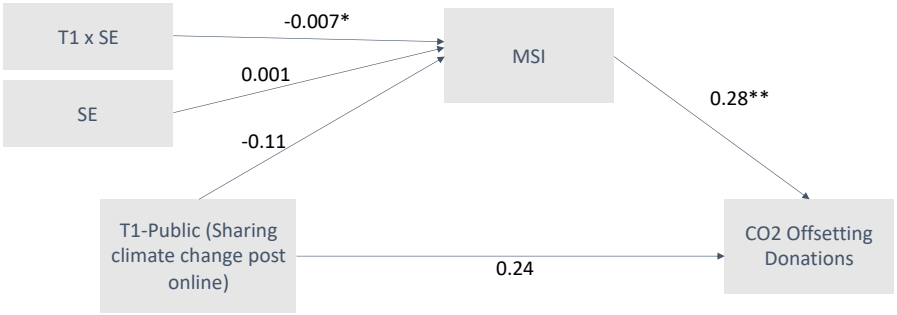
Figure 39. Conceptual Diagram: Results of the Moderated Mediation Model



In the model, we are specifically testing whether the indirect effect of sharing a public climate change post online (versus private and control) on CO2 offsetting donations through MSI is moderated by SE. To examine and reject the null hypothesis that there is no moderation effect by SE on the mediation of MSI between T1 and CO2 offsetting donation, we look at the *Index of Moderation Mediation* (IMM) in the STATA final output. IMM quantifies the degree to which the conditional indirect effect is a linear function of the moderator (Hayes, 2015). IMM is "a_3b_1", indicating the change in the moderating variable's conditional indirect effect (a3) per unit increase.

IMM = -0.007* (0.28) = -0.002. Figure 40 presents the statistical diagram with the results.

Figure 40. Statistical Diagram: results of the moderated mediation model



Based on 10000 bootstrap confidence interval testing, IMM is statistically significant (-0.0022, Bootstrap error=-0.003, 95% CI [-0.004, -0.000]. The negative effect of IMM indicates that as we move from lower SE to higher SE, the indirect effect becomes increasingly negative. In other words, the higher the SE, the more negative the influence of sharing a public climate change post online (versus private and control) on MSI becomes, thus decreasing the donations. **This explains that social expectations moderate the indirect effect of sharing climate change posts via social media (versus pledging in private or control, which is reading a post on social media without taking action) through MSI on donations.** After showing a

significant IMM (i.e., a significant moderated mediation effect), we will look more closely at the Conditional Indirect Effects (CIE). To quantify this indirect effect, Hayes (2015) express it in the following equation (3):

Conditional Indirect Effect (CIE) = $(a_1b_1 + a_3b_1W)$

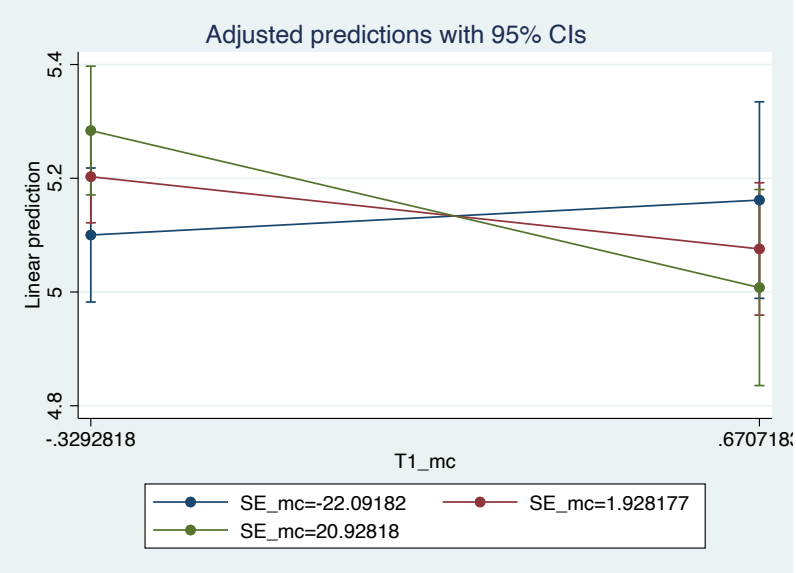
Using the values from our regression outputs, the function for the conditional indirect effect is f(W)= (-0.11+(-0.007*SE) *0.28. The model provides a different effect for various levels of SE. The model tests three levels of CIE (CIElow for -1sd, CIEmed for mean, and CIEhigh for +1sd). Table 26 highlights the three levels, the output of the Bootstrap results, and the 95% confidence interval.

Table 26. Moderated Mediation: Bootstrap Results

	SE level	Observed Coefficient	Bootstrap error	[95% Confidence Intervals]		Significance
CIElow (-1sd)	Low	0.017	0.031	-0.042	0.085	Non-significant
CIEmed (mean)	Medium	-0.035	0.023	-0.092	0.000	Significant
CIEhigh (+1sd)	High	-0.078	0.040	-0.170	-0.015	Significant

The results show that SE has a significant conditional indirect effect at SE levels higher than the mean. Moreover, the moderated mediation effect is significant only when SE is above the mean (Figure 41).

Figure 41. Moderated Mediation Results



Hypothesis 5: Moderated Mediation T2-Private, SE, MSI and Donations

For hypothesis 5, we look at the second treatment. Before running the moderated mediation model, we ran a moderation analysis to check if SE moderates the effect of T2 on MSI. Hence, we computed an interaction analysis with Social Expectations (SE)[18] as the moderator (it ranges from 0-100). SE represent how much participants expect others to engage in the given behaviour and approve of it. SE does not moderate the effect of *Private, i.e., pledging in private to engage in pro-environmental behaviours* on MSI (b=-0.005, SE= 0.0034, t=1.48, p=0.14). Therefore, social expectations do not influence the effect of pledging on MSI. **Regardless of what participants expect others to do or approve of, they experience higher MSI when privately pledging to engage in pro-environmental behaviour** (Table 27).

Table 27. The Effects of T2-Private x Social Expectations Interaction on MSI Variable (T1 and C baseline)

Variable	b	SE b	95% CI	t	p
DV = Moral Self-Image (MSI)					
Constant	5.07	0.12	[4.8341, 5.3226]	40.80	0
T2-Private	-0.060	0.212	[-0.4781, 0.3570]	-0.28	0.77
SE	0.000	0.002	[-0.0039,0.0040]	0.01	0.99
T2-Private x SE	0.005	0.0034	[-0.0016, 0.0117]	1.48	0.14

Note. R2 = 0.0160

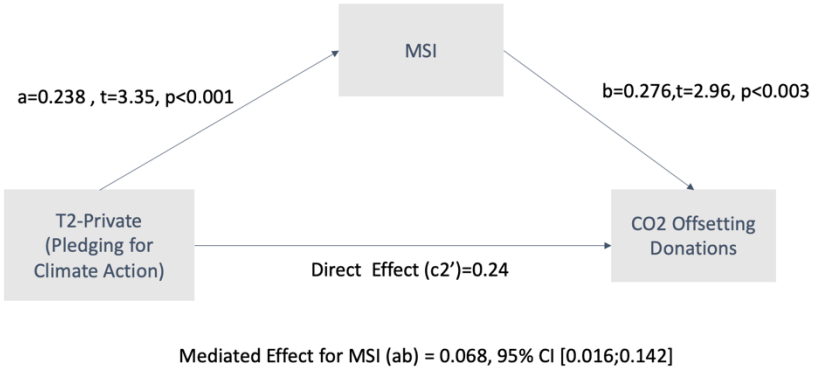
We no longer need to compute a moderated mediation model because we did not find a significant moderation effect between T2, SE and MSI. However, we still have to explore whether MSI indirectly explains the impact of T2 on the amount of CO2 offsetting. Therefore, we computed a mediation model to examine whether MSI indirectly explains the effect of pledging on the amount of CO2 offsetting.

Following Hayes' (2013) Macro Process visa bootstrapping method, the results showed a significant indirect effect between pledging for climate change action and CO2 offsetting by an increase in MSI. In other words, MSI positively mediates the effect of pledging for climate action on the amount of CO2 offsetting (b=0.068; 95% [0.016;0.142]) (Figure 42). **Participants experienced more positive moral self-regard after pledging compared to**

[18] Social Expectations (SE) is a continuous variable computed from the average of EE and NE (Empirical and Normative Expectations)

sharing/ liking a post on social media, and this increase explained the higher donations made to offset CO2 emissions. Therefore, pledging and committing in private increases the likelihood of staying consistent and increasing the positive spillover effect through an increase in MSI.

Figure 42. Mediation Model – Pledging, MSI, Donation



5. Discussion

As predicted in the framework, a significant difference exists between the reported MSI between the private condition and the other two conditions. Participants who shared a post on climate change on their social media experienced lower MSI. Unlike what we predicted in hypothesis 1, sharing a post on social media regarding climate change compared to pledging in private or just reading a post made participants feel less moral since they were reminded of the negative consequences of climate change on the environment. We hypothesised that participants would feel more positive about their moral self-image when sharing or liking a post (as this could make them feel that they have supported the cause or signal their virtues). However, MSI significantly increased among the second treatment group, where participants were asked to pledge in private and commit to engaging in pro-environmental behaviour. Pledging significantly increased participants' moral self-image, as they felt more positive after pledging and committing to take positive actions.

For hypothesis 2, when it comes to the second subsequent behaviour, which is the average donations made by participants for CO2 offsetting, there is no significant difference between the two conditions. This highlights the need to explore the underlying behavioural spillover

effect further and understand which group experienced compensation or consistency compared to the control group. In fact, as predicted in hypothesis 3, MSI significantly explained the relationship between taking the first behaviour and the subsequent one related to the donations made for CO2 offsetting. The latter indicates the effect of moral self-regard on the spillover effect and highlights the role that moral self-image plays in participants' subsequent decisions. It further explains the Moral Credit Model (Merritt et al. 2010, Miller and Effron 2010), in which people accumulate moral credits in a hypothetical moral bank account. Moral credits are used to either engage in moral actions or offset immoral actions (Merritt et al. 2010, Miller and Effron 2010).

In addition, when it comes to hypothesis 4, the moderated mediation model tested the relationships between social expectations, MSI, and the first treatment, i.e., public behaviour related to climate change, such as sharing or liking a post on social media. The positive and significant results showed that social expectations explain when people feel licensed versus consistent with their previous actions. Social expectations significantly **negatively** moderated the effect of sharing a social media post related to climate change on MSI, and then MSI **positively** predicted the effect of posting on CO2 offsetting. In other words, sharing a post on social media decreases the reported moral self-regard only when people expect others to engage and approve of pro-environmental behaviour (high social norms). This decreased moral self-image indirectly explained the amount donated for CO2 offsetting. **This moderated mediation model explains the negative spillover effect, where individuals experienced a decrease in their moral credit (through the low MSI), which led to a decrease in donation.**

Finally, for hypothesis 5, when it comes to pledging in private and committing to engage in pro-environmental behaviours to combat climate change, participants experienced positive moral self-image (i.e., increased in their moral credit) regardless of the social expectations (social norms do not moderate or influence this relationship). In addition, the increase in moral self-image significantly explains the increase in the amount of donations for CO2 offsetting. **As predicted, this explains that pledging in private leads to a positive spillover effect and consistency in subsequent behaviour. The positive spillover is explained through the increase in the reported moral self-image.**

6. Limitation

The limitation of the present study is that it is an online experiment rather than a field experiment on social media. However, the pilot study we ran at the beginning helped us ensure that the participants would visit their social media accounts and engage in the requested behaviour. Future research will then need to probe the mechanism using actual behaviour that participants get to perform on social media. The second limitation is that we are unable to track whether participants liked a post or shared it, and in the control, whether they read a post on social media.

7. Conclusion

This study makes different contributions to several literature strands. First, the experiment showed two different behavioural spillover effects. Second, it explained that moral self-image could empirically explain the spillover mechanisms. Third, the first treatment presents new evidence that virtue signalling on social media for climate action leads to a positive spillover effect rather than moral licensing. To contribute to the debate in the literature that discusses whether signalling one's virtues by showing support for a social cause would cause moral licensing (transgression) or consistency. In this research, we showed that virtue signalling by sharing a post on social media led to a decrease in moral self-image (moderated by social norms), decreasing their likelihood of donating for CO2 offsetting. However, the latter is only true when pro-environmental behaviours are socially expected (i.e. when individuals expect others to engage in pro-environmental behaviour and approve of it). So, it is crucial to instil these positive social norms while increasing the number of posts on social media. In this study, we also tested positive spillover effects; taking pro-environmental action (e.g., pledging) in private led to an increase in moral self-image (regardless of the social expectations) and eventually to consistency in donating for CO2 offsetting. Future research needs to replicate these findings in different countries to see the effect of virtue signalling in public and pledging in private on behavioural spillovers.

8. References

Ajzen, I. (2012). The theory of planned behaviour. In *Handbook of Theories of Social Psychology* (Vol. 1, pp. 438–459). https://doi.org/10.4135/9781446249215.n22

Aronson, E. (1969). The Theory of Cognitive Dissonance: A Current Perspective11 This article and all the original work reported in it were supported by a grant from the National Science Foundation. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 4, pp. 1–34). Academic Press. https://doi.org/10.1016/S0065-2601(08)60075-1

Bal, A., Archer-Brown, C., Robson, K., & Hall, D. (2013). Do good, goes bad, gets ugly: Kony 2012. *Journal of Public Affairs*, *13*. https://doi.org/10.1002/pa.1475

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bem, D. J. (1972). Self-Perception Theory11Development of self-perception theory was supported primarily by a grant from the National Science Foundation (GS 1452) awarded to the author during his tenure at Carnegie-Mellon University. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 6, pp. 1–62). Academic Press. https://doi.org/10.1016/S0065-2601(08)60024-6

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms* (pp. xvi, 260). Cambridge University Press.

Bicchieri, C. (2016). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Blanken, I., van de Ven, N., Zeelenberg, M., & Meijers, M. H. C. (20140602). Three attempts to replicate the moral licensing effect. *Social Psychology*, *45*(3), 232. https://doi.org/10.1027/1864-9335/a000189

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, *47*, 1–16. https://doi.org/10.1016/j.joep.2014.12.003

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behaviour Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140. https://doi.org/10.1177/001872675400700202

Gawronski, B., & Strack, F. (2012). Cognitive consistency is a fundamental principle of social information processing. *Cognitive Consistency: A Fundamental Principle in Social Cognition.*, 1–16.

Grace, D., & Griffin, D. (2009). Conspicuous donation behaviour: Scale development and validation. *Journal of Consumer Behaviour*, *8*(1), 14–25. https://doi.org/10.1002/cb.270

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1–22. https://doi.org/10.1080/00273171.2014.962683

Hayes, A. F. (2018). Partial, conditional, and moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, *85*(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100

Heider, F. (1958). The naive analysis of action. *The Psychology of Interpersonal Relations.*, 79–124. https://doi.org/10.1037/10628-004

Kraft-Todd, G. T., Kleiman-Weiner, M., & Young, L. (n.d.). Virtue Discounting: Observability Reduces Moral Actors' Perceived Virtue. *OPEN MIND*.

Kristofferson, K., White, K., & Peloza, J. (2014). The Nature of Slacktivism: How the Social Observability of an Initial Act of Token Support Affects Subsequent Prosocial Action. *Journal of Consumer Research*, *40*(6), 1149–1166. https://doi.org/10.1086/674137

Krpan, D., Galizzi, M. M., & Dolan, P. (2019). Looking at Spillovers in the Mirror: Making a Case for "Behavioral Spillunders". *Frontiers in Psychology*, *10*, 1142. https://doi.org/10.3389/fpsyg.2019.01142

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral Self-Licensing: When Being Good Frees Us to Be Bad. *Social and Personality Psychology Compass*, *4*(5), 344–357. https://doi.org/10.1111/j.1751-9004.2010.00263.x

Miller, D. T., & Effron, D. A. (2010). Psychological License. In *Advances in Experimental Social Psychology* (Vol. 43, pp. 115–155). Elsevier. https://doi.org/10.1016/S0065-2601(10)43003-8

Monin, B., & Miller, D. (2001). Moral Credentials and the Expression of Prejudice. *Journal of Personality and Social Psychology*, *81*, 33–43. https://doi.org/10.1037//0022-3514.81.1.33

Noland, A. (2017). *Social Media Activists: Analyzing the Relationship between Online Activism and Offline Attitudes and Behaviors*.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behavioral Research*, *42*(1), 185–227. https://doi.org/10.1080/00273170701341316

Wallace, E., Buil, I., & Chernatony, L. (2020). 'Consuming Good' on Social Media: What Can Conspicuous Virtue Signalling on Facebook Tell Us About Prosocial and Unethical Intentions? *Journal of Business Ethics*, *162*. https://doi.org/10.1007/s10551-018-3999-7

White, K., & Peloza, J. (2009). Self-benefit versus other-benefit marketing appeals: Their effectiveness in generating charitable support. *Journal of Marketing*, *73*(4), 109–124. https://doi.org/10.1509/jmkg.73.4.109

Whitmarsh, L., & Lorenzoni, I. (2010). Perceptions, behaviour and communication of climate change. *Wiley Interdisciplinary Reviews: Climate Change*, *1*, 158–161. https://doi.org/10.1002/wcc.7

Whitmarsh, L., Poortinga, W., & Capstick, S. (2021). Behaviour change to address climate change. *Current Opinion in Psychology*, *42*. https://doi.org/10.1016/j.copsyc.2021.04.002

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

VII. General Discussion

1. Overview of Findings

Researchers and practitioners need to understand and predict the different possible behavioural spillovers when discussing the sustainability of behavioural change interventions. To do so, this thesis has aimed to investigate the link between social norms and moral self-image on behavioural spillovers, specifically the moral licensing effect. The main and overarching contribution of this thesis is bridging these different literature strands into one theoretical and empirical model. In other words, linking social norms, moral identity, and behavioural spillovers into one model and testing it in different applications and domains to answer two main research questions: **1) How do social expectations influence individuals' moral self-image and affect?** and **2) How does moral self-image influence subsequent behaviour (behavioural spillovers)?** To address the questions, we first looked at how social expectations moderate the effect of engaging or not engaging in behaviours on moral self-image and experienced affect. Second, we explored whether the difference in the experienced affect and moral self-regard influences subsequent behaviours (behavioural spillovers).

1.1 Summary of the Findings

The findings and contributions of this thesis were achieved through four empirical studies and papers that tested the main research questions (Figure 43).

Figure 43. Overview of the four empirical studies in this thesis

In the first part of the thesis, we looked at social norms and expectations' effect on how we feel and perceive our moral self-image. To do so, we suggested a new theoretical framework that differentiates between behaviours that are socially expected through social norms versus others that are exceptions. We successfully showed that people experience different affect and moral self-regard depending on their social expectations about what others approve of. We tested the effect of standing in line compared to getting an electric car on affect and moral self-image. These findings raise interesting implications for interventions in public policy based on social norms (discussed in the section below).

In the second paper, we replicated the same experiment and added another behaviour to test the generalizability of the framework. The behaviour was related to the COVID-19 rapid test. We also tested the spillover effect on subsequent behaviour related to donating to the UK Cancer Research Centre. We presented a new moderated mediation model that tested the effect of social expectations (moderator) and moral self-image (mediator) on the main effect of behaviour 1 (taking the rapid test) on behaviour 2 (donation). The model explained the moral licensing effect and the underlying influence of social expectations and moral self-image on this compensatory behaviour.

In the third paper, we replicated the moderated mediation model on another domain related to pro-environmental behaviour (i.e., using the train instead of the plane to reduce emissions). The findings further presented that moral licensing was a negative behavioural spillover for

participants in the train condition. More specifically, participants who imagined taking the train and had higher social expectations around this behaviour experienced higher moral self-image and lower donations for CO2 offsetting.

In the fourth paper, we explored the concept of virtue signalling in the domain of climate action to examine whether virtue signalling would lead to negative spillover effects compared to pledging in private. The findings differed; however, we captured interesting behavioural spillover effects with moral self-image as a mediator. First, participants in the virtue signalling condition experienced lower moral self-image after sharing/ liking a post on social media. Social expectations moderated this decrease in moral self-image and indirectly explained the decrease in donations compared to the other conditions. The latter was also successfully examined using our suggested moderated mediation model. Second, participants in the pledging (private) condition experienced a positive spillover effect. They behaved consistently with previous actions, and donations were significantly mediated by increased moral self-image after pledging. However, the increase in moral self-image was not moderated by social expectations. This means that, regardless of what other people in the social network do or approve of, participants experienced a positive moral self-image compared to the other conditions when pledging in private.

1.2 Summary of Methodological Findings

The present thesis contributes to the literature as it builds on existing scales and frameworks from the literature, including social norms (Bicchieri, 2006, 2016; Bicchieri and Mercier, 2014), moral self-image (Jordan et al., 2015), and behavioural spillovers (Dolan and Galizzi, 2015; Krpan et al., 2019; Galizzi and Whitmarsh, 2019; Nash et al., 2017) such as moral licensing (Blanken et al., 2015) and moral credit model (Merritt et al., 2010).

First, in addition to the research findings outlined in the next section of this chapter, one of the present research's main strengths is how it builds on the existing literature from three different literature strands. The behavioural science literature has widely used the concept of social norms as a tool to induce behavioural change, mainly informing people about what others are doing. In this thesis, we delve more into diagnosing social norms and, more specifically, measuring different types of social expectations (Bicchieri, 2006; 2016). This research offers

an innovative tool to use the research methods of Bicchieri (2006) in diagnosing norms in behavioural experiments to test behavioural spillovers.

Second, the existing research in the literature related to moral self-image and its scale (Jordan et al., 2015) seems distant from moral balancing (Monin and Miller, 2001; Zhong and Liljenquist, 2006; Sachdeva et al., 2009; Jordan et al., 2011, 2015) and spillover research (Dolan et al., 2015; Krpan et al., 2019). Whereas several studies showed moral balancing effects, more is needed about the empirical validity of the moral credits model. Across the empirical studies in this thesis, we leveraged an existing construct by Jordan et al. (2015) that measures the malleable state of an individual's moral self-image and tested whether it could explain the moral credit model (Merritt et al., 2010).

2. Research Implications

In sum, the present thesis offers several interesting findings with theoretical and practical/ policy implications. We will summarise each in the following section.

2.1 Theoretical Implications

In addition to exploring and testing different behavioural spillovers in empirical settings (Galizzi and Whitmarsh, 2019), this thesis shows that examining and measuring behavioural spillovers is important. More specifically, this research produces several findings, which are presented in the previous section. However, to better understand the value of the present findings, it is necessary to understand their implications for the related theoretical accounts or literature strands. To do so, we will present theoretical implications followed by practical/ policy implications. For the theoretical implications, we will link the implications back to the literature gap and respective research strands in the literature (Table 28)

Table 28. Theoretical Contribution of the Thesis Mapped to Literature Gaps

	Theoretical Contributions of the Thesis	**Literature Gaps**	**Related Literature Strands**
a	**Effect of social norms and social expectations on moral self-image and affect:** ***Theoretical framework (Figure 4, Chapter 1)***	Social norms are explored as a tool to change behaviour, and no research looked at how social norms influence moral self-perceptions and affect.	- Social norms and expectations - Morality and Moral Self-Image

	- Social expectations lead to a significant difference in the self-reported experienced affect and moral self-image. - Social expectations (as a contextual effect) moderate the effect of engaging versus not in behaviour on the experienced moral self-image. In other words, with different social expectation levels, there is a significant difference in moral self-image when engaging or not in certain behaviour.	The effect of behaving in congruence with a social norm (interdependent behaviour) versus personal normative beliefs (interdependent behaviour) has not been investigated. Is there a difference in interdependent and independent behaviours on moral self-image and affect?	
b	**Moral Self-Image (MSI) as a mediator that explains behavioural spillovers (particularly moral licensing and moral consistency)** - MSI was convincingly presented as a mediator that indirectly explains behavioural spillover effects. - Empirically tested the moral credit model in behavioural spillovers and quantified the moral credit. - Whereas MSI predicts the occurrence of behavioural spillovers, affect did not. Although, affect significantly differed across conditions, however, they did not mediate any the effect of behaviour 1 on behaviour 2. Affect does not mediate or indirectly explain moral lceinsing effects.	No study looked at how moral self-image (which is dynamic and context-dependent) influence behavioural spillovers (particularly moral licensing and moral consistency). No study looked at quantifying the moral credit to explain moral licensing and cleansing.	- Morality and Moral Self-Image - Behavioural Spillovers
c	**Effect of social norms on behavioural spillovers through moral self-image:** ***Moderated Mediation Model (Figure 5, Chapter 1)***	What influence consistency and compensatory behaviour? When are people more likely to stay consistent with their initial action?	- Social Norms - Morality and Moral Self-Image - Behavioural Spillovers

	- The effect of moral self-image on behavioural spillovers is moderated by social expectations when the behaviour is performed in public. - Social expectations do not moderate the effect of pledging in private on MSI and subsequent behaviour.		

a. Effect of Social Norms on Moral Self-Image and Affect: Theoretical framework

So far, the growing literature on social norms focuses on differentiating between descriptive and injunctive norms. Descriptive norms or empirical expectations refer to common behaviours and describe what people do (Cialdini et al., 1990; Reno et al., 1993; Cialdini et al., 2006; Bicchieri et al., 2014), whereas injunctive norms or normative expectations refer to approved behaviours. Injunctive norms could take two forms: prescriptive norms to define the right behaviour (what is "ought to") and proscriptive norms to explain the wrong behaviour (what is "ought not to") (Janoff-Bulman et al., 2009). However, little is known about how conforming to different social norms influences how we perceive our actions or moral self-regard. In this thesis, we contributed to the literature on social norms, where we suggested a new theoretical framework that looks at the effect of "Expectation" versus "Exception" on how affect and self-perceived morality. The difference between the expectation and the exception is the descriptive norm or empirical expectation (describing how much we expect others to undertake or engage in the behaviour) (Cialdini et al., 1990; Reno et al., 1993; Cialdini et al., 2006; Bicchieri et al., 2014). In addition to the contribution of this framework to understanding behavioural spillovers, it also provides a novel contribution to the social norms' literature strand. For instance, in the pro-environmental behaviour domain, not engaging in the exception (i.e., positive behaviour with a high injunctive norm but a low descriptive norm) leads to neutral, experienced affect.

b. Moral Self-Image Explains Moral Credit Model

An important contribution of this thesis is related to empirically testing the moral credit model (Merritt et al., 2010) through moral self-image (Jordan et al., 2015). The literature on moral behaviour posits evidence that, on the one hand, people engage in morally positive behaviours. On the other hand, there is also evidence that people often engage in immoral acts (Griffin et

al., 2015). In this sense, acting in a moral or immoral behaviour represents a conflict of competing motivations; one motive is to benefit from the selfish benefits of engaging in immoral acts, and another motive strives to maintain a positive moral self-image (Baumeister and Alghamdi, 2015; Woolley and Fishbach, 2015; Mazar et al., 2008; Sheldon and Fishbach, 2015). These competing motivations are called moral balancing, which could be translated into either moral licensing or moral compensation (Merritt et al., 2010; Effron and Conway, 2015; Merritt et al., 2010; Mullen and Monin, 2016). To date, the theoretical account widely explaining the moral balancing effects is the moral credit model (Merritt et al., 2010), which presents the idea of a hypothetical moral credit bank account. However, it is a relatively young research stream and has yet to be more empirically tested to quantify these moral credits in an empirical setting. Hence, the present research constitutes the most in-depth exploration of moral self-image as a mediator that indirectly explains the effect of engaging in the first behaviour on the second one.

This research shows that when moral self-image decreases after engaging in a behaviour, individuals are more likely to compensate by engaging in a positive behaviour afterwards. However, when moral self-image increases after engaging in positive initial behaviour, this increase leads to moral licensing or transgression in subsequent behaviour, significantly when social expectations and social norms moderate the change in moral self-image. This is the case of moral licensing captured after taking the rapid test or taking the train instead of the plane… In these instances, social expectations play a role in the increase in moral self-image (i.e., when people expect others to engage in the behaviour and approve of it, they experience a higher moral self-image). Therefore, the increase in moral self-image is not purely intrinsic but mainly influenced by social norms and the social desirability effect. In paper 4, we provide evidence that moral self-image increases after pledging in private, but this increase is independent of social expectations. In other words, regardless of what other people do or approve of, participants are more likely to experience positive moral self-regard after pledging in private. So, foot-in-the-door and cognitive dissonance (Freedman and Fraser, 1966; Fointiat, 2006) prevail when the behaviour is conducted in private with low public observability, leading to positive and consistent subsequent behaviour. Understanding the underlying mechanisms behind moral self-image makes it easier to predict which behavioural spillovers will take place subsequently.

c. Moderated Mediation Model of Social Norms and Moral Self-Image

Another theoretical contribution of this thesis is that it constitutes one of the first research that looks at a comprehensive and nuanced examination of the mechanisms behind the effect of social norms, morality, and moral licensing. Previous research focused on the moderators of moral credentials (Merritt et al., 2012; Merritt et al., 2010; Miller and Effron, 2010), such as construal level, identification with the social cause, cost of initial behaviour, observability of the behaviour, among others (Blanken et al., 2015). However, to the best of our knowledge, studies still need to provide an in-depth examination of the mechanism behind moral licensing in a moderated mediation model. Krpan, Galizzi and Dolan (2019) tested a similar statistical model to examine future spillunders (Krpan et al., 2021). More specifically, the focus was on how general self-efficacy moderates the influence of expected exercise on intellectual performance.

This thesis also tests a similar statistical model related to moderated mediation analyses. **We successfully test how social norms moderate the influence of positive moral behaviour on moral self-image and how moral self-image indirectly explains the licensing effect in subsequent behaviour.** This finding contributes to the literature on testing and examining underlying mechanisms and pathways behind behavioural spillovers. It presents a methodological way to account for social norms and moral self-image. This model could be replicated in other contexts using the same scales. Moreover, different types of moderators could be tested to see if moral licensing could also be explained more.

2.2 Practical and Policy Implications

a. Collective Action Problems and Pro-Environmental Behaviours

In 2015, United Nations member states committed to achieving an ambitious global agenda with 17 Sustainable Development Goals (SDGs) (Whitmarsh, 2023). The comprehensive agenda recognises poverty, corruption, transparent institutional governance, inequality, environment, and diversity. One of the common characteristics of all these developmental challenges is the need for every individual to coordinate and contribute to achieve collective benefit and social change. Understanding collective action problems and their underlying behavioural motives is key to helping governments and international organisations realise the SDGs and build sustainable and happy societies. This requires zooming in to understand the minuscule individual choices that people make daily and instilling behavioural change, which

will help in zooming out and cascading these changes into collective and sustainable transformation (leading to a snowball effect to solve collective action problems) (Whitmarsh, 2023). Many of these challenges have always been tackled at the macro level, focusing on institutional, legal, structural, and political, with little focus on psychological motives. So far, most efforts have been invested in reforming policies and regulations, fostering international partnerships and agreements, developing technical solutions and innovative models, designing incentives and economic tools, and promoting strategic communication and informative campaigns. These indispensable regulatory tools and technical solutions are beneficial in transitioning towards more inclusive, transparent, green, diverse, productive, healthy and happy societies. However, "[…] achieving the sustainable development goals requires a critical understanding of how people make decisions and act on them, how they think about, influence, and relate to one another, and how they develop beliefs and attitudes" (UNDP, 2016, pp. 1–2). In addition to understanding behaviour and designing behaviourally informed strategies, it is equally important to keep an eye on the "ripple effect", behavioural spillovers and spillunders (Dolan and Galizzi, 2015; Krpan et al., 2019) from one domain to the other or within the same domain.

Within this policy context, the findings of the present thesis highlight the importance of psychological mechanisms behind social norms and expectations. Solving such complex challenges requires understanding the norms that encourage collective responsibility. The assumption that people will act sustainably when they perceive others doing the same does not hold. In some cases, if the motive is to engage in the act without intrinsic motivation to contribute to the social or pro-environmental cause, then people are more likely to transgress. Usually, the individual cost of contributing to these collective action problems (e.g., taking the train instead of a plane to save emissions, buying an electric vehicle instead of a conventional one, or recycling at home) is perceived as higher than the individual benefit. However, the findings of this research showed that participants with high social expectations experienced higher moral self-image after engaging in positive and moral behaviour. The increase in moral self-image is somehow a moral credit (Merritt et al., 2001), which is considered an individual benefit. Therefore, people are more likely to perceive these moral credits as a license to transgress afterwards. Overall, this research shed light on capturing the individual motives, experienced affect, and affect after engaging in pro-social behaviour. This is particularly relevant for policymakers and practitioners devising policy programs intending to achieve long-lasting behavioural change.

b. Social Norms and Experienced Affect

The literature on normative behaviour focuses primarily on defining what constitutes social norms. Bicchieri (2006, 2016) examines in her research the nature and evolution of social norms, how to measure and diagnose them, and what tools to use to influence social norms and achieve social change. In this thesis, we primarily adopt her definition of social norms and her approach to diagnosing social norms through normative and empirical expectations (Bicchieri, 2014). However, this present thesis contributes to how people feel and perceive themselves and their actions after engaging in behaviours congruent with social expectations. For policymakers, this offers a new way of looking at the normative behaviours in their societies, whether they are moral or immoral in nature. To further explain, the empirical studies in this thesis focus on positive behaviours that contribute to collective good (e.g., pro-environmental behaviour to combat climate change, green transport, rapid testing during COVID-19, and standing in line). When governments and policymakers have ambitious visions and national objectives to achieve Sustainable Development Goals (UN, 2013), understanding how people feel and perceive their morality after engaging in normative behaviours is crucial. The findings of this research showed that people do not feel guilty, ashamed or nervous when not engaging in positive and moral behaviour that is not descriptive (e.g., buying an electric car). In this case, communicating information that increases their likelihood of feeling guilty or instilling negative feedback about inaction might lead to changes in experienced affect and morality, leading to behavioural change.

c. Normative Interventions and Unintended Consequences

Social norms interventions have also been widely used to apply behavioural insights to solve systemic policy challenges. Numerous behavioural interventions used social norms messaging to induce the desired change by communicating descriptive norms (Goldstein et al., 2008; Cialdini et al., 1990). The findings from this literature shed light on the potential unintended consequences of using normative messaging to induce sustainable behavioural change. This is particularly relevant regarding collective action problems that require small individual behavioural change to achieve a long-term impact. For instance, individuals who imagined taking the train only experienced a positive moral self-image when social expectations were high, and this increase in moral self-image led to a decrease in donations for CO2 offsetting.

For policymakers, this means highlighting to people what others are doing or approving of, which might lead to licensing effects and, eventually, unintended consequences.

d. Behavioural Spillovers: Unintended Consequences of Behavioural Interventions

In the past decades, applying behavioural insights into policies and programs has been gaining huge momentum (Dolan et al., 2010). Behavioural interventions offer governments a powerful tool for improving public policy outcomes by understanding and influencing citizen behaviour. By incorporating insights from behavioural economics, governments can design more effective policies that align with how individuals make decisions. These applications of behavioural interventions enhance the efficacy of governmental initiatives, fostering positive behavioural change and improving overall public welfare. As this field matures, there is a growing focus on the long-term sustainability of these behavioural interventions and lasting behavioural change. For policymakers and behavioural science practitioners, capturing these behavioural spillovers could maximise the program or policy's benefits and reduce unintended consequences. Within this context, the findings from the present research contribute to a better understanding of the mechanisms behind possible behavioural spillover effects, specifically moral licensing. From a policymaking and practical perspective, moral licensing seems to be one of the most counteractive behavioural spillovers. This is because governments and organisations are devising interventions to improve ethical and pro-social behaviour that have collective benefit; ensuring sustained positive behavioural change is essential. For instance, if a company emphasises its commitment to sustainability, it should be cautious about potential moral licensing effects that might lead consumers or employees to justify less environmentally friendly choices.

The findings of this current research shed light on the impact of social norms on moral licensing and, crucially, highlight the flexibility of moral self-image. This research underscores to policymakers the dynamic nature of individuals' moral self-regard, emphasising that an individual's actions may lean towards morality or immorality based on motives, drives, and other contextual factors in the social environment. These insights highlight the need for nuanced and context-specific policymaking to address and influence individuals' pro-social and ethical behaviours effectively.

3. Limitations

Ecological Validity

It is essential to take note of the limitations of the present research. One of the main limitations of this thesis is the external ecological validity and robustness (Brewer, 2000). Ecological validity presents the level to which a psychological construct tested in the lab can occur in the real world. Historically, researchers have questioned the external validity of laboratory experiments and studies as to whether they could be replicated in actual settings in the field (Brunswick, 1956). However, more recent social psychologists are more interested in whether "the psychological processes that occur in an experiment are the same as psychological processes that occur in everyday life" (Brewer, 2000, p. 12). Hence, to check whether the present research has ecological validity, it is important to discuss whether the mechanisms tested in the studies differ in the lab compared to the real world. We can argue that the tasks and nature of the behaviours tested in the empirical studies are very simple and close to real-world behaviours. For instance, asking participants to imagine standing in a queue or taking a rapid COVID-19 test is very easy and arguably possible for every participant to engage in. Due to the limitation of conducting experiments in the lab, we had to rely on hypothetical scenarios to test how participants react to them. It is important to note that these behaviours are very relatable in nature and easy to imagine. In addition, the carbon emission task, where participants were asked to choose an amount to buy CO2 certificated that offset their emissions (Berger and Wyss, 2021; Hough-Guldberg et al., 2019), is a good proxy for actual behaviour. Finally, the literature and existing empirical research backed all the measures and mechanisms tested in the empirical studies.

Robustness

Another important aspect of external validity is robustness (Brewer, 2000). This construct refers to whether "the general effect holds up in the face of wide variations in subject populations and settings" (Brewer, 2000, p. 10). In the context of the present research, it is important to discuss the extent to which my findings may generalise to different populations and experimental stimuli. In the first study, we had a random sample from the UK and India, allowing us to see the effect of different social expectations. For example, how people perceive normative behaviours in India differs from those in the UK (for instance, how normative it is to stand in line or drive an electric car). However, in the other subsequent studies, we limited the sample to the UK. This is because

in study 2, we tested a specific behaviour related to taking the rapid test, and back then, it was timely as the UK government announced a regulation prohibiting access to public spaces without proof of getting tested. In study 3, we particularly looked at a scenario where participants imagined taking the train from London to Amsterdam. We were looking to present behaviour very close to real-life decisions; therefore, it was ideal to contextualise and limit the sample to a specific population. In the last study, the complexity ensured that participants would visit their social media accounts and read posts. Therefore, we ran a pilot study asking participants to report a description of the post they read. Hence, limiting the sample to the same population was more convenient.

Self-reported measures of morality (social desirability)

Another limitation that all the empirical studies share is the social desirability responses. Even though the moral self-image scale is in line with the morality literature (Jordan et al., 2015), it poses a limitation given the importance of morality in how others perceive people (Goodwin, 2015). Even though the studies were anonymous, people are still more likely to report a positive moral self-image because of the potential positive effect and perception it might have on others (i.e., the experimenter). This is directly related to the warm glow effect and social desirability bias (Lopez-Becerra and Alcon, 2021). However, despite this limitation, we could detect significant differences between moral self-image across the conditions, depending on the nature of the task or the behaviour they engaged in. This shows that even though it is a self-reported measure, it captured nuances and differences in the experience of moral self-image across participants who engaged in different tasks. Moreover, we showed the mediation role of moral self-image in explaining the behavioural spillovers and moral balancing effects. Hence, it is reasonable to assume that the results from the self-reported measures in the empirical studies are good representations of the expected mechanisms, such as moral self-image.

4. Future Directions

So far, I have discussed the theoretical and methodological implications of the current research and presented some research limitations. Hence, the final task is to highlight the most critical points I have discussed so far and propose future directions for this research to be further expanded and elaborated. I suggest that some interesting and potential next research steps could be to (1) measure the effect of social norms and social expectations on affect and moral self-image in different domains and cultures such as non-WEIRD countries; (2) measure the proposed moderated mediation model of social norms and moral self-image on behavioural spillovers in different domains, predominantly negative or harmful behaviours (e.g., corruption), and (3) explore the

effect of virtue signalling on different social causes and measure behavioural spillovers; and (4) explore other moderators and mediators that could further explain behavioural spillovers (such as the role of self-identity with the social cause and the difficulty of initial behaviour).

First, to construct a clearer view of the effect of social norms on experienced affect and moral self-image, it is crucial to run more studies that test the effect of similar behaviours in different contexts. For example, instead of manipulating two different behaviours that have various social norms and expectations, it would be interesting to take, for instance, one behaviour (such as standing in the queue) and compare whether confirming it or not influences affect and moral self-image differently in two contexts or cultures where social norms by itself differ. Moreover, future research could replicate the current study design in study 1 but for other behaviours, such as accepting bribery and supporting social causes such as climate change or gender or racial discrimination.

Second, Future research could explain the newly suggested moderated mediation model in different domains. It would be particularly interesting to test the effect of social norms and moral self-image on spillover effects in domains such as corruption, moral dilemmas, collection action and social activism. A potential future experiment could include identity/personal norm issues, thinking through opting out of expected social behaviour (e.g., not being corrupt), and how moral self-image mediates the impact on spillovers.

Third, to build more on existing research related to virtue signalling (Levy, 2021; Kraft-Todd et al., 2023; Bradley et al., 2018) and slacktivism (Greijdanus et al., 2020; Cabrera et al., 2017; Kristofferson et al., 2014), it is necessary to test the effect of public observability of an initial pro-social behaviour and signalling virtue to others on behavioural spillovers. People feel motivated to act pro-social when they know others can witness or acknowledge their actions. It is essential to explore the mechanisms through which public observability and virtue signalling exert influence. Social norms, peer influence, and the desire for social approval may mediate the relationship between these variables. Understanding these mechanisms in specific pro-social behaviours could provide insights into how behaviours spread within social networks. By delving into these aspects, researchers could contribute valuable insights into the complex interplay between public observability, virtue signalling, and behavioural spillovers, advancing our understanding of behaviour in social contexts.

Finally, another promising methodological line of research about behavioural spillovers is further examining mechanisms and constructs that moderate and mediate different behavioural spillover effects. In this present research, we focused on social norms and moral self-image. This could be further expanded to include, for instance, environmental self-identity (Van der Werff et al., 2014) or the difficulty and amount of effort consumed in the initial behaviour (Truelove et al., 2014). Furthermore, it would be interesting to explore the latter on spillunder effects too (Krpan et al., 2019) and run systematic and extensive research to capture the occurrence of behavioural spillunders and their underlying mechanisms.

5. Concluding Remarks

Behavioural spillovers have become a popular research area with demonstrations of different spillover effects and theoretical explanations widespread in the literature (Dolan and Galizzi, 2015; Galizzi and Whitmarsh, 2019; Nash et al., 2017), such as the model of sequential moral behaviour, moral credential model, and model credit model, among others. However, more research needs to be conducted with direct empirical support to understand the underlying mechanism behind these explanations. More specifically, there is a considerable debate on what drives moral balancing and compensation (Monin and Miller, 2001). So far, research in numerous domains has shown that people do not always act consistently with previous actions and several moderators (such as the observability of the initial behaviour, the cost of the initial behaviour, perception of commitment versus progress, or personal commitment to the social cause) were examined in an attempt to explain the underlying mechanisms behind behavioural spillovers (Blanken et al., 2015). Although different empirical studies look at behavioural spillovers, there needs to be more evidence of the role social norms and moral self-image play in influencing behavioural spillovers. Going beyond the previous findings, the present research shows the effect of social expectations and norms on affect and moral self-image and how the change in experienced moral self-image explains different behavioural spillovers. The present thesis successfully tests a **novel theoretical framework and statistical model** that links social norms and expectations, moral self-image, and behavioural spillovers. Indeed, it presents a **moderated mediation model** that empirically tests and provides evidence of the proposed mechanism in different domains (e.g., rapid test, sustainable transport, climate change action). **In short, the moderated mediation model explains the negative spillover effect, where the influence of engaging in the first behaviour on MSI is moderated by social norms (i.e., the**

influence is stronger for people who score higher on social expectations). MSI, in turn, negatively predicts the occurrence of the second/ subsequent behaviour.

In addition, this thesis explained positive spillover effects; taking pro-environmental action (e.g., pledging) in private led to an increase in moral self-image (regardless of the social expectations) and eventually to consistency in donating for CO2 offsetting.

Furthermore, given that moral self-image (Jordan et al., 2015) is a dynamic and malleable phenomenon that is influenced by situational factors, the thesis presents that moral self-image could explain moral balancing (Monin and Miller, 2001) and quantifies explicitly the moral credits building on the moral credit model. In other words, we show that moral self-image indirectly predicts moral balancing (such as the moral licensing effect). **Based on the results of four empirical experiments, the thesis provides several theoretical implications related to capturing underlying moderating and mediating spillover pathways.** It also offers essential practical and policy-relevant implications that aid practitioners in predicting possible behavioural spillovers. Finally, we present avenues for future research further to explore new empirical mechanisms behind behavioural spillover effects and replicate the model in other behavioural domains as well as contexts and cultures.

6. References

Aarts, H., & Dijksterhuis, A. (2003). The silence of the library: Environment, situational norm, and social behavior. *Journal of Personality and Social Psychology*, *84*(1), 18–28. https://doi.org/10.1037/0022-3514.84.1.18

Abrahamse, W., & Steg, L. (2013). Social influence approaches to encourage resource conservation: A meta-analysis. *Global Environmental Change*, *23*(6), 1773–1785. https://doi.org/10.1016/j.gloenvcha.2013.07.029

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

Ajzen, I. (2012). The theory of planned behavior. In *Handbook of Theories of Social Psychology* (Vol. 1, pp. 438–459). https://doi.org/10.4135/9781446249215.n22

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, *100*(401), 464–477. https://doi.org/10.2307/2234133

Aquino, K., McFerran, B., & Laven, M. (2011). Moral identity and the experience of moral elevation in response to acts of uncommon goodness. *Journal of Personality and Social Psychology*, *100*(4), 703–718. https://doi.org/10.1037/a0022540

Aquino, K., & Reed, A. (2002). The self-importance moral identity. *Journal of Personality and Social Psychology*, *83*, 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Aronson, E. (1969). The Theory of Cognitive Dissonance: A Current Perspective11This article and all the original work reported in it were supported by a grant from the National Science Foundation. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 4, pp. 1–34). Academic Press. https://doi.org/10.1016/S0065-2601(08)60075-1

Aronson, E. (1997). The theory of cognitive dissonance: The evolution and vicissitudes of an idea. In *The message of social psychology: Perspectives on mind in society* (pp. 20–35). Blackwell Publishing.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Bal, A., Archer-Brown, C., Robson, K., & Hall, D. (2013). Do good, goes bad, gets ugly: Kony 2012. *Journal of Public Affairs*, *13*. https://doi.org/10.1002/pa.1475

Bamberg, S., Hunecke, M., & Blöbaum, A. (2007). Social context, personal norms and the use of public transportation: Two field studies. *Journal of Environmental Psychology*, *27*(3), 190–203. https://doi.org/10.1016/j.jenvp.2007.04.001

Bandura, A. (1999). *SOCIAL COGNITIVE THEORY OF PERSONALITY*. 81.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). *Mechanisms of Moral Disengagement in the Exercise of Moral Agency*. 11.

Barkan, R. (2015). Ethical Dissonance, Justifications and Moral Behavior. *Current Opinion in Psychology*, *6*, 157–161.

Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, *141*(4), 757–773. https://doi.org/10.1037/a0027588

Baron, R. M., & Kenny, D. A. (1986a). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Baron, R. M., & Kenny, D. A. (1986b). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Barth, M., Jugert, P., & Fritsche, I. (2016). Still underdetected – Social norms and collective efficacy predict the acceptance of electric vehicles in Germany. *Transportation Research Part F: Traffic Psychology and Behaviour*, *37*, 64–77. https://doi.org/10.1016/j.trf.2015.11.011

Baumeister, R. F., & Alghamdi, N. (2015a). Resource-based interventions in the workplace: Integration, commentary, and recommendations. *Journal of Occupational and Organizational Psychology*, *88*(3), 623–629. https://doi.org/10.1111/joop.12130

Baumeister, R. F., & Alghamdi, N. G. (2015b). Role of self-control failure in immoral and unethical actions. *Current Opinion in Psychology*, *6*, 66–69. https://doi.org/10.1016/j.copsyc.2015.04.001

Bazerman, M. H., & Gino, F. (2012). Behavioral Ethics: Toward a Deeper Understanding of Moral Judgment and Dishonesty. *Annual Review of Law and Social Science*, *8*(1), 85–104. https://doi.org/10.1146/annurev-lawsocsci-102811-173815

Beaman, A. L., Cole, C. M., Preston, M., Klentz, B., & Steblay, N. M. (1983). Fifteen Years of Foot-in-the Door Research: A Meta-Analysis. *Personality and Social Psychology Bulletin*, *9*(2), 181–196. https://doi.org/10.1177/0146167283092002

Bem, D. J. (1972). Self-Perception Theory11Development of self-perception theory was supported primarily by a grant from the National Science Foundation (GS 1452) awarded to the author during his tenure at Carnegie-Mellon University. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 6, pp. 1–62). Academic Press. https://doi.org/10.1016/S0065-2601(08)60024-6

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bentham, J. (1823). *An Introduction to the Principles of Morals and Legislation*. W. Pickering.

Berger, S., & Wyss, A. M. (2021). Measuring pro-environmental behavior using the carbon emission task. *Journal of Environmental Psychology*, *75*. https://doi.org/10.1016/j.jenvp.2021.101613

Berkowitz, L. (1972). Frustrations, comparisons, and other sources of emotion arousal as contributors to social unrest. *Journal of Social Issues*, *28*(1), 77–91. https://doi.org/10.1111/j.1540-4560.1972.tb00005.x

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms* (pp. xvi, 260). Cambridge University Press.

Bicchieri, C. (2016). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Bicchieri, C., Fatas, E., Aldama, A., Casas, A., Deshpande, I., Lauro, M., Parilli, C., Spohn, M., Pereira, P., & Wen, R. (2020). *In Science we (should) trust: Expectations and compliance during the COVID-19 pandemic* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-106840/v1

Bicchieri, C., Lindemans, J. W., & Jiang, T. (2014). A structured approach to a diagnostic of collective practices. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01418

Bicchieri, C., & Mercier, H. (2014). Norms and Beliefs: How Change Occurs. In M. Xenitidou & B. Edmonds (Eds.), *The Complexity of Social Norms* (pp. 37–54). Springer International Publishing. https://doi.org/10.1007/978-3-319-05308-0_3

Biel, A., Johansson-Stenman, O., & Nilsson, A. (2011). The willingness to pay–willingness to accept gap revisited: The role of affect and moral satisfaction. *Journal of Economic Psychology*, *32*(6), 908–917. https://doi.org/10.1016/j.joep.2011.07.010

Bissing-Olson, M. J., Fielding, K. S., & Iyer, A. (2016). Experiences of pride, not guilt, predict pro-environmental behavior when pro-environmental descriptive norms are more positive.

Journal of Environmental Psychology, *45*, 145–153. https://doi.org/10.1016/j.jenvp.2016.01.001

Bissing-Olson, M. J., Iyer, A., Fielding, K. S., & Zacher, H. (2013). Relationships between daily affect and pro-environmental behavior at work: The moderating role of pro-environmental attitude: DAILY AFFECT AND PRO-ENVIRONMENTAL BEHAVIOR. *Journal of Organizational Behavior*, *34*(2), 156–175. https://doi.org/10.1002/job.1788

Black, J. S., Stern, P. C., & Elworth, J. T. (1985). Personal and contextual influences on household energy adaptations. *Journal of Applied Psychology*, *70*(1), 3. https://doi.org/10.1037/0021-9010.70.1.3

Blake, J., & Davis, K. (1964). Norms, values, and sanctions. In *Handbook of modern sociology* (pp. 456–484).

Blamey, R. (1998). The Activation of Environmental Norms: Extending Schwartz's Model. *Environment and Behavior*, *30*(5), 676–708. https://doi.org/10.1177/001391659803000505

Blanken, I., Van De Ven, N., & Zeelenberg, M. (2015). A Meta-Analytic Review of Moral Licensing. *Personality and Social Psychology Bulletin*, *41*(4), 540–558. https://doi.org/10.1177/0146167215572134

Blanken, I., van de Ven, N., Zeelenberg, M., & Meijers, M. H. C. (20140602). Three attempts to replicate the moral licensing effect. *Social Psychology*, *45*(3), 232. https://doi.org/10.1027/1864-9335/a000189

Blasi, A. (1983). Moral cognition and moral action: A theoretical perspective. *Developmental Review*, *3*(2), 178–210. https://doi.org/10.1016/0273-2297(83)90029-1

Boegershausen, J., Aquino, K., & Reed, A. (2015). Moral identity. *Current Opinion in Psychology*, *6*, 162–166. https://doi.org/10.1016/j.copsyc.2015.07.017

Borsari, B., & Carey, K. B. (2003). Descriptive and injunctive norms in college drinking: A meta-analytic integration. *Journal of Studies on Alcohol*, *64*(3), 331–341. https://doi.org/10.15288/jsa.2003.64.331

Burger, J. M. (1999). The Foot-in-the-Door Compliance Procedure: A Multiple-Process Analysis and Review. *Personality and Social Psychology Review*, *3*(4), 303–325. https://doi.org/10.1207/s15327957pspr0304_2

Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality? *Proceedings of the Royal Society B: Biological Sciences*, *285*(1875), 20180116. https://doi.org/10.1098/rspb.2018.0116

Brewer, M. B. (2000). Research design and issues of validity. *Handbook of Research Methods in Social and Personality Psychology.*, 3–16.

Brunstein, J. C., & Gollwitzer, P. M. (1996). Effects of failure on subsequent performance: The importance of self-defining goals. *Journal of Personality and Social Psychology*, *70*(2), 395–407. https://doi.org/10.1037/0022-3514.70.2.395

Cabrera, N. L., Matias, C. E., & Montoya, R. (2017). Activism or slacktivism? The potential and pitfalls of social media in contemporary student activism. *Journal of Diversity in Higher Education*, *10*(4), 400–415. https://doi.org/10.1037/dhe0000061

Cascio, J., & Plant, E. A. (2015). Prospective moral licensing: Does anticipating doing good later allow you to be bad now? *Journal of Experimental Social Psychology*, *56*, 110–116. https://doi.org/10.1016/j.jesp.2014.09.009

Choi, B. (Peter), & Crandall, C. S. (2008). Permission to Be Prejudiced: Legitimacy Credits in the Evaluation of Advertisements With Black and White Models. *ACR North American Advances*, *NA-35*. https://www.acrwebsite.org/volumes/13145/volumes/v35/NA-35

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. *Social Influence*, *1*(1), 3–15. https://doi.org/10.1080/15534510500181459

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 24, pp. 201–234). Academic Press. https://doi.org/10.1016/S0065-2601(08)60330-5

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015. https://doi.org/10.1037/0022-3514.58.6.1015

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In *The handbook of social psychology, Vols. 1-2, 4th ed* (pp. 151–192). McGraw-Hill.

Cialdini, R., & Goldstein, N. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, *55*, 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Conway, P., & Peetz, J. (2012). When Does Feeling Moral Actually Make You a Better Person? Conceptual Abstraction Moderates Whether Past Moral Deeds Motivate Consistency or Compensatory Behavior. *Personality and Social Psychology Bulletin*, *38*(7), 907–919. https://doi.org/10.1177/0146167212442394

Cooter, R. (1996). Normative Failure Theory of Law. *Cornell Law Review*, *82*, 947.

Cornelissen, G., Dewitte, S., & Warlop, L. (2011). Are Social Value Orientations Expressed Automatically? Decision Making in the Dictator Game. *Personality and Social Psychology Bulletin*, *37*(8), 1080–1090. https://doi.org/10.1177/0146167211405996

Curry, O. S. (2016). Morality as Cooperation: A Problem-Centred Approach. In T. K. Shackelford & R. D. Hansen (Eds.), *The Evolution of Morality* (pp. 27–51). Springer International Publishing. https://doi.org/10.1007/978-3-319-19671-8_2

Curry, O. S., Jones Chesters, M., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

Cushman, F., & Young, L. (2009). The Psychology of Dilemmas and the Philosophy of Morality. *Ethical Theory and Moral Practice*, *12*, 9–24. https://doi.org/10.1007/s10677-008-9145-3

Damon, W., & Hart, D. (1992). Self-understanding and its role in social and moral development. In *Developmental psychology: An advanced textbook, 3rd ed* (pp. 421–464). Lawrence Erlbaum Associates, Inc.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629–636. https://doi.org/10.1037/h0046408

Dolan, P. (2014). *Happiness by Design: Finding Pleasure and Purpose in Everyday Life*. Penguin UK.

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, *47*, 1–16. https://doi.org/10.1016/j.joep.2014.12.003

Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). *MINDSPACE: Influencing behaviour for public policy* [Monograph]. Institute of Government. http://www.instituteforgovernment.org.uk/publications/

Doris, J. M., & Doris, J. M. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge University Press.

Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring Whites. *Journal of Experimental Social Psychology*, *45*(3), 590–593. https://doi.org/10.1016/j.jesp.2009.02.001

Elster, J. (1989). Social Norms and Economic Theory. *The Journal of Economic Perspectives*, *3*(4), 99–117.

Epley, N., & Dunning, D. (2000). Feeling 'holier than thou': Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology*, *79*(6), 861–875. https://doi.org/10.1037/0022-3514.79.6.861

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review*, *112*(4), 912–931. https://doi.org/10.1037/0033-295X.112.4.912

Evans, B. E., Greaves-Lord, K., Euser, A. S., Tulen, J. H. M., Franken, I. H. A., & Huizink, A. C. (2013). Determinants of physiological and perceived physiological stress reactivity in children and adolescents. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061724

Fagundes, D. (2017). The Social Norms of Waiting in Line. *Law & Social Inquiry*, *42*(4), 1179–1207. https://doi.org/10.1111/lsi.12256

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Ferguson, M. A., & Branscombe, N. R. (2010). Collective guilt mediates the effect of beliefs about global warming on willingness to engage in mitigation behavior. *Journal of Environmental Psychology*, *30*(2), 135–142. https://doi.org/10.1016/j.jenvp.2009.11.010

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140. https://doi.org/10.1177/001872675400700202

Fishbein & Ajzen (1975). (n.d.). Retrieved 1 April 2021, from https://people.umass.edu/aizen/f&a1975.html

Fointiat, V. (2006). "YOU'RE HELPFUL" VERSUS "THAT'S CLEAR". SOCIAL VERSUS FUNCTIONAL LABEL IN THE FOOT-IN-THE-DOOR PARADIGM. *Social Behavior and Personality: An International Journal*, *34*(5), 461–466. https://doi.org/10.2224/sbp.2006.34.5.461

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*(2), 195–202. https://doi.org/10.1037/h0023552

Galizzi, M. M., & Whitmarsh, L. (2019). How to Measure Behavioral Spillovers: A Methodological Review and Checklist. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/article/10.3389/fpsyg.2019.00342

Gawronski, B., & Strack, F. (2012). Cognitive consistency as a basic principle of social information processing. *Cognitive Consistency: A Fundamental Principle in Social Cognition.*, 1–16.

Gelfand, M. J., Harrington, J. R., & Jackson, J. C. (2017). The Strength of Social Norms Across Human Groups. *Perspectives on Psychological Science*, *12*(5), 800–809. https://doi.org/10.1177/1745691617708631

Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. (2012). Paying to Be Nice: Consistency and Costly Prosocial Behavior. *Management Science*, *58*(1), 179–187. https://doi.org/10.1287/mnsc.1110.1437

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research*, *35*(3), 472–482. https://doi.org/10.1086/586910

Goodwin, G. P. (2015). Moral character in person perception. Current Directions in Psychological Science, 24(1), 38–44. https://doi.org/10.1177/0963721414550709

Grace, D., & Griffin, D. (2009). Conspicuous donation behaviour: Scale development and validation. *Journal of Consumer Behaviour*, *8*(1), 14–25. https://doi.org/10.1002/cb.270

Graham, J., Nosek, B., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. (2011). Mapping the Moral Domain. *Journal of Personality and Social Psychology*, *101*, 366–385. https://doi.org/10.1037/a0021847

Gregory-Smith, D., Smith, A., & Winklhofer, H. (2013). Affect and dissonance in 'ethical' consumption choices. *Journal of Marketing Management*, *29*(11–12), 1201–1223. https://doi.org/10.1080/0267257X.2013.796320

Greijdanus, H., De Matos Fernandes, C. A., Turner-Zwinkels, F., Honari, A., Roos, C. A., Rosenbusch, H., & Postmes, T. (2020). The psychology of online activism and social movements: Relations between online and offline collective action. *Current Opinion in Psychology*, *35*, 49–54. https://doi.org/10.1016/j.copsyc.2020.03.003

Griffin, D. J., Bolkan, S., & Goodboy, A. K. (2015). Academic Dishonesty Beyond Cheating and Plagiarism: Students' Interpersonal Deception in the College Classroom. *Qualitative Research Reports in Communication*, *16*(1), 9–19. https://doi.org/10.1080/17459435.2015.1086416

Guagnano, G., Stern, P., & Dietz, T. (1995). Influences on Attitude-Behavior RelationshipsA Natural Experiment with Curbside Recycling. *Environment and Behavior - ENVIRON BEHAV*, *27*, 699–718. https://doi.org/10.1177/0013916595275005

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. https://doi.org/10.1037/0033-295x.108.4.814

Haidt, J. (2003). The moral affect. In *Handbook of affective sciences* (pp. 852–870). Oxford University Press.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion* (pp. xvii, 421). Pantheon/Random House.

Haidt, J., & Joseph, C. (2004). Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues. *Daedalus*, *133*(4), 55–66. https://doi.org/10.1162/0011526042365555

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2014) *The Behavioralist As Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance*. 45.

Harth, N. S., Leach, C. W., & Kessler, T. (2013). Guilt, anger, and pride about in-group environmental behaviour: Different affect predict distinct intentions. *Journal of Environmental Psychology*, *34*, 18–26. https://doi.org/10.1016/j.jenvp.2012.12.005

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1–22. https://doi.org/10.1080/00273171.2014.962683

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, *85*(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100

Heider, F. (1958). The naive analysis of action. *The Psychology of Interpersonal Relations.*, 79–124. https://doi.org/10.1037/10628-004

Herman, C. P., & Mack, D. (1975). Restrained and unrestrained eating. *Journal of Personality*, *43*(4), 647–660. https://doi.org/10.1111/j.1467-6494.1975.tb00727.x

Herrmann, T. (1971). 'Prescriptive' and 'proscriptive' conformity as a result of childhood training. *European Journal of Social Psychology*, *1*(1), 140–140. https://doi.org/10.1002/ejsp.2420010111

Hertwich, E. G., & Peters, G. P. (2009). Carbon Footprint of Nations: A Global, Trade-Linked Analysis. *Environmental Science & Technology*, *43*(16), 6414–6420. https://doi.org/10.1021/es803496a

Hofmann, W., Wisneski, D., Brandt, M., & Skitka, L. (2014). Morality in everyday life. *Science (New York, N.Y.)*, *345*, 1340–1343. https://doi.org/10.1126/science.1251560

Hoegh-Guldberg, O., Jacob, D., Taylor, M., Guillén Bolaños, T., Bindi, M., Brown, S., Camilloni, I. A., Diedhiou, A., Djalante, R., Ebi, K., Engelbrecht, F., Guiot, J., Hijioka, Y., Mehrotra, S., Hope, C. W., Payne, A. J., Pörtner, H.-O., Seneviratne, S. I., Thomas, A., … Zhou, G. (2019a). The human imperative of stabilizing global climate change at 1.5°C. *Science*, *365*(6459), eaaw6974. https://doi.org/10.1126/science.aaw6974

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, *85*(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1–22. https://doi.org/10.1080/00273171.2014.962683

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology*, *96*(3), 521–537. https://doi.org/10.1037/a0013779

Joosten, A., van Dijke, M., Van Hiel, A., & De Cremer, D. (2013). Feel Good, Do-Good!? On Consistency and Compensation in Moral Self-Regulation. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-013-1794-z

Jordan, A. H., Monin, B., Dweck, C. S., Lovett, B. J., John, O. P., & Gross, J. J. (2011). Misery Has More Company Than People Think: Underestimating the Prevalence of Others' Negative Affect. *Personality and Social Psychology Bulletin*, *37*(1), 120–135. https://doi.org/10.1177/0146167210390822

Jordan, J., Leliveld, M. C., & Tenbrunsel, A. E. (2015). The Moral Self-Image Scale: Measuring and Understanding the Malleability of the Moral Self. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01878

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the Moral Self: The Effects of Recalling Past Moral Actions on Future Moral Behavior. *Personality and Social Psychology Bulletin*, *37*(5), 701–713. https://doi.org/10.1177/0146167211400208

Judah, G., Aunger, R., Schmidt, W.-P., Michie, S., Granger, S., & Curtis, V. (2009). Experimental pretesting of hand-washing interventions in a natural setting. *American Journal of Public Health*, *99 Suppl 2*, S405-411. https://doi.org/10.2105/AJPH.2009.164160

Kahneman, D. (2011). *Thinking, fast and slow* (p. 499). Farrar, Straus and Giroux.

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A Focus Theory of Normative Conduct: When Norms Do and Do not Affect Behavior. *Personality and Social Psychology Bulletin*, *26*(8), 1002–1012. https://doi.org/10.1177/01461672002610009

Kant, I. (1785). *The Metaphysics of Morals*. Cambridge University Press.

Kercher, K. (1992). Assessing Subjective Well-Being in the Old-Old: The PANAS as a Measure of Orthogonal Dimensions of Positive and Negative Affect. *Research on Aging*, *14*(2), 131–168. https://doi.org/10.1177/0164027592142001

Kernis, M. H., & Goldman, B. M. (2006). A Multicomponent Conceptualization of Authenticity: Theory and Research. In *Advances in Experimental Social Psychology* (Vol. 38, pp. 283–357). Elsevier. https://doi.org/10.1016/S0065-2601(06)38006-9

Kernis, M. H., & Johnson, E. K. (1990). Current and typical self-appraisals: Differential responsiveness to evaluative feedback and implications for affect. *Journal of Research in Personality*, *24*(2), 241–257. https://doi.org/10.1016/0092-6566(90)90019-3

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement Learning Signal Predicts Social Conformity. *Neuron*, *61*(1), 140–151. https://doi.org/10.1016/j.neuron.2008.11.027

Köbis, N. C., Verschuere, B., Bereby-Meyer, Y., Rand, D., & Shalvi, S. (2019). Intuitive honesty versus dishonesty: Meta-analytic evidence. *Perspectives on Psychological Science*, *14*(5), 778–796. https://doi.org/10.1177/1745691619851778

Kohlberg, Lawrence, 1927-1987. (1981). The philosophy of moral development : moral stages and the idea of justice. San Francisco :Harper & Row,

Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development*, *12*(2), 93–120. https://doi.org/10.1159/000270857

Kraft-Todd, G. T., Kleiman-Weiner, M., & Young, L. (2023). Virtue Discounting: Observability Reduces Moral Actors' Perceived Virtue. *OPEN MIND*.

Kristofferson, K., White, K., & Peloza, J. (2014). The Nature of Slacktivism: How the Social Observability of an Initial Act of Token Support Affects Subsequent Prosocial Action. *Journal of Consumer Research*, *40*(6), 1149–1166. https://doi.org/10.1086/674137

Kroll, J., & Egan, E. (2004). Psychiatry, Moral Worry, and the Moral Affect. *Journal of Psychiatric Practice*, *10*(6), 352–360. https://doi.org/10.1097/00131746-200411000-00003

Krpan, D., & Dolan, P. (2022). You Must Stay at Home! The Impact of Commands on Behaviors During COVID-19. *Social Psychological and Personality Science*, *13*(1), 333–346. https://doi.org/10.1177/19485506211005582

Krpan, D., Galizzi, M. M., & Dolan, P. (2019). Looking at Spillovers in the Mirror: Making a Case for "Behavioral Spillunders". *Frontiers in Psychology*, *10*, 1142. https://doi.org/10.3389/fpsyg.2019.01142

Lapinski, M. K., & Rimal, R. N. (2005). An Explication of Social Norms. *Communication Theory*, *15*(2), 127–147. https://doi.org/10.1111/j.1468-2885.2005.tb00329.x

Levy, N. (2021). Virtue signalling is virtuous. *Synthese*, *198*(10), 9545–9562. https://doi.org/10.1007/s11229-020-02653-9

Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2018). The role of a "common is moral" heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General*, *147*(2), 228. https://doi.org/10.1037/xge0000365

Lopez-Becerra, E. I., & Alcon, F. (2021). Social desirability bias in the environmental economic valuation: An inferred valuation approach. *Ecological Economics*, *184*, 106988. https://doi.org/10.1016/j.ecolecon.2021.106988

MacCormick, N. (1998). Norms, Institutions, and Institutional Facts. *Law and Philosophy*, *17*(3), 301–345. https://doi.org/10.2307/3504883

Mackie, G. (1996). Ending Footbinding and Infibulation: A Convention Account. *American Sociological Review*, *61*(6), 999–1017. https://doi.org/10.2307/2096305

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the Positive and Negative Affect Schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences*, *27*(3), 405–416. https://doi.org/10.1016/S0191-8869(98)00251-7

Mallett, R. K., Melchiori, K. J., & Strickroth, T. (2013). Self-Confrontation via a Carbon Footprint Calculator Increases Guilt and Support for a Proenvironmental Group. *Ecopsychology*, *5*(1), 9–16. https://doi.org/10.1089/eco.2012.0067

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*, 63–78. https://doi.org/10.1037/0022-3514.35.2.63

Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, *45*(6), 633–644. https://doi.org/10.1509/jmkr.45.6.633

Mazar, N., & Zhong, C.-B. (2010). Do Green Products Make Us Better People? *Psychological Science*, *21*(4), 494–498. https://doi.org/10.1177/0956797610363538

Meijers, M. H. C., & Rutjens, B. T. (2014). Affirming belief in scientific progress reduces environmentally friendly behaviour: Progress and environmental behaviour. *European Journal of Social Psychology*, *44*(5), 487–495. https://doi.org/10.1002/ejsp.2009

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral Self-Licensing: When Being Good Frees Us to Be Bad. *Social and Personality Psychology Compass*, *4*(5), 344–357. https://doi.org/10.1111/j.1751-9004.2010.00263.x

Milgram, S., Liberty, H. J., Toledo, R., & Wackenhut, J. (1986). Response to intrusion into waiting lines. *Journal of Personality and Social Psychology*, *51*(4), 683–689. https://doi.org/10.1037/0022-3514.51.4.683

Miller, D. T., & Effron, D. A. (2010). Psychological License. In *Advances in Experimental Social Psychology* (Vol. 43, pp. 115–155). Elsevier. https://doi.org/10.1016/S0065-2601(10)43003-8

Miller, D. T., & Prentice, D. A. (2016). Changing Norms to Change Behavior. *Annual Review of Psychology*, *67*, 339–361. https://doi.org/10.1146/annurev-psych-010814-015013

Monin, B., & Jordan, A. H. (2009). The dynamic moral self: A social psychological perspective. In *Personality, identity, and character: Explorations in moral psychology* (pp. 341–354). Cambridge University Press. https://doi.org/10.1017/CBO9780511627125.016

Monin, B., & Miller, D. (2001). Moral Credentials and the Expression of Prejudice. *Journal of Personality and Social Psychology*, *81*, 33–43. https://doi.org/10.1037//0022-3514.81.1.33

Moore, C., & Gino, F. (2013). Ethically adrift: How others pull our moral compass from true North, and how we can fix it. *Research in Organizational Behavior*, *33*, 53–77. https://doi.org/10.1016/j.riob.2013.08.001

Moral cognition and moral action: A theoretical perspective. (1983). *Developmental Review*, *3*(2), 178–210. https://doi.org/10.1016/0273-2297(83)90029-1

Mukhopadhyay, A., Sengupta, J., & Ramanathan, S. (2008). Recalling Past Temptations: An Information-Processing Perspective on the Dynamics of Self-Control. *JOURNAL OF CONSUMER RESEARCH*, 14.

Mullen, E., & Monin, B. (2016). Consistency Versus Licensing Effects of Past Moral Behavior. *Annual Review of Psychology*, *67*(1), 363–385. https://doi.org/10.1146/annurev-psych-010213-115120

Narvaez, D., & Lapsley, D. K. (2009). *Personality, Identity, and Character: Explorations in Moral Psychology*. Cambridge University Press.

Nash, N., & Whitmarsh, L. (2023). *One thing leads to another? Pro-environmental behavioural spillover* (pp. 63–77). https://doi.org/10.4337/9781800882133.00011

Nash, N., Whitmarsh, L., Capstick, S., Hargreaves, T., Poortinga, W., Thomas, G., Sautkina, E., & Xenias, D. (2017). Climate-relevant behavioral spillover and the potential contribution of social practice theory. *WIREs Climate Change*, *8*(6). https://doi.org/10.1002/wcc.481

Nash, N., Whitmarsh, L., Capstick, S., Thøgersen, J., Gouveia, V., de Carvalho Rodrigues Araújo, R., Harder, M. K., Wang, X., & Liu, Y. (2019). Reflecting on Behavioral Spillover in Context: How Do Behavioral Motivations and Awareness Catalyze Other Environmentally Responsible Actions in Brazil, China, and Denmark? *Frontiers in Psychology*, *10*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.00788

Noland, A. (2019). *Social Media Activists: Analyzing the Relationship between Online Activism and Offline Attitudes and Behaviors*.

Olson, M. (1971). *The Logic of Collective Action: Public Goods and the Theory of Groups, Second Printing with a New Preface and Appendix*. Harvard University Press.

Onwezen, M. C., Antonides, G., & Bartels, J. (2013). The Norm Activation Model: An exploration of the functions of anticipated pride and guilt in pro-environmental behaviour. *Journal of Economic Psychology*, *39*, 141–153. https://doi.org/10.1016/j.joep.2013.07.005

Ostrom, E., & Ostrom, V. (2014). *Choice, Rules and Collective Action: The Ostrom's on the Study of Institutions and Governance*. ECPR Press.

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network. *Journal of Personality and Social Psychology*, *103*(6), 899–915. https://doi.org/10.1037/a0030015

Pavey, L., Sparks, P., & Churchill, S. (2018). Proscriptive vs. Prescriptive Health Recommendations to Drink Alcohol Within Recommended Limits: Effects on Moral Norms, Reactance, Attitudes, Intentions and Behaviour Change. *Alcohol and Alcoholism*, *53*(3), 344–349. https://doi.org/10.1093/alcalc/agx123

Perkins, H. W., Craig, D. W., & Perkins, J. M. (2011). Using social norms to reduce bullying: A research intervention among adolescents in five middle schools. *Group Processes & Intergroup Relations*, *14*(5), 703–722. https://doi.org/10.1177/1368430210398004

Perugini, M., & Leone, L. (2009). Implicit self-concept and moral action. *Journal of Research in Personality*, *43*(5), 747–754. https://doi.org/10.1016/j.jrp.2009.03.015

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behavioral Research*, *42*(1), 185–227. https://doi.org/10.1080/00273170701341316

Rai, T., & Fiske, A. (2011). Moral Psychology Is Relationship Regulation: Moral Motives for Unity, Hierarchy, Equality, and Proportionality. *Psychological Review*, *118*, 57–75. https://doi.org/10.1037/a0021867

Reed, A. (2002). Social identity as a useful perspective for self-concept–based consumer research. *Psychology & Marketing*, *19*(3), 235–266. https://doi.org/10.1002/mar.10011

Reno, R., Cialdini, R., & Kallgren, C. (1993). The Transsituational Influence of Social Norms. *Journal of Personality and Social Psychology*, *64*, 104–112. https://doi.org/10.1037/0022-3514.64.1.104

Rimal, R. N., & Real, K. (2005). How Behaviors are Influenced by Perceived Norms: A Test of the Theory of Normative Social Behavior. *Communication Research*, *32*(3), 389–414. https://doi.org/10.1177/0093650205275385

Rivis, A., & Sheeran, P. (2003). Descriptive Norms as an Additional Predictor in the Theory of Planned Behaviour: A Meta-Analysis. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, *22*(3), 218–233. https://doi.org/10.1007/s12144-003-1018-2

Rotella, A., & Barclay, P. (2020). Failure to replicate moral licensing and moral cleansing in an online experiment. *Personality and Individual Differences*, *161*, 109967. https://doi.org/10.1016/j.paid.2020.109967

Ryan, R., & Deci, E. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *The American Psychologist*, *55*, 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). *PSYCHOLOGICAL SCIENCE Research Article Sinning Saints and Saintly Sinners The Paradox of Moral Self-Regulation*.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science (0956-7976)*, *18*(5), 429–434. https://doi.org/10.1111/j.1467-9280.2007.01917.x

Schwartz, S. H. (1977). Normative Influences on Altruism11This work was supported by NSF Grant SOC 72-05417. I am indebted to L. Berkowitz, R. Dienstbier, H. Schuman, R.

Simmons, and R. Tessler for their thoughtful comments on an early draft of this chapter. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 221–279). Academic Press. https://doi.org/10.1016/S0065-2601(08)60358-5

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-Serving Justifications: Doing Wrong and Feeling Moral. *Current Directions in Psychological Science*, *24*(2), 125–130. https://doi.org/10.1177/0963721414553264

Sheikh, S., & Janoff-Bulman, R. (2010). The "Shoulds" and "Should Nots" of Moral Affect: A Self-Regulatory Perspective on Shame and Guilt. *Personality and Social Psychology Bulletin*, *36*(2), 213–224. https://doi.org/10.1177/0146167209356788

Sheldon, O. J., & Fishbach, A. (2015). Anticipating and Resisting the Temptation to Behave Unethically. *Personality and Social Psychology Bulletin*, *41*(7), 962–975. https://doi.org/10.1177/0146167215586196

Ariely, D. (2009). Predictably Irrational: The Hidden Forces That Shape Our Decisions Dan Ariely (2009). *International Journal of Advertising*, *28*(4), 743–745. https://doi.org/10.2501/S0265048709200862

Smith, J. R., & Louis, W. R. (2008). Group norms and the attitude-behaviour relationship. *Social and Personality Psychology Compass*, *3*(1), 19–35. https://doi.org/10.1111/j.1751-9004.2008.00161.x

Sparkman, G., & Walton, G. M. (2017). Dynamic Norms Promote Sustainable Behavior, Even if It Is Counternormative. *Psychological Science*, *28*(11), 1663–1674. https://doi.org/10.1177/0956797617719950

Sparkman, G., & Walton, G. M. (2019). Witnessing change: Dynamic norms help resolve diverse barriers to personal change. *Journal of Experimental Social Psychology*, *82*, 238–252.

Steg, L., Perlaviciute, G., & Van Der Werff, E. (2015). Understanding the human dimensions of a sustainable energy transition. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00805

Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology*, *29*(3), 309–317. https://doi.org/10.1016/j.jenvp.2008.10.004

Steinhorst, J., Klöckner, C. A., & Matthies, E. (2015). Saving electricity – For the money or the environment? Risks of limiting pro-environmental spillover when using monetary framing. *Journal of Environmental Psychology*, *43*, 125–135. https://doi.org/10.1016/j.jenvp.2015.05.012

Stets, J. E., & Burke, P. J. (2000). Identity Theory and Social Identity Theory. *Social Psychology Quarterly*, *63*(3), 224–237. https://doi.org/10.2307/2695870

Strack, F., Pauli, P., & Weyers, P. (2016). Editorial: Emotion and Behavior. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.00313

Susewind, M., & Hoelzl, E. (2014). A matter of perspective: Why past moral behavior can sometimes encourage and other times discourage future moral striving. *Journal of Applied Social Psychology*, *44*(3), 201–209. https://doi.org/10.1111/jasp.12214

Sunstein, C. (2019). *Conformity*.

Tajfel, H. (1972). Some developments in European social psychology. *European Journal of Social Psychology*, *2*(3), 307–321. https://doi.org/10.1002/ejsp.2420020307

Tajfel, Turner. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations*, 33–37.

Tangney, J. P. (1991). Moral affect: The good, the bad, and the ugly. *Journal of Personality and Social Psychology*, *61*(4), 598–607. https://doi.org/10.1037/0022-3514.61.4.598

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt* (pp. xvi, 272). Guilford Press. https://doi.org/10.4135/9781412950664.n388

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral Affect and Moral Behavior. *Annual Review of Psychology*, *58*, 345–372. https://doi.org/10.1146/annurev.psych.56.091103.070145

Tankard, M. E., & Paluck, E. L. (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review*, *10*(1), 181–211. https://doi.org/10.1111/sipr.12022

Tappin, B. M., & McKay, R. T. (2017). The Illusion of Moral Superiority. *Social Psychological and Personality Science*, *8*(6), 623–631. https://doi.org/10.1177/1948550616673878

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870. https://doi.org/10.1037//0022-3514.78.5.853

Thøgersen, J. (1999). Spillover processes in the development of a sustainable consumption pattern. *Journal of Economic Psychology*, *20*(1), 53–81. https://doi.org/10.1016/S0167-4870(98)00043-9

Thøgersen, J. (2006). Norms for environmentally responsible behaviour: An extended taxonomy. *Journal of Environmental Psychology*, *26*(4), 247–261. https://doi.org/10.1016/j.jenvp.2006.09.004

Thøgersen, J., & Crompton, T. (2009). *Simple and Painless? The Limitations of Spillover in Environmental Campaigning*. https://doi.org/10.1007/S10603-009-9101-1

Thøgersen, J., & Ölander, F. (2003). Spillover of environment-friendly consumer behaviour. *Journal of Environmental Psychology*, *23*(3), 225–236. https://doi.org/10.1016/S0272-4944(03)00018-5

Tomasello, M., & Vaish, A. (2012). Origins of Human Cooperation and Morality. *Annual Review of Psychology*, *64*. https://doi.org/10.1146/annurev-psych-113011-143812

Tracy, J. L., & Robins, R. W. (2007). The psychological structure of pride: A tale of two facets. *Journal of Personality and Social Psychology*, *92*(3), 506–525. https://doi.org/10.1037/0022-3514.92.3.506

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., & Vandenbergh, M. P. (2014). Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, *29*, 127–138. https://doi.org/10.1016/j.gloenvcha.2014.09.004

Tu, J.-C., & Yang, C. (2019). Key Factors Influencing Consumers' Purchase of Electric Vehicles. *Sustainability*, *11*(14), 3863. https://doi.org/10.3390/su11143863

Van der Werff, E., Steg, L., & Keizer, K. (2014). Follow the signal: When past pro-environmental actions signal who you are. *Journal of Environmental Psychology*, *40*. https://doi.org/10.1016/j.jenvp.2014.07.004

Vlaev, I., & Dolan, P. (2015). Action Change Theory: A Reinforcement Learning Perspective on Behavior Change. *Review of General Psychology*, *19*(1), 69–95. https://doi.org/10.1037/gpr0000029

Wallace, E., Buil, I., & Chernatony, L. (2020). 'Consuming Good' on Social Media: What Can Conspicuous Virtue Signalling on Facebook Tell Us About Prosocial and Unethical Intentions? *Journal of Business Ethics*, *162*. https://doi.org/10.1007/s10551-018-3999-7

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037//0022-3514.54.6.1063

White, K., & Peloza, J. (2009). Self-benefit versus other-benefit marketing appeals: Their effectiveness in generating charitable support. *Journal of Marketing*, *73*(4), 109–124. https://doi.org/10.1509/jmkg.73.4.109

Whitmarsh, L. (2023). Net zero: direct costs of climate policies aren't a major barrier to public support, research reveals. The Conversation. https://theconversation.com/net-zero-direct-costs-of-climate-policies-arent-a-major-barrier-to-public-support-research-reveals-210851

Whitmarsh, L. (2009). Behavioural responses to climate change: Asymmetry of intentions and impacts. *Journal of Environmental Psychology*, *29*(1), 13–23. https://doi.org/10.1016/j.jenvp.2008.05.003

Whitmarsh, L., & Lorenzoni, I. (2010). Perceptions, behavior and communication of climate change. *Wiley Interdisciplinary Reviews: Climate Change*, *1*, 158–161. https://doi.org/10.1002/wcc.7

Whitmarsh, L., & O'Neill, S. (2010). Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. *Journal of Environmental Psychology - J ENVIRON PSYCHOL*, *30*, 305–314. https://doi.org/10.1016/j.jenvp.2010.01.003

Whitmarsh, L., Poortinga, W., & Capstick, S. (2021). Behaviour change to address climate change. *Current Opinion in Psychology*, *42*. https://doi.org/10.1016/j.copsyc.2021.04.002

Woolley, K. & Ayelet Fishbach. (2015). The experience matters more than you think: People value intrinsic incentives more inside than outside an activity. *Journal of Personality and Social Psychology*, *109*(6), 968–982. https://doi.org/10.1037/pspa0000035

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175. https://doi.org/10.1037/0003-066X.35.2.151

Zhang, T., Gino, F., & Bazerman, M. H. (n.d.). *Exploring Simple Fixes to Our Moral Bugs*. 55.

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

Zhong, C.-B., & Liljenquist, K. (2006). Washing Away Your Sins: Threatened Morality and Physical Cleansing. *Science*, *313*(5792), 1451–1452. https://doi.org/10.1126/science.1130726

Zhong, C.-B., Liljenquist, K., & Cain, D. M. (2009). Moral self-regulation: Licensing and compensation. In *Psychological perspectives on ethical behavior and decision making* (pp. 75–89). Information Age Publishing, Inc.

VIII. Appendices

1. Appendix A: Moral Self Image Scale

Moral Self-Image Scale, Jordan et al (2015)[19]

1. Compared to the caring person I want to be, I am:

1	2	3	4	5	6	7	8	9
Much less caring than the person I want to be				Exactly as caring as the person I want to be				Much more caring than the person I want to be

2. Compared to the compassionate person I want to be, I am:

1	2	3	4	5	6	7	8	9
Much less compassionate than the person I want to be				Exactly as compassionate as the person I want to be				Much more compassionate than the person I want to be

3. Compared to the fair person I want to be, I am:

1	2	3	4	5	6	7	8	9
Much less fair than the				Exactly as fair as the person				Much fairer than the

[19] https://doi.org/10.3389/fpsyg.2015.01878

person I want to be	I want to be	person I want to be

4. Compared to the friendly person I want to be, I am:

1	2	3	4	5	6	7	8	9
Much less friendly than the person I want to be				Exactly as friendly as the person I want to be				More friendly than the person I want to be

5. Compared to the generous person I want to be, I am:

1	2	3	4	5	6	7	8	9
Much less generous than the person I want to be				Exactly as generous as the person I want to be				Much more generous than the person I want to be

6. Compared to the hard-working person I want to be, I am:

1	2	3	4	5	6	7	8	9
Much less hard-working than the person I want to be				Exactly as hard-working as the person I want to be				Much more hard-working than the person I want to be

7. Compared to the helpful person I want to be, I am:

1	2	3	4	5	6	7	8	9

Much less helpful than the person I want to be	Exactly as helpful as the person I want to be	Much more helpful than the person I want to be

8. Compared to the <u>honest</u> person I want to be, I am:

1 2 3 4 5 6 7 8 9

Much less honest than the person I want to be	Exactly as honest as the person I want to be	Much more honest than the person I want to be

9. Compared to the <u>kind</u> person I want to be, I am:

1 2 3 4 5 6 7 8 9

Much less kind than the person I want to be	Exactly as kind as the person I want to be	Much more kind than the person I want to be

2. Appendix B: Social Expectations in Study 1 and 2

1 **Imagine you decide to buy a new car 🚗.**
1.1 **How likely would you opt for an electric car? 🚗🚙🚘**

- ○ Extremely likely (1)
- ○ Somewhat likely (2)
- ○ Neither likely nor unlikely (3)
- ○ Somewhat unlikely (4)
- ○ Extremely unlikely (5)

1.2 **If we asked citizens in your social network the following:**

Do you have electric car?

In your view, what percentage answer YES and therefore have an electric car? 💭💭🤔

0 10 20 30 40 50 60 70 80 90 100

% of citizens who have an electric car ()	

1.3 **In your opinion, buying an electric car is:**

- ○ Extremely good (1)
- ○ Somewhat good (2)
- ○ Neither good nor bad (3)
- ○ Somewhat bad (4)
- ○ Extremely bad (5)

1.4 **If we asked citizens in your social network the following:**

Do you think that buying an electric car is the right thing to do?

In your view, what percentage of your network believe it is the right thing to do? 💭💭🤔

0 10 20 30 40 50 60 70 80 90 100

% of citizens who believe that buying an electric car is good ()	

Start of Block: Social Expectations Queue

1

On a beautiful Monday morning ☀️ , imagine you head out to start your day. You pass by your favourite coffee shop to grab your morning coffee ☕☕. You go there every morning for your latte and you know all the staff there.

It seems there are few customers who arrived before you.

1.1 **How likely would you stand in line and wait for your turn to grab your coffee?** ☕

- Extremely likely (1)
- Somewhat likely (2)
- Neither likely nor unlikely (3)
- Somewhat unlikely (4)
- Extremely unlikely (5)

1.2 **If we asked citizens in your social network the following:**

Would you wait in line for your turn?

In your view, what percentage answer YES and therefore wait in line? 💭💭🤔

0 10 20 30 40 50 60 70 80 90 100

% of citizens who wait in line ()	

1.3 **In your opinion, standing in queue or waiting in line is:**

- ○ Extremely good (1)
- ○ Somewhat good (2)
- ○ Neither good nor bad (3)
- ○ Somewhat bad (4)
- ○ Extremely bad (5)

1.4 **If we asked citizens in your social network the following:**
Do you think that waiting in line is the right thing to do?

In your view, what percentage of your social network believe it is the right thing to do?
💭 💭 🤔

	0 10 20 30 40 50 60 70 80 90 100
% of citizens who believe that waiting in line is the right thing to do ()	

End of Block: Social Expectations Queue

Start of Block: Social Expectations - Rapid Test

Now that social distancing limits have been lifted in England, Scotland and Wales. It is hoped that the COVID-19 rapid tests - known as lateral flow tests - can help stop individual cases from becoming outbreaks. Some businesses and public venues (e.g. **universities**) are also asking customers/ visitors to test before using their services.

Around 1 in 3 people with COVID-19 do not have symptoms. Rapid lateral flow tests help to find cases in people who may have no symptoms but are still infectious and can give the virus to others. The test could be taken at home, usually involves taking a sample from your throat and nose, or from your nose only, using a swab. You get a result in 15 to 30 minutes depending on the type of rapid lateral flow test you've taken.

It is recommended by the UK government to take a rapid lateral flow test twice a week (every 3 or 4 days).

Q160 **How likely are you willing to get the rapid test every 3 - 4 days before visiting or accessing public venue?**

- ○ Extremely likely (1)
- ○ Somewhat likely (2)
- ○ Neither likely nor unlikely (3)
- ○ Somewhat unlikely (4)
- ○ Extremely unlikely (5)

Q161 **If we asked citizens in your social network the following:**

Are you taking the rapid lateral flow test every 3-4 days (or twice a week) before visiting public venue?

In your view, what percentage answer YES and therefore take the rapid lateral flow test every 3-4 days? 💭💭🤔

	0 10 20 30 40 50 60 70 80 90 100
% of citizens who take a rapid test every 3-4 days ()	

Q162 **In your opinion, taking a COVID-19 rapid lateral flow test (every 3-4 days) is:**

- ○ Extremely good (1)
- ○ Somewhat good (2)
- ○ Neither good nor bad (3)
- ○ Somewhat bad (4)
- ○ Extremely bad (5)

Q163 **If we asked citizens in your social network the following:**

Do you think that taking the rapid lateral flow test (every 3-4 days) is the right thing to do?

In your view, what percentage of your social network believe it is the right thing to do? 💭💭🤔

0 10 20 30 40 50 60 70 80 90 100

% of citizens who believe that taking a rapid test twice a week is the right thing to do. ()	

End of Block: Social Expectations - Rapid Test

3. Appendix C: Donation Task (Behaviour 2) in Study 2

Start of Block: Donation

Q103 Thank you for participating in this study.

Before you go 🤳 .... we would like to announce the opportunity to win **a bonus of 10 GBP** 💰💰.

We will conduct a draw at the end of the study, and ***3*** *participants will get the chance to win 10 GBP each*.

If you were among the 3 winners, how much would you be willing to donate to **Cancer Research UK** ***https://www.cancerresearchuk.org/***

Please type the amount (from the 10 GBP) you would be willing to donate below (we will donate on your behalf and send you a confirmation).

4. Appendix D: Social Expectations in Study 3

Start of Block: SE

SE.1 **From 1 to 5, how likely are you to take a train instead of a plane (when possible) to save CO2 emissions and support climate change cause?**

- ○ Extremely unlikely 1 (1)
- ○ Somewhat unlikely 2 (2)
- ○ Neither likely nor unlikely 3 (3)
- ○ Somewhat likely 4 (4)
- ○ Extremely likely 5 (5)

SE.2 **Now, what about other people around you? What do you think they do?**

Out of 100 citizens in your social network, how many are likely to take a train instead of a plane (when possible) to save CO2 emissions and support climate change cause?

	0 10 20 30 40 50 60 70 80 90 100
% of your social network who do take trains instead of planes to save CO2 emissions. ()	

SE.3 **In your view, taking the train instead of a plane (when possible) to save CO2 emissions is the right thing to do.**

- ☐ Strongly disagree (1)
- ☐ Somewhat disagree (2)
- ☐ Neither agree nor disagree (3)
- ☐ Somewhat agree (4)
- ☐ Strongly agree (5)

SE.4 **Now, what do you think others approve of?**

If we ask citizens from your social network the following:

"Do you think that taking the train instead of a plane (when possible) to save CO2 emissions is the right thing to do?

In your view, what percentage of your network believes it is the right thing to do and approve of this behaviour?

	0 10 20 30 40 50 60 70 80 90 100
% of your social network that approve of saving CO2 emissions ()	

End of Block: SE

5. Appendix E: Behaviour 1 in Study 3

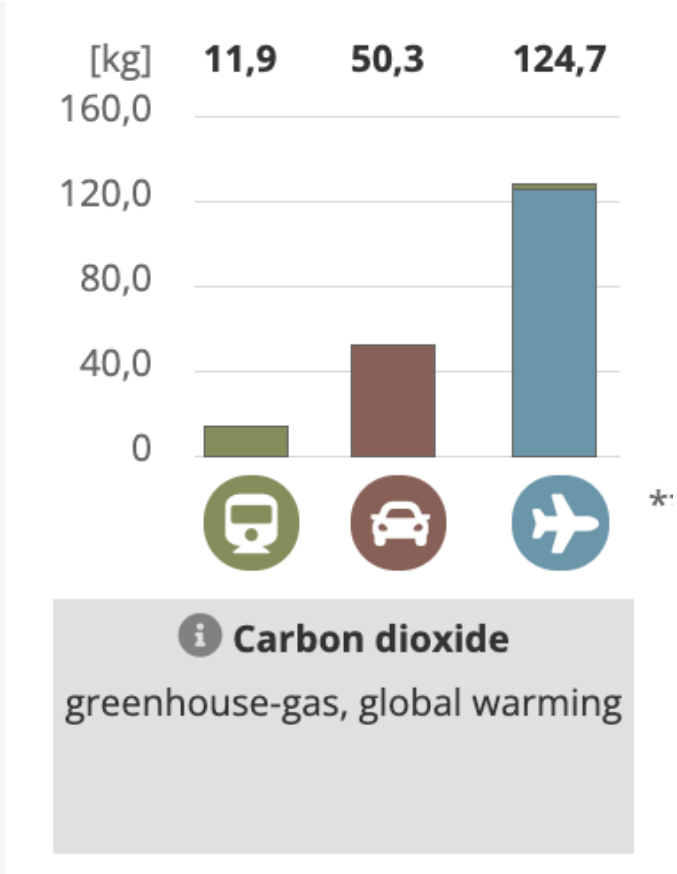
Plane Condition:

Now... we want you to imagine 💭💭 you are planning a holiday from London 🇬🇧 to the Netherlands 🇳🇱.

On average, It takes 1 hour and 15 min to fly from London to Amsterdam by plane ✈️ (without airport formalities).

However, it takes 4 hours to get to Amsterdam from London by train 🚂🚆 , but it saves 19 times on CO2 emissions ♻️ Air travel ✈️🛬 is now the fastest-growing contributor to global warming ☀️ and avoiding a flight when there are easy alternatives is often the most significant single thing that any individual can do to cut their carbon footprint and limit their impact on the environment ♻️ .

Travelling from London 🇬🇧 to Amsterdam 🇳🇱 , by plane ✈️ leads to approximately 124 Kg of CO2 emissions, compared to almost 11 Kg of CO2 emissions by train 🚆 .



<u>Imagine 💭 you decide to take the plane ✈️✈️</u> instead of the train from London to the Netherlands to save time.

By doing so, you would be responsible of producing almost 100 Kg of CO2 emissions!

Train Condition:

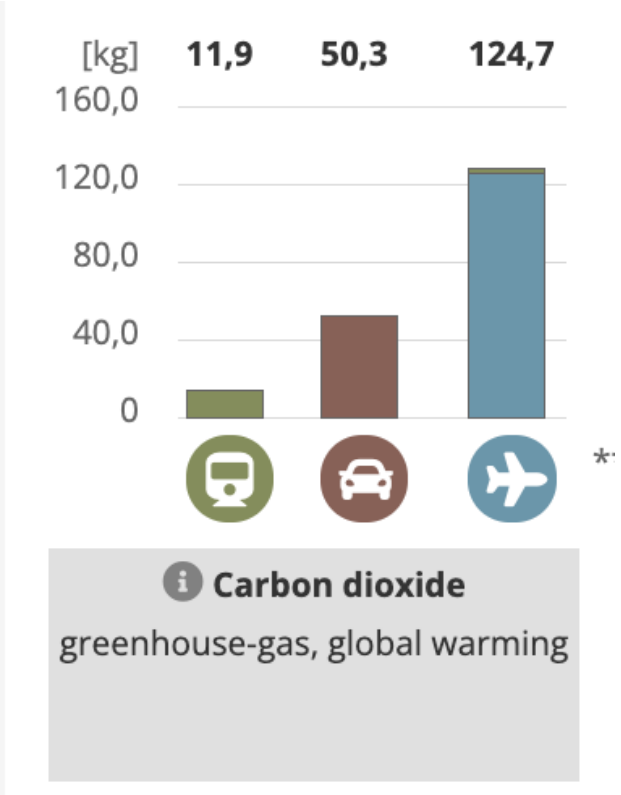
Now... we want you to imagine 💭💭 you are planning a holiday from London 🇬🇧 to the Netherlands 🇳🇱

On average, It takes 1 hour and 15 min to fly from London to Amsterdam by plane ✈️ (without airport formalities).

However, it takes 4 hours to get to Amsterdam from London by train 🚂🚊 , but it saves 19 times on CO2 emissions.

Air travel ✈️🛬 is now the fastest-growing contributor to global warming ☀️ and avoiding a flight when there are easy alternatives is often the most significant single thing that any individual can do to cut their carbon footprint and limit their impact on the environment ♻️ .

Travelling from London 🇬🇧 to Amsterdam 🇳🇱 , by plane ✈️ leads to approximately 124 Kg of CO2 emissions, compared to almost 11 Kg of CO2 emissions by train 🚂 .



Imagine 💭 you decide to take the train 🚂 instead of the plane from London to the Netherlands to save on CO2 emissions and take action to combat climate change ♻️ 🌍 .

By doing so, you would have saved almost 100 Kg of CO2 emissions 👏 ⭐ 🌍

6. Appendix F: Carbon Emission Task (Behaviour 2) in Study 3 and 4

Start of Block: Carbon Emission Task (CET)

Q28 Thank you for participating in this study.

Before you leave, we are planning to have a draw and choose 10 participants and give them a bonus as it is the holiday season. **The bonus is 10 GBP.**

We will conduct the draw at the end of the study, randomly choose 10 participants, and give the ten lucky winners a bonus of 10 GBP each.

Moreover, we are also giving a chance for the winners to choose whether they want to donate part or all of their bonus for carbon offsetting by buying CO2 certificates.

What is Carbon offsetting or buying CO2 certificates?

Carbon Offsetting/ buying CO2 certificates allows you to compensate for the carbon dioxide and other greenhouse gas emissions you produce by reducing emissions elsewhere. To compensate for your emission footprint, you can purchase carbon offset credits equivalent to the amount of emissions you were responsible for. Your credits will be invested in certified carbon reduction projects.

The table 29 below shows the impact of different options on the amount of CO2 offset and the eligible bonus.

Table 29. Carbon Emission Task (CET)

You are willing to forego (GBP)	£0	£1	£2	£3	£4	£5	£6	£7	£8	£9	£10
Equivalent CO2 Certificates	No CO2 certificate	80 Kg of CO2	150 Kg of CO2	230 Kg of CO2	300 Kg of CO2	380 Kg of CO2	450 Kg of CO2	530 Kg of CO2	600 Kg of CO2	680 Kg of CO2	750 Kg of CO2
Eligible Bonus	£10	£9	£8	£7	£6	£5	£4	£3	£2	£1	£0

Based on your input, we will buy the CO2 certificates and give you the rest as a

bonus. If you were one of the lucky winners, out of the 10 GBP, how much are you willing to forego in carbon offsetting by buying CO2 certificates?

- ○ 0 (1)
- ○ 1 (2)
- ○ 2 (3)
- ○ 3 (4)
- ○ 4 (5)
- ○ 5 (6)
- ○ 6 (7)
- ○ 7 (8)
- ○ 8 (9)
- ○ 9 (10)
- ○ 10 (11)

End of Block: Carbon Emission Task (CET)

7. Appendix G: Pilot Study 4

Climate Change_VS_Public_Pilot

PN.1 Do you think that taking action to support the climate change cause is the right thing to do?

- ○ Strongly disagree (5)
- ○ Somewhat disagree (6)
- ○ Neither agree nor disagree (7)
- ○ Somewhat agree (8)
- ○ Strongly agree (9)

SE.2 **Now, what do you think others approve of?**

If we ask citizens from your social network the following:

"Do you think that taking action to support the climate change cause is the right thing to do?

In your view, what percentage of your network believes it is the right thing to do and approves of this behaviour?

	0 10 20 30 40 50 60 70 80 90 100
% of your social network that approve of taking action to support climate change cause ()	

SE.2 **Now, what do you think others actually do?**

Out of 100 citizens in your social network, how many are likely to take actions that support climate change cause?

	0 10 20 30 40 50 60 70 80 90 100
% of your social network who take actions to support the climate change cause ()	

End of Block: Social Expectations

Start of Block: Public

Q164
Now... we would like you to please do the following steps:

1- Visit your preferred social media account (e.g., facebook, twitter, instagram, or others)
2- Look for posts related to #climatchange.
3- Like or reshare a post to show your support for this cause on your social network.
4- Once you do so, please come back to the survey to complete the questions.

Q39 **Did you share or like or reposted a social media post related to climate change?**

- ○ No (21)
- ○ Yes (22)

Q136 **Once you have re-shared or liked a post supporting climate change, please write down below what the post was about?**

_____

Q32 **Thank you for participating in this survey.**

8. Appendix H: Behaviour 1 in Study 4

Public Treatment

We would like you to please do the following steps:

1- Visit your preferred social media account (e.g., Facebook, Twitter, Instagram, or others)

2- Look for posts related to #climatechange.

3- LIKE or POST or SHARE a post to show your support for this cause on your social network.

4- Once you have done so, please come back to the survey to complete the questions.

Control

We would like you now to please do the following steps:

1- Visit your preferred social media account (e.g., Facebook, Twitter, Instagram, or any other)

2- Look for a post related to #climatechange and read it

3- Once you have done so, please come back to the survey to complete the questions.

Private Treatment

A pledge is a personal commitment to abide by a rule or to live up to a particular standard of behaviour. For example, "I will recycle more in the coming year".

Recycling and Climate Change

As both climate change and waste are ever-growing problems around the world, it is of great importance to understand how these two major issues are also affecting each other.

Recycling is beneficial to the climate crisis in two main ways: by limiting the number of raw materials being used and limiting the amount of waste going into landfills.

Estimates show that recycling between 2020 and 2050 will reduce emissions by 5.5 - 6.02 gigatons of carbon dioxide (equivalent to taking over 1 billion cars off the streets for one year). That is a major reduction from a simple lifestyle change, which makes recycling an effective, yet easy change to make to help curb greenhouse gas emissions and help limit the climate crisis.

Now, after reading all of the above, we ask that you consider pledging to significantly increase the extent to which you recycle in the coming year.

If you agree to this pledge, you will be stating that you are committed to recycling more aluminium cans, plastic bottles, paper products, and other recyclable materials in the current year.

If you agree to take this pledge, please click "I pledge" below.

If you do not agree to take this pledge, please click "I do not pledge" below.