Health Messages, Mechanisms & Measurement: Essays in Behavioural & Experimental Economics

Sarah Bowen

University of Nottingham

School of Economics

October, 2022

Thesis submitted to the University of Nottingham in accordance with the requirements for the degree of

Doctor of Philosophy

This work was supported by Economic and Social Research Council

[Grant Number: ES/P000711/1]

Sarah Bowen: *Health Messages, Mechanisms & Measurement: Essays in Behavioural & Experimental Economics*, PhD in Economics, Supervised by Professor Chris Starmer and Professor Abigail Barr © June 2022.

Intellectual Property Statement

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

First, I would like to thank my supervisors, Chris Starmer and Abigail Barr for their mentorship, guidance, and kindness during this PhD. Thank you both, for your patience and your expertise. Under your joint supervision, I have developed not only my skills as a researcher, but also my confidence. I also wish to thank Eamonn Ferguson and Tracey Thornley, who have both been wonderful collaborators and mentors over the duration of the PhD.

Without the financial support from the ESRC, CeDEx, NIBS, and the School of Economics, this thesis would not have been possible. Importantly, I wish to acknowledge and thank Suzanne Robey, José Guinot Saporta, Hayley McCalla, and all the administrative staff in the School of Economics for their incredible support.

To the friends old and new who supported me at every stage of the PhD (the good, the bad, and the ugly), I have only the deepest gratitude. For every crisis of confidence, you were there. I cannot thank you enough for the laughs, adventures, and memories; I will carry them with me always. To my wonderful extended PhD cohort, Anna Hochleitner, Candice Ford, Chris Dawes, Lara Suraci, Malte Baader, Pedro Luís Marques Correia da Silva, Richard Mills, and Yuliet Verbel, you have been the best part of this PhD. It takes a village to raise a thesis, so I wish to thank, Alice Neale, Orla Gilson, Pippin Burkett, Sarah Carling, Georgie Palmer, Laurie Dempsey, and Betsy Stirland for their encouragement and friendship over the years. I would like to thank Kiki for being an excellent conversationalist and companion. Also, a special thanks to Will for the endless supply of peppermint tea and for telling me I could do it until I believed it too.

To Mum, Dad, and Greg, I love you to the end of all the numbers plus a thousand. You are the backbone of everything I have done and will do. Thank you for your endless support, helping me when I could not help myself, and for being my most enthusiastic champions.

Finally, I wish to acknowledge and express my thanks to the members of the 1752 group for affording me catharsis in my darkest moments. This thesis would not exist if not for your advocacy and activism.

Contents

List of Figures

List of Tables

Introduction

This thesis contributes to the development and evaluation of robust message interventions to improve patient outcomes within public health. Across three chapters, I use experimental methods to examine message framing effects and investigate the validity of self-reports to evaluate message intervention effectiveness. The broad objective of this thesis is to study mechanisms that mediate the effects of message framing on both actual behaviour and self-reports of behaviour. Hence, while the application domain is public health, the results may interest behavioural scientists and policymakers interested in going beyond studying "what works". By doing so, I contribute to expanding literature studying the mechanisms underpinning the effects of messages designed to change behaviour.

The thesis is divided into two parts. The two-part structure reflects the structure of my PhD, where I explored two research agendas differing in terms of aims, objectives, and research domains. Part I is comprised of Chapters 1 and 2. Chapter 1 investigates a potential weakness of self-reports as a measurement tool to evaluate the true behavioural effect of a nudge message intervention. This is relevant specifically, but not exclusively, to the development of behavioural interventions to support greater patient adherence to long-term medications. In the absence of a gold-standard measure of medication adherence, self-reports are often the most, and sometimes the only, feasible measure of adherence.

Furthermore, we demonstrate a novel experimental methodology to measure the direct effect of a nudge message on self-reports when medication-taking behaviour cannot yet have changed due to a nudge message. We find, using three experiments run over three years, that nudge messages can produce a replicable and systematic upward

bias in self-reports of adherence which we will later refer to as a "nudge bias". Additionally, in this chapter, we demonstrate that facilitating subject recall by time-framing the self-report can reduce, although not entirely prevent, nudge bias in self-report data.

Motivated by observations of heterogeneity in Chapter 1: that different nudge messages produced different amounts of nudge bias in self-reports, Chapter 2 investigates the moderating and mediating mechanisms behind nudge bias. As potential moderators, we experimentally investigate social desirability concerns, individual processing style (concerning affect and cognition), and predictors of information processing via a central or peripheral route. Affective and cognitive evaluations of the messages are also investigated as mediators of nudge bias. The results of Chapter 2 provide no evidence to suggest that social desirability bias, individual processing style, or predictors of information processing route can fully explain nudge bias. However, we do find evidence that different factors drive nudge bias in time-framed and non-time-framed self-reports. For example, we find subjects' preference for engaging with stimuli that provoke emotional response moderates the effect of some nudge messages on non-time-framed self-reports, although not in the direction consistent with nudge bias. On the other hand, we find that for some nudges, personal involvement with the issue of improving medication adherence moderates nudge bias in time-framed self-reports.

In summary, Part I of the thesis develops and presents a novel experimental design to detect and measure nudge bias in self-reports of adherence. Further, we demonstrate that facilitating recall at the point of self-report can reduce nudge bias. However, the determinants of nudge bias remain unexplained. Our investigation reveals that controlling for social desirability bias in the analysis is insufficient to control for nudge bias in self-report data. The results have implications for developing a robust evaluation of behaviourally informed communications designed to support behavioural change. More specifically, policymakers should consider and control for the potential effect of nudge bias in self-reports to ensure the robust evaluation of nudge messages as effective behaviour change interventions. Therefore, our study contributes to the existing research on nudge interventions by highlighting the risk of nudge bias inherent

in self-report data and further providing policymakers with a methodology to measure and control nudge bias directly.

Part II of the thesis builds on the theme of designing and evaluating behaviourally informed messages, however, with an application to a different public health domain and a different messaging objective. The aim of Chapter 3 is to identify the most effective combination of frames to communicate the recent change in the UK's blood donor selection policy. In June 2021, the UK blood service began evaluating blood donor eligibility based on individual sexual behaviour rather than sexuality. Before the implementation of these changes, we conducted an online experiment to investigate the effectiveness and interaction between altruism (donor vs recipient vs both) and risk frames (reduced risk vs increased safety). We adopt an approach-avoidance framework of motivation to evaluate the effects of message framing on motivations to donate. The main finding is that altruism frames combined with communications that highlight the effect of the policy change on improved safety reduce deterrence from donating blood relative to non-altruistic and risk-framed communications. The results of Chapter 3 are particularly relevant considering that blood services in many other countries are expected to follow the UK and Canada's (in September 2022) example by transitioning towards a more individualized selection policy.

In summary, Part II of this thesis provides timely evidence that, where possible, blood services should communicate policy changes using altruistic frames with a focus on improvements to safety over donor and risk reduction focused frames. Our findings support and are consistent with growing evidence that altruism frames can enhance health behaviours requiring cooperation.

Part I

Chapter 1

Nudging self-reports of adherence

1.1 Motivation

Poor adherence to medications, or the frequent deviation from the agreed recommendations between patient and prescriber, is an unmet behavioural public health challenge (Kleinsinger, 2018).[1] Although, medicines adherence is a primary determinant of treatment success (Jimmy & Jose, 2011), it is estimated that as many as 50% of patients sub-optimally adhere to their prescribed regimes (De Geest & Sabaté, 2003).

As medicines adherence is, in-part, a behavioural problem, many researchers have asked whether patients can be "nudged" to improve their adherence (Dai et al., 2017; Glasgow et al., 2021; Heinrich & Kuiper, 2012; Jachimowicz et al., 2019; Kamal et al., 2015; Kimmel et al., 2012; Kwan et al., 2020; Luong et al., 2021; Raiff et al., 2016; Ramanath et al., 2012; Reese et al., 2016; Roseleur et al., 2019; Tao et al., 2015; Volpp et al., 2008; Waughtal et al., 2021; Zhao et al., 2012). A nudge is defined as "any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler & Sunstein, 2008). To evaluate whether a nudge indeed increases adherence, researchers often rely on patient self-reports. Objective measures of medication adherence are of course, also used to evaluate adherence interventions (such as pill counts, electronic caps/containers, blood tests, etc.), however, these measures are expensive, difficult to access, and not suitable for all therapy types. The basic assumption underlying studies using self-report measures is that, within an experimental setting, only the behavioural response to an intervention can lead to improved self-reports. However, a concern is that nudges can increase self-reports of adherence without first influencing medication-taking behaviour. If this is true, reliance on self-report data can result in spurious correlation, overestimating the effectiveness of nudges and leading to sub-optimal policy recommendations.

[1] Note that the degree to which a patient's behaviour must match the agreed recommendations of their prescriber to be classified as "adherent" can vary across therapy type and illness group (Burnier, 2019). In general, a good level of adherence is defined as taking medicines exactly as prescribed $> 80\%$ of the time (De Geest & Sabaté, 2003; Osterberg & Blaschke, 2005).

So far, previous literature has not addressed this issue in a systematic way. It is, therefore, crucial to develop a better understanding of how reliance on self-reports affects evaluation outcomes. This chapter aims to close this gap and test whether self-reports of medication-taking behaviour accurately evaluate nudge interventions designed to improve adherence. Using a series of online experiments, we find that self-reports can indeed suffer from a nudge bias (i.e., where people increase their level of self-reported adherence without changing their behaviour). While we find heterogeneity in the effect of different nudges on self-reports, our results suggest policy makers should be cautious when evaluating interventions based on self-reported outcomes alone.

Previous research has long acknowledged the central role of medication adherence for patient outcomes, making identifying cost-effective interventions a public health priority (Haynes et al., 2002). Despite the demand, few effective strategies have been identified. Of 182 relevant randomised controlled trials (RCTs), only 17 studies had a low risk for methodological bias and only 5 interventions resulted in improvements in adherence (Nieuwlaat et al., 2014).[2] However, effect sizes are small, the interventions complex, with multiple components, and there are no common characteristics between them (Nieuwlaat et al., 2014). These mixed results show there is still demand for research informing the development of adherence interventions and highlights the importance of methodological foundations to increase comparability and advance knowledge.

As mentioned above, one approach identified as a potential cost-effective adherence intervention is the behavioural change intervention (BCI), with nudges as one example (Cross et al., 2020). In general, BCIs are characterised by their utilisation of psychological and behavioural economic tools in intervention content and delivery (Beard et al., 2019) and cost-effectiveness (Gordon et al., 2007; National Institute for Health and Clinical Excellence (NICE), 2007; Shepherd et al., 2010). Broadly defined, nudging techniques can include the act of sending a message. Where a message is "a

[2]Methodological weaknesses include poor and variable measurement (making a comparison across studies difficult), sub-optimal sample design (i.e., not just recruiting participants because they are willing, but because their adherence is low), and a lack of theoretical underpinning.

short piece of information that you give to a person when you cannot speak to them directly" (Cambridge University Press, n.d.). There are many features of a message that can be designed to more effectively trigger psychological processes leading to behaviour change, such as the content of the message, the message sender, the frequency and timing of message delivery, and the mode of delivery (i.e., a text message vs a letter).

Nudge message interventions have been shown to influence behaviour (e.g., compliance with COVID-19 government guidance (Hume et al., 2020; Sasaki et al., 2021)). While many studies have evaluated the effectiveness of nudge message interventions designed to support medicines adherence, it is difficult to draw general conclusions about the characteristics of effective interventions due to important methodological limitations. The biggest problem is measurement quality, with many studies relying on self-reports. Self-reports in general, have been used to evaluate the effect of nudges on adherence alone (Heinrich & Kuiper, 2012; Jachimowicz et al., 2019; Kamal et al., 2015) or in combination with other self-report instruments (Ramanath et al., 2012).

Although objective measures of medication adherence exist (i.e., by proxy measures such as Medication Adherence Monitoring Systems (MEMs) and blood tests, or secondary data measures such as pharmacy refill rates and dispensary data), normally the most straightforward measure is a patient's self-report (Brian Haynes et al., 1980; Jimmy & Jose, 2011; Walsh et al., 2002). The reason for this is that objective measures can be expensive (MEMs), difficult to access (pharmacy data), and may not be suitable for all medication types and poly-pharmacy patients (Arnet et al., 2013).[3] As a result of these measurement challenges, no gold standard exists for measuring medication-taking behaviour (National Institute for Health and Clinical Excellence (NICE), 2009). However, the benefits of using a self-report tool over an objective measure should not be understated. In addition to being low-cost and easy to implement, self-reports can provide additional contextual information about behaviour not often captured by ob-

[3]For example, Trackcaps (or SmartCaps) are not suitable for liquid medications or large prescription pills.

jective measures. For instance, self-reports are currently the only way to assess and differentiate distinct types of non-adherence across patients. One widely used dichotomy to characterise non-adherence is unintentional vs intentional non-adherence. Unintentional (e.g., forgetfulness) and intentional non-adherence (e.g., skipping a dose) are driven by different factors, including patient beliefs, demographic characteristics, and habit strength (Bae et al., 2016; Furniss et al., 2014; Gadkari & McHorney, 2012; Müller et al., 2015; Thorneloe et al., 2018; Wroe, 2002). As such, self-reports offer practitioners a valuable diagnostic and evaluative tool to tailor interventions to patient-specific barriers.

Given the advantages of self-reports, it is surprising that there are no nudge evaluation studies (as far as we are aware) that measure nudges' effect on adherence using self-reports and objective measures. As such, there is no data to assess the validity of the self-report as a measurement tool to evaluate nudge effectiveness. In this chapter, we argue that self-reports can be a valid measure of behaviour so long as self-reported behavioural improvements are evidence of actual behaviour improvement. There is a theoretical concern that a nudge might encourage a person (consciously or unconsciously) to change their response to a self-report without there being a corresponding change in actual behaviour (e.g., a nudge highlights the harm an individual's non-adherence can have on others; the patient feels more uncomfortable about revealing the extent of their adherence failure).[4] To test the validity of self-reports, we design an environment where a nudge can change self-reports of medication adherence, but not

[4] Another way of modelling the potential effect of a nudge on self-reports is to understand self-reports as a low-cost opportunity for identity performance (Brenner & DeLamater, 2016). People are motivated to perform valued identities (Brenner & DeLamater, 2014). There are many ways in which a person can perform an identity. A respondent who values medication adherence and sees themselves as the "kind of person" who is adherent may not enact the adherent identity at a rate consistent with the identity given the costs of its enactment. Therefore, if a respondent interprets a question about their adherence (self-report) as one about identity rather than behaviour, over-reporting may be influenced by a desire for consistency between the "ideal" and "actual" self. A nudge in this setting may increase this desire for identity consistency via increasing the salience of an individual's identity disparity or increasing the salience of the self-report as a low-cost opportunity to perform the "ideal self".

behaviour.

There are two outcomes of our experimental test. From this point onwards, we will refer to self-reports as "nudge-robust" if they are insensitive to nudges in an environment where behaviour cannot yet change. Alternatively, we will refer to a "nudge bias" if nudges systematically influence self-reports in an environment where behaviour cannot yet change.

The self-report tool we investigate in this study is the Medication Adherence Rating Scale (MARS-5) (Chan et al., 2020). The MARS-5 is among the most widely used self-report adherence measures and has been validated across patient groups and languages (Lee et al., 2017; Mahler et al., 2010; Manullrdby et al., 2012; Scribano et al., 2019; Sjölander et al., 2013; Stentzel et al., 2018; Stone et al., 2021; Thompson et al., 2000).[5] Unlike other self-report measures, the scale is free and does not require a license to use (Tesfaye & Peterson, 2022). In addition, it allows us to differentiate between different patient-specific barriers to adherence, namely intentional and unintentional adherence.

Across 3 online experiments, we investigate whether self-reports collected with the MARS-5 are nudge robust. Methodologically, we start from an environment where a nudge bias seems most likely to emerge and then try to identify a nudge-robust self-report. Study 1 is designed as a fundamental test for the existence of a nudge bias in self-reports. We use a more open-ended version of the MARS-5 to facilitate greater flexibility in a subject's memory retrieval. To the extent that a nudge bias is a realisation of established response biases, such as socially desirable reporting and recall bias, open-endedness in the implementation (i.e. not facilitating respondent recall) of self-reports may encourage nudge bias in self-report data. We refer to the self-report in Study 1 as non-time-framed (NTF-self-report).

Following Study 1, Study 2 provides a tougher test for the presence of nudge bias. We achieve this by implementing a version of the self-report generally considered less

[5]There are a small number of exceptions. The MARS is not a valid measure of adherence in the following patient groups: children with Asthma (Garcia-Marcos et al., 2016) and COPD patients taking inhalation medication (Tommelein et al., 2014).

susceptible to response bias distortions. We ask respondents to report behaviour within a recent and specific-time frame (TF-self-reports). Facilitating recall is one of many implementation recommendations shown to reduce the potential for response bias (Stirratt et al., 2015) and increase the accuracy of self-report data (Clarke et al., 2008; Dalziel et al., 2018; Kjellsson et al., 2014; Stull et al., 2009).[6] Therefore, in terms of nudge bias, we expect TF-self-reports to reduce the influence of selective memory as a vehicle for socially desirable reporting. Therefore, TF-self-reports should be more nudge-robust than NTF-self-reports.

Finally, Study 3 tests whether we can replicate findings from Studies 1 and 2 and provides a well-powered test for the validity of self-reports. Using the results of Study 1 and Study 2 as pilots, we pay special attention to the sample design in Study 3 such that the test is powered to detect the marginally significant effects in Study 1 and 2. Finally, we pool the data across all studies to provide evidence of a systematic bias in self-reports: subjects tend to overstate their adherence after being nudged. While TF-self-reports appear to reduce the size of the nudge bias observed, they are not completely robust.

The remainder of this chapter is organised as follows: we present the experimental design, procedures, and results for Experiments 1, 2, and 3 conducted on independent samples in Section 1.2, 1.3, and 1.4, respectively. Then, to increase reliability and achieve greater statistical power, in Section 1.5, we pool and analyse the independent samples across each experiment. Finally, in Section 1.6, we provide a general discussion.

1.2 Study 1

1.2.1 Design

Study 1 tests for nudge bias in self-reports of medicines adherence. In a typical nudge evaluation experiment, the intervening period (time between intervention and outcome measurement) must be sufficiently long to allow behaviour to change to detect an inter-

[6] Alongside using a validated scale and employing technological delivery.

vention effect with the measurement instrument. Our approach differs (see Figure 1.1). As in the standard approach, we introduce a nudge designed to stimulate changes in (future) behaviour. However, at this point, we divert from the standard approach and immediately ask participants to report their (past) adherence behaviour. If nudges systematically influence self-reports, improved self-reports cannot reflect an actual change in behaviour and instead reveal a weakness of the measure. In other words, any difference between nudge and control group subjects indicates self-reports are impacted directly by the nudge in a context where, by design, there can have been no adjustment in adherence behaviour. The design allows us to assess the following hypothesis:

H1: NTF-self-reports of adherence are nudge biased.

Consistent with this hypothesis, we expect to observe a systematic improvement in self-reported adherence between nudged subjects and the control group.

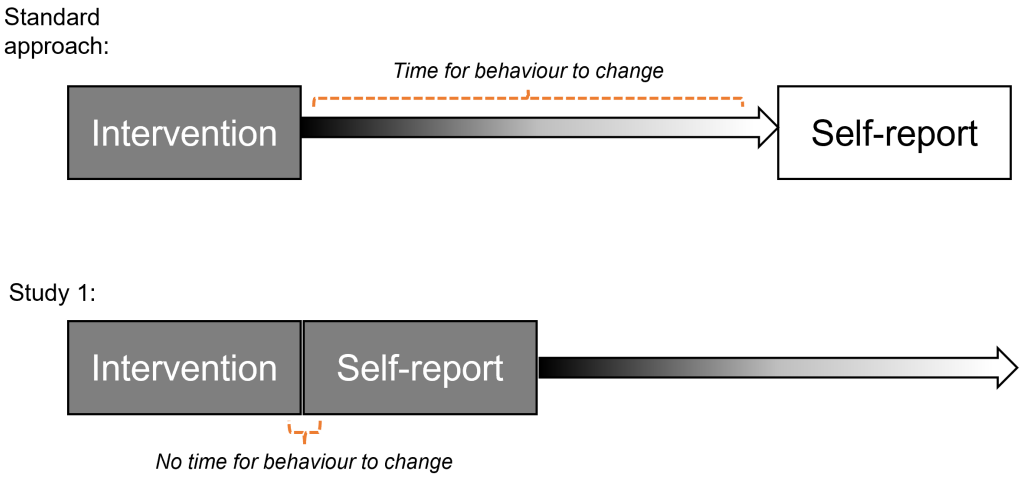


Figure 1.1: Experimental design as compared to the standard intervention evaluation approach

We exogenously manipulate the message shown to subjects immediately before measuring medicines adherence via self-report by evenly randomising subjects to a message condition at the beginning of the experiment. The between-subjects design allows us to investigate the consistency and degree of potential nudge bias across nudges

employed by different behavioural change techniques.

The messages are summarised in Table 1.1.

Label	Message
Control	Medication Adherence Experiment
Health*	Not taking medication as prescribed could endanger patient health
£-NHS*	The NHS loses £300 million per year from wasted medication
Burden-NHS	Not taking medication as prescribed places a significant burden on our NHS
Descriptive norm	Results from a recent experiment found that more than 70% of patients with chronic illnesses, take their medication as prescribed.

* adapted from Jachimowicz et al. (2019)

Table 1.1: Message conditions

In Study 1, the messages are taken either from existing medicines adherence nudge interventions or are informed by the behaviour change literature. We show subjects in the control group the name of the study instead of a nudge.[7]

The health and £-NHS messages are adapted from a prior behavioural intervention designed to support adherence within a patient's first months of a newly prescribed long-term medication (Jachimowicz et al., 2019). In this study, patients in the treatment group received a commitment sticker alongside their first dispensed prescription of a new medication, either augmented with a behavioural message or not.[8] The authors find that the commitment sticker increased adherence only when presented with a message highlighting the consequence of non-adherence incurred by the patient to their health. Patients who received the health-message commitment were 1.59 times more likely to report being adherent than patients who received no intervention at all.[9] The commitment intervention without a message and with a message highlighting the

[7]For more detail about how messages appeared to subjects in the experiment, see Appendix A.1.1.

[8]The authors designed the intervention to test whether making a commitment to take medication exactly as prescribed at the start of a new regime would lead to better adherence.

[9]The effect of the intervention on patient adherence was assessed via patient self-reports using the Morisky Medication Adherence Scale (Morisky & DiMatteo, 2011).

financial cost of wasted medication to the NHS did not significantly improve adherence.

The burden-NHS message is informed by the behavioural change technique "salience of consequences"; in that motivation to act is increased when the costliness of non-adherence behaviour is made salient (Carey et al., 2018; Lindenberg & Papies, 2019). The difference between the health, £-NHS and the burden-NHS message is in how we frame the currency (financial vs health vs burden) and bearer of cost (the patient (self) vs the NHS (self and others)). The nudge messages health, £-NHS, and burden-NHS thus highlight different but highly correlated aspects of the same attribute (Ungemach et al., 2018).[10]

Finally, the descriptive norm message uses social norms to promote adherent behaviour. Normative peer behaviours and attitudes are key predictors of health behaviours (Ajzen, 1985; Gerrard et al., 2008). Therefore, increasing the salience of a new or existing normative behaviour is a strategy widely applied to behaviour change interventions. Social normative interventions work on the premise that behaviour is determined in part by beliefs about what other people do (descriptive norm) and what they ought to do (prescriptive norm) (Bicchieri, 2006; Dempsey et al., 2018). The descriptive norm message informs subjects about the behaviour (being adherent to medication) of most of their peers (patients with long-term conditions). The 70% statistic used in the message is adapted from research assessing patients' medication adherence in the New Medicine Service (NMS) (Elliott et al., 2016).[11]

Self-report measure

Self-reports of medication adherence are measured using the 5-item Medication Adherence Report Scale (MARS-5) (Chan et al., 2020). The MARS-5 assesses how adherent subjects are based on the frequency (on a 5-point Likert scale from "Never" to "Always") with which they diverge from their prescribed regime. The items are designed to measure separately intentional (change dose, stop medication, skip a dose, use less

[10] We did not additionally test a version of the message highlighting the patient's personal financial cost of non-adherence for statistical power reasons.

[11] The NMS was the setting of the Jachimowicz et al. (2019) study mentioned above.

than prescribed) and unintentional non-adherent behaviour (forget dose) (Horne & Weinman, 2002). In Study 1, we implement the MARS-5 with no time-frame (NTF) recall instructions (see Appendix A.1.2.).

Adherence is usually assessed as the sum of responses to each item (continuous score) or by a cut-off score to code individuals as either adherent or non-adherent (dichotomous score). Continuous scores are bound between 5 (perfect non-adherence) and 25 (perfect adherence), where a higher score indicates a greater degree of adherence to medication.[12] Under a dichotomous scoring method, respondents with a score of 25 are typically coded as adherent, and those with a MARS score of less than 25, as non-adherent. The cut-off score used to code respondents as adherent is arbitrary and can vary. In most cases, the threshold score falls between a continuous score of 20 and 25. In this experiment, we use a cut-off of 25 to create a dichotomous adherence score. In general, continuous scores are the most favoured MARS-5 scoring method. However, an argument against using a continuous score is that the MARS-5 does not measure a single underlying construct. It can be difficult to interpret the correlation between the continuous score and an objective measure (Sandy & Connor, 2015). In addition to overall adherence, the MARS-5 allows for separately assessing intentional and unintentional adherence. Intentional non-adherence scores are calculated as the mean response to the four intentional items: change, stop, skip, and use less. As unintentional non-adherence is assessed using a single item (forget dose), the response to this item is used as a score of unintentional non-adherence. Intentional and unintentional scores range between 1 and 5. Notably, as with most self-reports of adherence, the MARS-5 instrument is validated only as a measure of overall non-adherence. There is limited validation literature on self-reports as measures of intentional and unintentional barriers to adherence (Müller et al., 2015). Therefore, we use intentional, unintentional, continuous, and dichotomous adherence scores as outcome measures to assess the nudge

[12] As the questionnaire is framed in terms of how often people do not adhere, the MARS-5 item are reverse coded as follows: 5 = Never, 4 = Sometimes, 3 = About half the time, 2 = Most of the time, 1 = Always.

robustness of the NTF MARS-5.

Additional measures

In addition to self-reports, we measure subject characteristics associated with medication-taking and reporting behaviour. Before the experimental manipulation and outcome measure (message exposure and self-report), we collect data on demographics and illness types. Post manipulation, we collect more information about subject's medication regime and assess subject's proclivity to respond in a socially desirable manner.

Pre-manipulation

Demographics: Subjects complete a demographic questionnaire to assess gender identity (Female/ Male/ Non-binary or non-conforming), age, the highest level of completed education (No qualifications/ GCSEs or equivalent/ A-levels or equivalent/ Undergraduate degree or equivalent/ Postgraduate degree or equivalent/ Other) and household income (up to £9,999/ £10,000 to £24,999/ £25,000 to £49,999/ £50,000 to £99,999/ £100,000 and above). In addition, we measure subjective socio-economic position using the MacArthur ladder scale of Subjective Social Status (Adler et al., 2000; Adler & Stewart, 2007). The MacArthur scale requires subjects to place themselves on a 10-rung ladder where 1 is the bottom rung, representing those in the UK who are the worst off in earnings, education, and employment, and 10 is the top rung, representing those who are the best off.

Illness type: We ask participants to indicate one illness for which they are prescribed long-term medication. We instructed subjects to think only of this illness and prescribed course of therapy when responding to questions about their medication-taking behaviour in the experiment. We provide respondents with a non-exhaustive list of chronic conditions and the option to indicate "Other" and self-specify using a free-text response.

Post-manipulation

Medication regime: We ask subjects how long they had been prescribed their medication (0-6 months/ 7-11 months/ 1-3 years/ 3+ years), how they access healthcare

in the UK (public (NHS) only/ Private only/ Both Public and Private), and how important strict adherence to their medication is for maintaining their health (Not at all important/ Moderately important/ Extremely important). Respondents are also asked to rate how complex their medication regime is on a scale from 1 "Extremely simple" to 5 "Extremely complex", whether they were prescribed more than one type of medication to treat their condition (Yes/ No), and whether they must remember to take on average more than one dose per day (Yes/ No).

Finally, we ask participants the frequency with which they travel away from home (Every week/ Every month/ Once in 3 months/ Once in 6 months/ Once in a year/ Never) and the extent to which travel increases the complexity of being adherent on a scale of 1 "Strongly disagree" to 5 "Strongly agree".

Social desirability bias: We measure the tendency to respond in a socially desirable manner via the 40-item Balanced Inventory of Desirable Responding scale (BIDR) (Hart et al., 2015; Paulhus, 1988; Stöber et al., 2002). The BIDR comprises two subscales: impression management (IM) and social-deceptive enhancement (SDE). IM refers to the conscious dissimulation of responses to create a socially desirable image (leading responses to be biased towards pleasing others). In contrast, SDE is a non-conscious inclination to perceive oneself favourably (leading to honest but overly positive responding). In the literature, IM scores are used to control for socially desirable responding in self-report data (Larson, 2019). We calculate IM and SDE scores as the mean response to the relevant sub-scale items. Scores range between 1 and 7 and are positively correlated ($\rho = 0.4089, p < 0.001$).[13]

[13]In Experiment 3, the short version of the BIDR scale (BIDR-16) was used to reduce the length of the experiment. The BIDR-16 comprises 16 items selected from the BIDR-40 and retains the 40-item scales 2-factor structure, reliability and validity (Hart et al., 2015). Thus, we use 8 items from each sub-scale in Studies 1 and 2 to calculate IM and SDE scores to maintain consistent scoring across Experiments 1, 2, and 3 for the pooled analysis.

1.2.2 Procedures

We ran the experiment in July 2018. The experiment was programmed in Qualtrics, and participants were recruited on Prolific. All participants were paid a £4 participation fee (£20/hour). The average experiment completion time was 12 minutes.

We required respondents to meet the following Prolific criteria: (i) to live in the UK and (ii) to be diagnosed with a chronic illness. Upon entering the experiment, respondents completed an additional screening question to ensure all participants were currently taking long-term medication prescribed for a chronic illness. Respondents who reported not to be taking a prescribed course of long-term medication completed a hypothetical version of the experiment and were excluded from the main analysis.[14]

1.2.3 Results

Summary statistics

Figure 1.2 summarises the number of subjects recruited, randomised, and included in the analysis.

In total, 601 subjects were recruited for the experiment. After excluding subjects currently not prescribed long-term medication for a chronic condition ($n = 100$) and those that did not complete all stages of the experiment ($n = 1$), there are 500 subjects included in the main analysis.

Table 1.2 summarises outcome measures and subject characteristics. Tests for joint orthogonality confirm respondent characteristics balanced across message conditions ($p = 0.1173$, see Appendix A.2.1). The average age of participants is 42 ($SD = 13.62$). Although the sample is not intended to be representative of the UK population, it is notable that, consistent with research on who participates in online experiments, female-identifying and highly educated individuals are over-represented (Curtin et al., 2000; Mulder & de Bruijne, 2019): 72% of subjects identify as female, and 58% report

[14] For the exact wording of the screening question in the experiment, see Appendix A.1.2, and for more details on the hypothetical version of the experiment, see Appendix A.1.3.

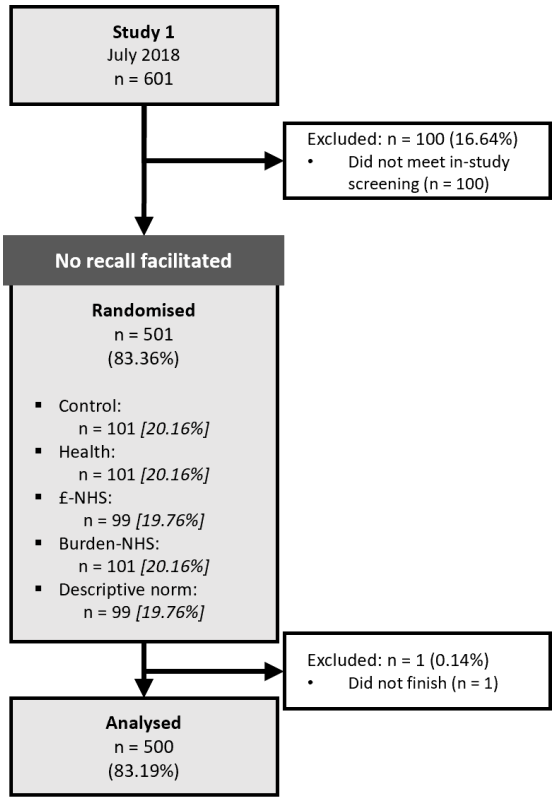


Figure 1.2: Study 1 trial profile

holding at least a university degree or equivalent. Average socio-economic scores are below the scale's mid-point ($M = 4.92, SD = 1.6$), and modal household income is in the bracket of £25,000 to £49,000.

Participants report medicines adherence for a wide range of illnesses, with the most represented in the sample being Asthma/COPD (20%), Diabetes (19%), and Hypertension (15%).

Regarding regime characteristics, 70% of respondents report adherence to a regime they had prescribed over 3 years ago, 61% are prescribed more than one medication to control the same illness, and 64% report taking medication more than once a day on average. The average importance of strict adherence to maintaining health is moderate to extremely important (M = 2.36, SD = 0.66). Most (92%) respondents report accessing health care through the NHS only.

Regarding self-reported adherence, the continuous score is highly skewed, ranging between 11 to 25, with a mean of 22.24 ($SD = 3.02$). According to the dichotomous scoring rule, 23% of subjects reported being adherent (<25 coded as non-adherent). Consistent with existing evidence, subjects report, on average, a significantly greater degree of unintentional non-adherence ($M = 4.21, SD = 0.78$) relative to intentional non-adherence ($M = 4.51, SD = 0.63$) (two-sided independent t-test, $p < 0.001$) (Molloy et al., 2014).

Nudge bias in self-reports

Overall, self-reports of adherence in the sample were negatively skewed (see Appendix A.2.2). Therefore, non-parametric tests are used to investigate whether nudges biased self-reports.

Table 1.3 summarises NTF-self-report scores across message conditions. Let us first consider the continuous score. Consistent with Hypothesis 1, we observe differences in the distributions of NTF-self-reports between the nudge and control group, which provides the first indication of nudge bias (Two-sample Kolmogorov-Smirnov test: $p = 0.055$). Looking at each nudge message separately, this result holds for the £-NHS,

Study 1

	M	SD		M	SD
Outcome measures			**Chronic illness**		
NTF continuous	22.24	3.02	Hypertension	0.15	0.35
NTF dichotomous	0.23	0.42	Cardiovascular disease	0.02	0.15
NTF unintentional	4.21	0.78	Stroke	0.02	0.15
NTF intentional	4.51	0.63	Asthma/COPD	0.20	0.40
			Diabetes	0.19	0.39
Demographics			Arthritis	0.02	0.13
Age	42.67	13.62	Cancer	0.00	0.06
Gender			Endometriosis	0.00	0.00
Female	0.72	0.45	Fibromyalgia	0.00	0.05
Male	0.27	0.44	IBD	0.01	0.09
Non-binary	0.01	0.09	Mood disorders	0.04	0.20
Prefer not to say	0.00	0.00	Multiple Sclerosis	0.01	0.09
Education			Hyperthyroidism	0.05	0.22
No qual	0.01	0.12	Autoimmune condition	0.03	0.17
GCSEs or eq.	0.18	0.38	Other	0.25	0.43
A-levels or eq.	0.23	0.42	Prefer not to say	0.01	0.12
University degree or eq.	0.43	0.50			
Post-grad degree	0.15	0.36	**Prescribed medication**		
Other	0.00	0.06	Time taking medication		
Prefer not to say	0.00	0.00	0-6 months	0.04	0.20
			7-11 months	0.04	0.20
Subjective socio-economic	4.92	1.53	1-3 years	0.22	0.42
Household income			3+ years	0.70	0.46
up to £9,999	0.10	0.30	Prefer not to say	0.00	0.00
£10000 to £24999	0.34	0.48			
£25000 to £49999	0.36	0.48	>1 medication prescribed	0.61	0.49
£50000 to £99999	0.18	0.38	>1 dose per day	0.64	0.48
£100000 +	0.02	0.13	Subjective complexity	1.93	1.06
			Importance of adherence	2.36	0.66
Social desirability bias			NHS only healthcare	0.92	0.28
SDE	3.83	0.94	Travel more than once per 3m	0.38	0.49
IM	4.09	1.07	Travel complexity	2.79	1.36

Note: n = 500.

Table 1.2: Study 1 summary of outcome and sample characteristics

burden-NHS, and descriptive norm messages. On average, subjects were shown one of these messages report greater adherence to their prescribed regimes. We observe the strongest nudge bias in the £-NHS message condition. By contrast, the health message has no detectable effect on NTF-self-reports.

Study 1: NTF-self-reports

			Continuous		Dichotomous		Unintentional		Intentional	
Message (n)	min	max	M (SD)	p	%	p	M (SD)	p	M (SD)	p
Control (101)	11	25	21.65 (3.19)		20%		4.09 (0.78)		4.39 (0.63)	
Health (100)	10	25	21.82 (3.42)	0.405	19%	0.886	4.09 (0.82)	0.848	4.43 (0.74)	0.168
£-NHS (99)	14	25	22.85 (2.30)	0.005	27%	0.213	4.35 (0.59)	0.041	4.62 (0.48)	0.004
Burden-NHS (101)	10	25	22.31 (3.43)	0.032	26%	0.314	4.24 (0.84)	0.143	4.52 (0.71)	0.018
Norm (99)	14	25	22.60 (2.44)	0.041	25%	0.356	4.28 (0.71)	0.115	4.58 (0.54)	0.014

Note: p-values displayed for continuous, unintentional, and intentional scores taken from non-parametric Wilcoxon ranksum tests between message treatment and control group. P-values next to dichotomous scores taken from $\chi 2$ tests. Two sample Kolmogorov Smirnov test between continuous NTF self-reports in nudge conditions and the control group: $d = 0.1493$, exact $p = 0.055$.

Table 1.3: Study 1 average continuous and dichotomous adherence scores

While we find support for a nudge bias in continuous scores, this does not hold for dichotomous scores. We observe the most nudge bias in the £-NHS message condition. Finally, when looking at different barriers to adherence, we find significant differences between the nudge and control group for intentional but not unintentional scores. Only in the £-NHS message condition does the nudge bias both unintentional and intentional adherence scores.

Next, we examine the robustness of our findings in a regression framework. The results are summarised in Table 1.4. Controlling for subject demographic, medication regime and social desirability scores, the nudge bias disappears for most messages. Only for the £-NHS message do we find a marginally significant effect for the continuous (Model 1: $\beta = 0.722, p = 0.050$) and intentional non-adherence score (Model 4: $\beta = 0.138, p = 0.073$).

While the £-NHS nudge message moved self-reports of adherence towards perfect adherence (the socially desirable response) and, were actual behaviour changing, the

Study 1: NTF-self-reports

	(1) Continuous β (se)	(2) Dichotomous dydx (se)	(3) Unintentional β (se)	(4) Intentional β (se)
Health	0.402	0.023	0.021	0.095
	(0.366)	(0.053)	(0.104)	(0.076)
£-NHS	0.722*	0.037	0.169	0.138*
	(0.368)	(0.053)	(0.104)	(0.077)
Burden-NHS	0.452	0.066	0.128	0.081
	(0.364)	(0.054)	(0.103)	(0.076)
Norm	0.485	0.019	0.142	0.086
	(0.369)	(0.051)	(0.105)	(0.077)
Constant	16.240***		2.742***	3.458***
	(1.142)		(0.338)	(0.239)
Controls	YES	YES	YES	YES
R^2 / Pseudo R^2	0.302	0.183	0.150	0.306
F/χ2	12.281	76.67	4.988	12.499
N	500	500	500	500

Note: Significance indicated: * p < 0.1, ** p < 0.05, *** p < 0.01. Models 1, 3, and 4 are OLS. Model 2 coefficients are average marginal effects estimated from Probit model run with robust standard errors. All models included controls for demographics (age, education), medication regime (recency, subjective complexity, importance), and social desirability bias scores (SDE, IM). Baseline group for message condition = control message.

Table 1.4: Study 1 regression analysis on NTF-self-reports

intended nudge effect, the main £-NHS effect size is small (Cohen's d = 0.42), and coefficients are only statistically significant at the 10% level.

Overall, our findings from Study 1 provide evidence of a nudge bias. While we find significant differences in self-reports between the control and the nudge group, most differences disappear after including subject-level controls. However, the results do reveal heterogeneity in nudge bias across messages, motivating further investigation to identify a more nudge-robust version of the self-report.

1.3 Study 2

1.3.1 Design & Procedures

Study 2 provides a tougher test for nudge bias relative to Study 1 via the implementation of the time-framed (TF) self-report. The experiment in Study 2 follows the same design as Study 1.[15] The difference between the experiments relates to the implementation of the self-report instrument. In Study 2, we explicitly instruct subjects to think only of the last 2-weeks when completing the MARS-5 questionnaire.[16] As previously discussed, we expect TF-self-reports to show a lower level of nudge bias, as evidence shows facilitating respondent recall reduces the effect of recall bias on self-reports (Clarke et al., 2008; Kjellsson et al., 2014; Lu et al., 2008; Stirratt et al., 2015; Stull et al., 2009). Given the mild evidence for nudge bias in Study 1, we do not expect to observe the bias in Study 2.

H2: TF-self-reports of adherence are nudge-invariant.

Consistent with this hypothesis, we expect no difference in self-reports of adherence between message conditions.

The experiment ran in July 2019. Again, the experiment was programmed in Qualtrics, and participants were recruited on Prolific. All participants received £4 for participating (£20/hour), and the average experiment completion time was 12 minutes.

As in Study 1, we require respondents to meet the following Prolific criteria: (i) to live in the UK and (ii) to be diagnosed with a chronic illness. To ensure sample independence between experiments, subjects who had participated in Study 1 could not participate in this experiment. Upon entering the experiment, we administered an additional screening question to ensure all participants were currently taking long-term medication prescribed for a chronic illness. Respondents who reported not taking a prescribed course of long-term medication were removed from the experiment and paid

[15] We also included additional questions in Study 2 to assess the intentionality of each type of non-adherence (defined by MARS-5 item) reported by subjects. For the wording of these questions, see Appendix A.1.4.

[16] For exact wording of the time-framed MARS-5, see Appendix A2.1

£0.40 for their interest in the study.

1.3.2 Results

Summary statistics

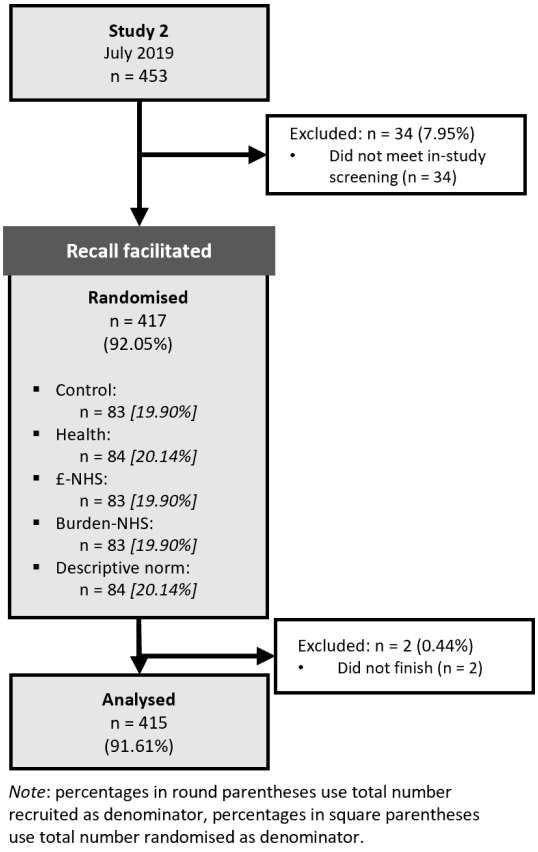


Figure 1.3: Study 2 trial profile

As shown in Figure 1.3, 453 subjects were recruited initially. However, after excluding subjects not currently prescribed long-term medication for a chronic condition (n = 34), 415 subjects are included in the main analysis.

Table 1.5 summarises the experiment's outcome measures and subject characteristics. Tests for joint orthogonality confirm respondent characteristics balanced across message conditions ($p = 0.5434$, see Appendix A.2.1). Moreover, subject characterist-

ics between the independent samples recruited for Study 1 and 2 did not differ (for comparisons across experimental samples, see Appendix A.2.3).

When looking at self-reports, we again find the distribution of adherence scores is highly skewed. Continuous scores range from 8 to 25, with a mean of 22.42 ($SD = 2.96$). According to the dichotomous scoring rule, 24% of subjects report being adherent (<25 coded as non-adherent). Also, as in Study 1, unintentional non-adherence is greater on average ($M = 4.25, SD = 0.75$) than intentional non-adherence ($M = 4.54, SD = 0.62$) ($t(414) = 9.66, p < 0.001$).

Study 2

	M	SD		M	SD
Outcome measures			**Chronic illness**		
TF continuous	22.42	2.96	Hypertension	0.10	0.30
TF dichotomous	0.24	0.43	Cardiovascular disease	0.03	0.17
TF unintentional	4.25	0.75	Stroke	0.04	0.20
TF intentional	4.54	0.62	Asthma/COPD	0.14	0.35
			Diabetes	0.23	0.42
Demographics			Arthritis	0.03	0.17
Age	42.74	13.21	Cancer	0.00	0.05
Gender			Endometriosis	0.00	0.00
Female	0.66	0.48	Fibromyalgia	0.00	0.05
Male	0.34	0.47	IBD	0.01	0.09
Non-binary	0.01	0.07	Mood disorders	0.03	0.18
Prefer not to say	0.00	0.00	Multiple Sclerosis	0.00	0.05
Education			Hyperthyroidism	0.07	0.26
No qual	0.01	0.12	Autoimmune condition	0.02	0.15
GCSEs or eq.	0.16	0.36	Other	0.28	0.45
A-levels or eq.	0.27	0.44	Prefer not to say	0.01	0.10
University degree or eq.	0.42	0.49			
Post-grad degree	0.14	0.35	**Prescribed medication**		
Other	0.01	0.07	Time taking medication		
Prefer not to say	0.00	0.00	0-6 months	0.07	0.25
			7-11 months	0.06	0.24
Subjective socio-economic	5.24	1.74	1-3 years	0.23	0.42
Household income			3+ years	0.65	0.48
up to £9,999	0.09	0.28	Prefer not to say	0.00	0.00
£10000 to £24999	0.33	0.47			
£25000 to £49999	0.37	0.48	>1 medication prescribed	0.66	0.47
£50000 to £99999	0.18	0.38	>1 dose per day	0.66	0.47
£100000 +	0.04	0.19	Subjective complexity	2.12	1.19
			Importance of adherence	2.40	0.64
Social desirability bias			NHS only healthcare	0.90	0.30
SDE	3.78	0.90	Travel more than once per 3m	0.41	0.49
IM	4.11	1.02	Travel complexity	2.93	1.38

Note: n = 415.

Table 1.5: Study 2 summary of outcome and sample characteristics

Nudge bias in self-reports

Table 1.6 summarises average adherence scores across message conditions in Experiment 2. Consistent with Hypothesis 2, we find no evidence of nudge bias in TF-self-reports.

There is no difference in the distribution of TF-self-reports between nudged and control group subjects (Two-sample Kolmogorov Smirnov test: $p = 0.895$).

Study 2: TF-self-reports

			Continuous		Dichotomous		Unintentional		Intentional	
Message (n)	min	max	M (SD)	p	%	p	M (SD)	p	M (SD)	p
Control (82)	8	25	22.24 (3.59)		27%		4.21 (0.83)		4.51 (0.73)	
Health (84)	12	25	22.26 (2.82)	0.376	18%	0.165	4.20 (0.77)	0.861	4.51 (0.60)	0.523
£-NHS (84)	10	25	22.59 (2.95)	0.704	33%	0.423	4.33 (0.83)	0.210	4.57 (0.60)	0.811
Burden-NHS (83)	8	25	22.71 (3.02)	0.598	23%	0.558	4.31 (0.62)	0.628	4.60 (0.69)	0.299
Norm (83)	15	25	22.34 (2.44)	0.236	20%	0.337	4.20 (0.66)	0.613	4.53 (0.48)	0.291

Note: p-values displayed for continuous, unintentional, and intentional scores taken from non-parametric Wilcoxon ranksum tests between message treatment and control group. P-values next to dichotomous scores taken from $\chi 2$ tests. Two sample Kolmogorov Smirnov test between continuous NTF self-reports in nudge conditions and the control group: $d = 0.0709$, exact $p = 0.867$.

Table 1.6: Study 2 average continuous and dichotomous adherence scores by message

The regression analysis in Table 1.7 confirms that TF-self-reports are nudge robust, even after including subject-level controls.

We, therefore, would conclude that TF-self-reports appear to be a more nudge-robust self-report measure. However, before making a final recommendation on the use of self-reports to evaluate medication-taking behaviour after nudging adherence, we aim to replicate the findings of Studies 1 and 2.

Study 2: TF-self-reports

	(1) Continuous β (se)	(2) Dichotomous dydx (se)	(3) Unintentional β (se)	(4) Intentional β (se)
Health	0.083	−0.071	0.008	0.019
	(0.441)	(0.062)	(0.113)	(0.094)
£-NHS	0.229	0.053	0.120	0.027
	(0.441)	(0.064)	(0.113)	(0.094)
Burden-NHS	0.476	−0.027	0.138	0.084
	(0.438)	(0.060)	(0.112)	(0.093)
Norm	0.146	−0.042	0.015	0.033
	(0.439)	(0.063)	(0.112)	(0.093)
Constant	16.362***		3.032***	3.333***
	(1.112)		(0.284)	(0.237)
Controls	YES	YES	YES	YES
R^2 / Pseudo R^2	0.165	0.176	0.133	0.145
F/χ2	4.310	61.93	3.365	3.724
N	413	413	413	413

Note: Significance indicated: * p < 0.1, ** p < 0.05, *** p < 0.01. Models 1, 3, and 4 are OLS. Model 2 coefficients are average marginal effects estimated from Probit model run with robust standard errors. All models included controls for demographics (age, education), medication regime (recency, subjective complexity, importance), and social desirability bias scores (SDE, IM). Baseline group message condition = control message.

Table 1.7: Study 2 regression analysis on TF-self-reports

1.4 Study 3

In Studies 1 and 2 we find for certain nudges (£-NHS) nudge bias can influence NTF-self-reports and that time-framing produces a nudge robust version of the self-report. Now we aim to replicate these results and improve our statistical power to increase confidence in these findings.

1.4.1 Design & Procedures

As in the previous experiments, subjects see a message before we elicit their medication adherence using the MARS-5 self-report tool. As the aim of Study 3 is to replicate Studies 1 and 2, we use a 2 (Measurement) x 5 (Message) between-subjects design. Within which subjects are randomly allocated to complete either the NTF- or the TF-self-report. The message shown to subjects during the experiment is determined at random.

As we expect to replicate the results of previous experiments, we test the same set of hypotheses:

H1: NTF-self-reports of adherence are nudge biased.

H2: TF-self-reports of adherence are nudge robust.

The experiment ran in June and July 2021. Due to the global COVID-19 pandemic, we did not ask subjects questions about travel. Instead, subjects were asked whether their ability and motivation to take medication as prescribed had been affected by the pandemic and whether they were reporting adherence to immunosuppressive medication.

Experiment 3 was programmed in Qualtrics, and participants were recruited on Prolific. Subjects earned £4 for their participation (£20/hour) in the main experiment. The average completion time was 12 minutes.

As in the previous experiments, respondents must meet the following Prolific criteria: (i) to live in the UK, (ii) to be diagnosed with a chronic illness, and additionally (iii) to not have participated in Experiment 1 or 2. The target sample size for Ex-

periment 3 was larger than the previous studies. We used a pre-screening stage to identify an eligible pool of subjects before the main experiment. In the pre-screening stage (23rd April- 2nd June 2021), subjects were asked if they currently take long-term medication for a chronic illness. Only subjects who responded "yes" were invited to the main experiment.[17]

As Study 3 was designed to replicate our previous results, extra attention was paid to sample design and power. Based on data from Study 1 and Study 2, we require an estimated minimum sample size of 91 subjects per message group to replicate the effect of the £-NHS message on NTF-self-reports (Cohen's d = 0.43) at the 5% level effect, using a non-parametric test (80% power). Then, to accommodate control group comparisons with multiple treatment groups, we assign a larger proportion of the sample to the control group (relative to the message treatment sub-samples). Participants in Experiment 3 were randomly allocated to the control group with 60% probability and to each message treatment group with 10% probability. In Experiment 3, the self-report administered (NTF or TF) was determined at random with equal probability.

1.4.2 Results

Summary statistics

Figure 1.4 provides an overview of the number of participants recruited, screened, and randomised into treatment in Study 3.

We recruited 3239 subjects at pre-screening. After excluding subjects not currently prescribed long-term medication for a chronic condition (n = 689), we invited 2518 subjects to the main experiment. There was an attrition rate of 17% between pre-screening and the main experiment. In the end, 2061 subjects were randomised into treatment and included in the main analysis: 1030 completed the NTF-self-report (control: n = 624, health: n = 95, £-NHS: n = 111, burden-NHS: n = 99, descriptive norm: n = 101) and 1031 the TF-self-report (control: n = 612, health: n = 113, £-NHS:

[17] All subjects recruited for the pre-screening stage were paid a £0.18 participation fee. See Appendix A.1.5, for more detail about the pre-screening survey

n = 96, burden-NHS: n = 105, descriptive norm: n = 105).

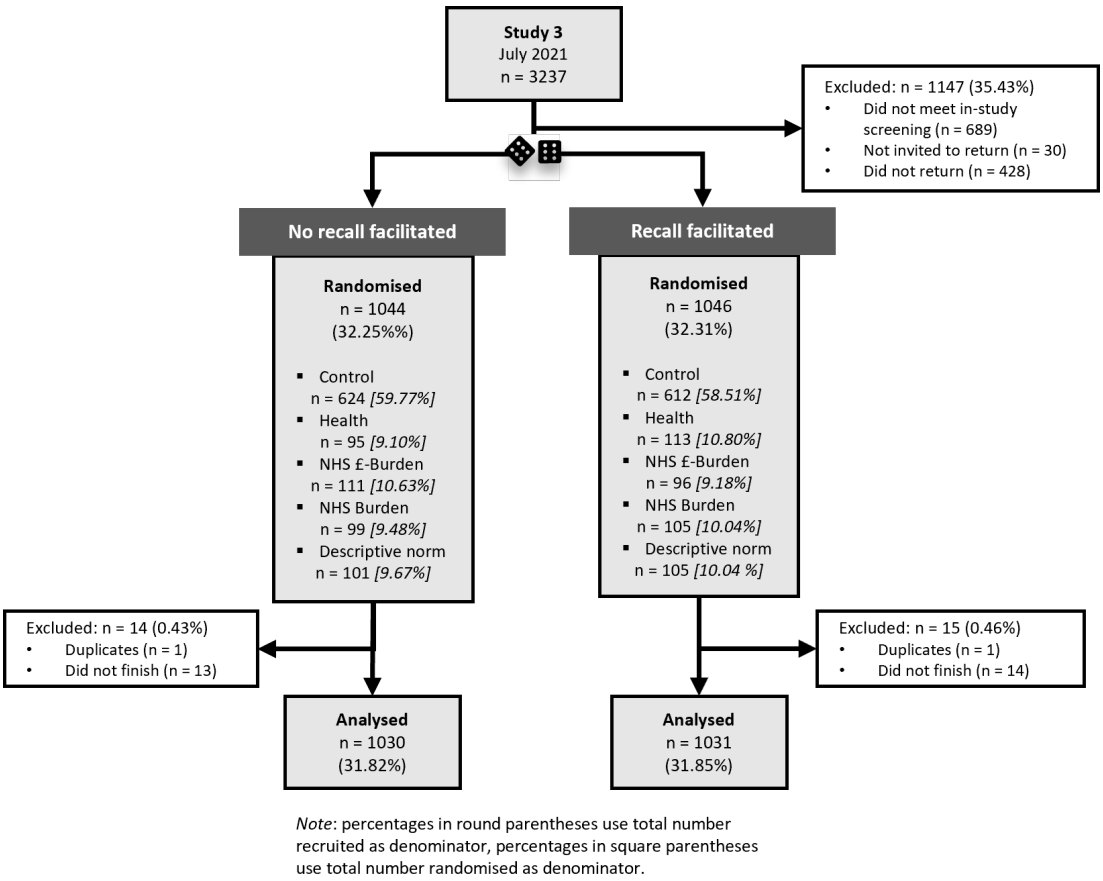


Figure 1.4: Study 3 trial profile

Table 1.8 summarises outcome measures and subject characteristics in Study 3. Tests for joint orthogonality confirm respondent characteristics balanced across message conditions (p = 0.6717) and measurement conditions (p = 0.0798) (see Appendix A.2.1). Compared to the previous experiments, there were some differences in the average subject characteristics recruited in Study 3 (for a full discussion of comparisons across experimental samples see Appendix A.2.3). However, we control for these characteristics in our regression analysis.

In addition, we asked respondents in Study 3 how the pandemic had affected their ability and motivation to take medication exactly as prescribed. On average,

respondents report a small reduction in ability ($M = -0.08, SD = 0.36$; t(2056) = -10.49, $p = 0.000$), but no change in motivation to be adherent to their medication ($M = 0.01, SD = 0.49$; t(2057) = 1.12, $p = 0.260$). Finally, 19% of respondents reported adherence to immunosuppressive medication.

Study 3

		M	SD		M	SD
Outcome measures				**Chronic illness**		
Continuous	NTF	22.54	2.78	Hypertension	0.08	0.27
	TF	22.11	2.77	Cardiovascular disease	0.04	0.21
Dichotomous	NTF	0.22	0.42	Stroke	0.01	0.11
	TF	0.43	0.49	Asthma/COPD	0.10	0.30
Unintentional	NTF	4.20	0.73	Diabetes	0.21	0.41
	TF	4.44	0.74	Arthritis	0.07	0.26
Intentional	NTF	4.59	0.58	Cancer	0.03	0.17
	TF	4.67	0.56	Endometriosis	0.02	0.14
				Fibromyalgia	0.05	0.21
Demographics				IBD	0.07	0.26
Age		45.69	14.38	Mood disorders	0.07	0.25
Gender				Multiple Sclerosis	0.02	0.15
Female		0.70	0.46	Hyperthyroidism	0.02	0.13
Male		0.29	0.45	Autoimmune condition	0.01	0.09
Non-binary		0.01	0.10	Other	0.19	0.40
Prefer not to say		0.00	0.05	Prefer not to say	0.01	0.07
Education						
No qual		0.01	0.11	**Prescribed medication**		
GCSEs or eq.		0.14	0.35	Time taking medication		
A-levels or eq.		0.24	0.43	0-6 months	0.04	0.19
University degree or eq.		0.45	0.50	7-11 months	0.03	0.18
Post-grad degree		0.16	0.36	1-3 years	0.19	0.39
Other		0.00	0.06	3+ years	0.73	0.44
Prefer not to say		0.00	0.04	Prefer not to say	0.00	0.05
Subjective socio-economic		5.28	1.58	>1 medication prescribed	0.72	0.45
Household income				>1 dose per day	0.66	0.47
up to £9,999		0.06	0.23	Subjective complexity	2.18	1.18
£10000 to £24999		0.25	0.44	Importance of adherence	2.40	0.62
£25000 to £49999		0.40	0.49	NHS only healthcare	0.89	0.32
£50000 to £99999		0.24	0.43			
£100000 +		0.05	0.21			
				COVID-19	-0.08	0.36
Social desirability bias				Immunosuppressive medication	-0.08	0.36
SDE		3.84	0.96	Ability to be adherent	0.01	0.49
IM		4.49	1.00	Motivation to be adherent	0.19	0.39

Note: n = 2061.

Table 1.8: Study 3 summary of outcome and sample characteristics

Nudge bias in self-reports

Study 3

	Message (n)			Cont. M (SD)	p	Dich. %	p	Unint. M (SD)	p	Int. M (SD)	p
NTF	Control (624)	7	25	22.43 (2.85)		20%		4.14 (0.74)		4.57 (0.59)	
	Health (95)	9	25	22.52 (3.05)	0.486	22%	0.694	4.23 (0.71)	0.221	4.57 (0.64)	0.735
	£-NHS (111)	14	25	23.03 (2.19)	0.045	29%	0.046	4.39 (0.65)	0.000	4.66 (0.47)	0.232
	Burden-NHS (99)	13	25	22.78 (2.45)	0.283	24%	0.376	4.30 (0.69)	0.021	4.62 (0.52)	0.714
	Norm (101)	12	25	22.51 (2.87)	0.456	25%	0.314	4.26 (0.76)	0.059	4.56 (0.59)	0.969
TF	Control (612)	6	25	22.94 (2.91)		38%		4.39 (0.77)		4.64 (0.59)	
	Health (113)	17	25	23.67 (1.79)	0.017	51%	0.010	4.58 (0.58)	0.026	4.77 (0.36)	0.057
	£-NHS (96)	13	25	23.53 (2.08)	0.081	47%	0.114	4.56 (0.66)	0.025	4.74 (0.41)	0.199
	Burden-NHS (105)	11	25	22.46 (2.64)	0.009	54%	0.002	4.54 (0.67)	0.044	4.73 (0.54)	0.032
	Norm (105)	9	25	22.79 (3.36)	0.956	44%	0.294	4.37 (0.82)	0.975	4.60 (0.67)	0.848

Note: p-values displayed for continuous, unintentional, and intentional scores taken from non-parametric Wilcoxon ranksum tests between message treatment and control group. P-values next to dichotomous scores taken from Chi2 tests. Two sample Kolmogorov-Smirnov test between continuous NTF self-reports in nudge conditions and the control group: d = 0.0579, exact p = 0.381; TF self-reports in nudge conditions and the control group: d = 0.1077, exact p = 0.006.

Table 1.9: Study 3 average continuous and dichotomous adherence scores by message

As shown in Table 1.9, we see a similar pattern of nudge bias for the £-NHS message in NTF-self-reports as in Study 1, with the exception that in this experiment, there is a nudge bias in unintentional adherence scores, but not intentional adherence scores. Interestingly, we find that in all other message conditions, NTF-self-reports are nudge-robust. Unsurprisingly, the absence of a nudge bias in most message conditions leads to no difference in the overall distribution of NTF-self-reports between nudged and control

subjects (Two-sample Kolmogorov-Smirnov test: p = 0.381).

Turning to TF-self-reports, we fail to replicate the findings from Study 2. Contrary to H2, we find most nudge messages (health, £-NHS, burden-NHS), produce a nudge bias in TF-self-reports (Two-sample Kolmogorov-Smirnov test: p = 0.006). Only in the case of the descriptive norm treatment are TF-self-reports nudge robust.

Regression analyses confirm the presence of a nudge bias for a number of messages and outcome measures (see Table 1.10). The health message increases TF-self-reports for continuous (Model 5, $\beta = 0.556, p = 0.030$), dichotomous (Model 6, $dydx = 0.110, p = 0.018$), and unintentional scores (Model 7, $\beta = 0.145, p = 0.037$). Moreover, subjects report higher adherence after exposure to the £-NHS message in continuous (Model 5, $\beta = 0.605, p = 0.029$), and unintentional scores (Model 7, $\beta = 0.188, p = 0.013$). Finally, exposure to the burden-NHS message significantly increases the probability subjects report to be perfectly adherent, all else held constant (Model 6, $dydx = 0.148, p = 0.003$).

	NTF-self-reports				TF-self-reports			
	(1) Cont. β (se)	(2) Dich. dydx (se)	(3) Unint. β (se)	(4) Int. β (se)	(5) Cont. β (se)	(6) Dich. dydx (se)	(7) Unint. β (se)	(8) Int. β (se)
Health	0.132	0.034	0.110	0.005	**0.556****	**0.108****	**0.145****	0.103*
	(0.274)	(0.045)	(0.076)	(0.057)	(0.255)	(0.047)	(0.069)	(0.052)
£-NHS	0.463*	0.0596	**0.217*****	0.062	**0.605****	0.075	**0.188****	0.104*
	(0.259)	(0.041)	(0.072)	(0.054)	(0.277)	(0.050)	(0.075)	(0.057)
Burden-NHS	0.210	0.0277	0.143*	0.017	0.360	**0.148*****	0.120*	0.060
	(0.270)	(0.041)	(0.075)	(0.057)	(0.266)	(0.050)	(0.072)	(0.055)
Norm	0.134	0.062	0.134*	0.000	−0.423	0.018	−0.077	−0.087
	(0.269)	(0.045)	(0.075)	(0.056)	(0.264)	(0.047)	(0.072)	(0.054)
Constant	17.473***		3.028***	3.611***	17.603***		2.946***	3.664***
	(0.698)		(0.194)	(0.146)	(0.615)		(0.167)	(0.126)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
R^2 / Psuedo R^2	0.176	0.129	0.110	0.167	0.158	0.1372	0.142	0.140
F/χ2	11.751	141.79	6.839	11.082	10.379	178.54	9.186	8.997
N	1011	1011	1011	1011	1016	1016	1016	1016

Note: Significance indicated: * p < 0.1, ** p < 0.05, *** p < 0.01. Models 1, 3, 4, 5, 7, and 8 are OLS. Models 2 and 6 coefficients are average marginal effects estimated from Probit model run with robust standard errors. All models included controls for demographics (age, education), medication regime (recency, subjective complexity, importance), and social desirability bias scores (SDE, IM). Baseline group for message condition = control message.

Table 1.10: Study 3 regression analysis

1.5 Pooled analysis

To increase statistical power, we pool the independent experimental samples across studies, and conduct regression analyses on the pooled sample (see Table 1.11).[18]

The results provide evidence of nudge bias in both NTF and TF-self-reports. However, in most cases, time-framing the self-report reduces the average size of detected nudge bias. For example, the £-NHS message increases NTF-continuous scores (Model 1,$\beta = 0.540, p = 0.008$) by more than TF-continuous scores (Model 5, $\beta = 0.459, p = 0.041$). Only in the case of dichotomous scores did we detect nudge bias in TF-self-reports, and not NTF-self-reports. While in both TF- and NTF-self-reports, we find no evidence of nudge bias on intentional scores significant at a 5% level, there is an effect on unintentional scores.

Having analysed data from 2,931 subjects, we find evidence that nudge messages can systematically influence self-reports of adherence. While not all messages and self-report scores are equally biased, policymakers should be cautious when evaluating intervention effectiveness based only on self-report data.

[18] For a discussion about the potential impact of the COVID-19 pandemic on our results see Appendix A.2.4.

	NTF-self-reports				TF-self-reports			
	(1) Cont. β (se)	(2) Dich. dydx (se)	(3) Unint. β (se)	(4) Int. β (se)	(5) Cont. β (se)	(6) Dich. dydx (se)	(7) Unint. β (se)	(8) Int. β (se)
Health	0.235	0.030	0.086	0.037	0.383*	0.053	0.101*	0.070
	(0.210)	(0.033)	(0.059)	(0.044)	(0.214)	(0.036)	(0.057)	(0.045)
£-NHS	**0.540*** **	0.052	**0.211*** **	0.082*	**0.459** **	**0.090** **	**0.160*** **	0.075
	(0.203)	(0.032)	(0.057)	(0.043)	(0.224)	(0.038)	(0.060)	(0.047)
Burden-NHS	0.212	0.045	**0.134** **	0.019	0.421*	**0.091** **	**0.133** **	0.072
	(0.207)	(0.032)	(0.058)	(0.043)	(0.219)	(0.037)	(0.059)	(0.046)
Norm	0.244	0.042	**0.143** **	0.025	−0.152	0.013	−0.029	−0.031
	(0.207)	(0.033)	(0.058)	(0.043)	(0.218)	(0.037)	(0.058)	(0.045)
Year								
2018	−0.294*	0.001	−0.016	**-0.069** **				
	(0.151)	(0.023)	(0.043)	(0.032)				
2019					**-0.700*** **	**-0.190*** **	**-0.200*** **	**-0.125*** **
					(0.164)	(0.026)	(0.044)	(0.034)
Constant	17.194***		2.922***	3.568***	17.417***		3.020***	3.599***
	(0.559)		(0.157)	(0.117)	(0.538)		(0.143)	(0.112)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
R^2 / Pseudo R^2	0.206	0.136	0.111	0.203	0.163	0.159	0.146	0.142
F/χ2	20.183	195.63	9.744	19.81	14.463	257.12	12.627	12.26
N	1502	1502	1502	1502	1429	1429	1429	1429

Note: Significance indicated: * p < 0.1, ** p < 0.05, *** p < 0.01. Models 1, 3, 4, 5, 7, and 8 are OLS. Models 2 and 6 coefficients are average marginal effects estimated from Probit model run with robust standard errors. All models included controls for demographics (age, education), medication regime (recency, subjective complexity, importance), and social desirability bias scores (SDE, IM). Baseline group for message condition = control message and Year = 2021.

Table 1.11: Pooled regression analysis

1.6 Discussion & Conclusion

In this chapter, we present a novel experimental design to investigate the effect of nudge messages on self-reports of medicines adherence independent from behaviour. In our set-up, we ask participants about their adherence immediately after seeing a nudge message, such that there is no possibility for the nudge to affect behaviour. By this logic, any difference between the nudge and control group is indicative of a direct influence of the nudge on self-reports (nudge bias). Our results reveal some nudge messages give rise to replicable biases in self-reports of medicines adherence.

Not all nudges move self-reports to the same degree. However, we find evidence that at least one nudge message (£-NHS message) consistently affects self-reports; establishing the potential presence of nudge bias in self-reports of medicines adherence. The £-NHS message produces nudge bias in both continuous and unintentional scores, regardless of whether the self-report is time-framed, or not. We replicate this finding across studies.

Analysis of the pooled sample reveals that while TF-self-reports are not nudge-robust, they do have the potential to reduce nudge bias. Therefore, nudge bias may be the result, at least partly, of established response biases (e.g., recall bias). Moreover, from a policy perspective, this implies a slight preference for TF-self-reports over NTF-self-reports.

The results suggest there is a reason for caution in the reliance on self-reports of adherence to evaluate the behavioural effects of message nudge interventions. As a result, we should interpret previous evidence regarding adherence nudge message effectiveness based only on self-report data with caution. Although notably, the results of this chapter lend credibility to the findings of Jachimowicz et al. (2019) - a nudge message behavioural evaluation using only patient self-reports - as we find self-reports are mostly robust to the nudge adapted from their winning intervention (health).

Investigation into mechanisms driving nudge bias in self-reports may offer strategies in implementation and analysis to reduce and control for the presence of nudge bias in

self-report data. Also, further work is required to understand why, relative to intentional behaviour, self-reports of unintentional behaviour are more sensitive to nudge bias.

Chapter 2

Investigating mechanisms of nudge bias

2.1 Motivation

In Chapter 1, we find nudge messages systematically influence self-reports despite behaviour not having changed. This so-called nudge bias demonstrates a weakness of self-reports as measures to evaluate the effects of nudge messages on medication-taking behaviour.[1] Within our experimental set-up, we observed heterogeneity in nudge bias across nudges; some messages produce a replicable bias (£-NHS), whereas others do not.[2] Furthermore, we find facilitating respondent recall, a technique often applied to reduce the presence of response biases before they can occur (Clarke et al., 2008; Kjellsson et al., 2014; Lu et al., 2008; Stirratt et al., 2015; Stull et al., 2009), reduces but does not eliminate nudge bias.

Building on these results, we investigate mechanisms driving nudge bias in self-reports of medicines adherence in this chapter. Clarifying the psychological processes underlying self-report distortions is not only relevant to the response bias literature in general (Furnham & Henderson, 1983; Hill et al., 2019; Latkin et al., 2016; van de Mortel, 2008), but from an applied public policy perspective, our investigation may inform refinements to self-reports as an evaluative tool. This applies to both implementing self-reports when evaluating the effectiveness of nudge messages on behaviour (i.e., to minimise bias before it occurs) and measuring and adjusting for overstated adherence in analysis.

In this chapter, we focus on two potential channels of nudge bias: (i) social desirability bias and (ii) message effectiveness. The two channels represent two possible interpretations of nudge bias. Under channel (i), nudge bias is orthogonal to actual medication-taking behaviour change, implying nudges can cause distortions in self-reports that mask the true relationship between nudges and behaviour. On the other

[1] As outlined in Chapter 1, we use the following terminology to describe two possible results of our experimental test: (1) self-reports are nudge-robust if they are insensitive to nudges in an environment where behaviour cannot yet change, (2) self-reports are nudge-biased if they are systematically influenced towards the socially desirable and intended behavioural outcome (greater medicines adherence) by nudges in an environment where behaviour cannot yet change.

[2] For a summary of all the messages tested, see Appendix A.1.1.

hand, under channel (ii), nudge bias may be an early signal of the incoming effect of a nudge on medication-taking behaviour.

While the experiments in Study 1 and Study 2 are designed to test for the presence of nudge bias in self-reports of medication adherence, the aim of Study 3 is to replicate these results and facilitate the investigation moderators of nudge bias in self-reports. Using results from Study 3, we find that nudge bias cannot yet be explained by social desirability bias or routes of persuasion predicting message effectiveness. As such, the psychological processes underlying nudge bias remain unexplained by the end of this chapter. However, we find some evidence that TF-self-reports are less influenced by subject characteristics unrelated to medication-taking behaviour and may provide a relatively more accurate measure of actual adherence relative to NTF-self-reports.

The structure of the chapter is as follows. First, in Section 2.1.1 and 2.1.2 we motivate the investigation of social desirability bias and routes of persuasion as mechanisms to explain nudge bias. Then, our empirical strategy and hypotheses are presented in Section 2.2. In Section 2.3, we provide more detail about the experimental design of Study 3 in terms of how mechanisms were measured. Finally, in Section 2.4, we present the results, followed by a general discussion and conclusion in Section 2.5.

2.1.1 Social desirability bias

Social desirability bias results from a "tendency to under-report socially undesirable attitudes and behaviours, and to over-report more desirable attributes" (Latkin et al., 2017). According to Paulhus' theory, there are two components of social desirability bias (Paulhus, 1984): Impression Management, the purposeful presentation of self to fit into a situation or please an audience; and Self-Deception Enhancement (SDE), the potentially unconscious motivation to maintain a positive self-image. Of the two components, social desirability response bias is more strongly attributed to Impression Management in the literature (Tourangeau & Yan, 2007). Regarding self-reports of medication adherence, social desirability bias has been shown to lead to overreporting. For example, when Nieuwkerk et al. (2010) measured the social desirability concerns

of Dutch HIV patients alongside a self-report and an objective measure (viral load) of medication adherence. The authors find the relationship between self-reports and viral load was statistically significant, but only for patients with low social desirability concerns.

This chapter aims to investigate the degree to which socially desirable reporting can explain nudge bias. Previous research has shown social desirability concerns tend to emerge, particularly when reporting issues that participants find sensitive and in situations where there are widely accepted norms (Grimm, 2010). The nudge messages in our experiments theoretically motivate greater medication adherence by increasing the salience of consequences incurred from non-adherence or social norms. However, given what we know about social desirability bias, it is possible that our nudges inadvertently increased socially desirable reporting in self-reports.

Starting with social norms, research has shown norms can prove a strong lever to influence behaviour via the individual's unwillingness to depart too far from group standards (Lewin, 2004). When the policymaker's goal is to encourage greater medication adherence, a nudge message should frame adherence as the standard behaviour. However, framing adherence as the norm may increase the social desirability concerns of diverting from this standard and lead people to under-report non-adherence (Stodel, 2015). Therefore, we might reasonably expect our social norms message (referred to in our experiment as the descriptive norm message) to lead to the overstatement of adherence due to socially desirable reporting. Regarding nudges highlighting the consequences of non-adherence, social desirability bias may be influenced by a different mechanism. Renzetti & Lee (1993) define a "sensitive" topic as one inherently linked with risks and costs, such as negative feelings of shame and embarrassment or negative consequences. The more sensitive a topic, the more we expect socially desirable responding to distort self-reports. Therefore, messages framing the consequences of non-adherent behaviour may increase socially desirable reporting by increasing the sensitivity of non-adherence.

Under the assumption that the nudge messages tested in our experiment can affect

both medication-taking behaviour and social desirability concerns, then, as behaviour cannot yet change in response to the nudge message at the point of self-report, any effect of the nudges on self-reports could be explained as increased socially desirable reporting. In other words, the nudges within our experimental context can be understood as social desirability manipulations. As such, our online experiment bears similarities to previous literature investigating the effects of social desirability on self-reports by experimentally manipulating aspects of the reporting environment (e.g., question introductions (Belli et al., 1999; Persson & Solevid, 2014) and information about the organisation conducting the survey (Lüke & Grosche, 2018)).

Social desirability bias in self-reports is normally dealt with by measuring the bias via the subject's proclivity to respond in a socially desirable way and then either proving bias is not a significant problem or controlling for the effects of the bias in analysis (Larson, 2019). As stated above, Impression Management is the component of social desirability widely considered to predict socially desirable responses.[3] Therefore, in Study 3, Impression Management is measured and investigated as a moderator of nudge bias. While we hypothesise that subjects with high Impression Management concerns will be more likely to overstate adherence in nudge message conditions, we find no evidence to suggest social desirability concerns explain nudge bias.

2.1.2 Nudge effectiveness

Suppose a nudge leads an individual to reflect on their own behaviour (their actual self) and form an intention to exert more effort to take medication correctly (ideal self). We would expect then that the individual will take actions to reduce the dissonance between their actual and ideal self and "perform" their valued identity (Brenner & DeLamater, 2014). Medication-taking behaviour is only one way a person can "perform" their identity. A respondent who values medication adherence and sees themselves as "the

[3]IM is a self-presentation strategy dependent on a respondent's second-order belief about the person asking the question. We might, therefore, expect self-reports completed online to be less affected by social desirability distortions. However, evidence shows that online environments do not reduce social desirability bias in self-reports (Antin & Shaw, 2012; Gnambs & Kaspar, 2016).

kind of person" who adheres to their medication may interpret the self-report as a low-cost opportunity to enact this identity (Brenner & DeLamater, 2016; Layder & Stryker, 1982). Therefore, overstating adherence in a self-report may be the by-product of mechanisms through which a nudge goes on to improve actual adherence behaviour. Under this model, nudge bias may be interpreted as an artefact of an effective nudge message and a signal of future behaviour change. To investigate whether nudge bias results from an effective message intervention (i.e., for which subjects are more likely to improve their adherence once the behaviour is allowed to change), we investigate determinants of message effectiveness as moderators and mediators of nudge bias.

Information processing style

Message effectiveness can depend on multiple factors (Wansink & Pope, 2015). In our experimental context, situational factors - those that relate to when and how nudges are presented to subjects - are held constant.[4] However, message characteristics (Cialdini et al., 2006; Ferraro & Price, 2013; Merritt et al., 2010; Powell et al., 2015), and dispositional factors such as demographics (Akerlof & Kennedy, 2013; Allcott, 2011; Beshears et al., 2015) and individual cognitive styles (Bao & Ho, 2015; Peer et al., 2020; Powell et al., 2019; Schöning et al., 2019) do vary and have all been shown to influence message effectiveness. So far in Chapter 1, we have shown that demographic factors cannot explain nudge bias. Therefore, in this chapter, we probe the extent to which individual cognitive styles and routes of persuasion moderate the effect of nudge messages on self-reports when medication-taking behaviour cannot yet change. Following this, we investigate the role of affective and cognitive reactions to the messages as mediators to better understand how nudges bias self-reports.

As discussed in Chapter 1, nudge messages can change behaviour by inducing positive or negative feelings (affect) or increasing the salience of relevant information to

[4] As the experiments were conducted online, we cannot control for our subjects' environment while completing the experiment. However, given our large sample size and randomisation into message treatment, we argue it is unlikely that situational factors can explain nudge bias.

motivate action (cognition). Studies on the determinants of persuasive communications (where persuasion describes how a person's attitudes and behaviour are changed by messages) acknowledge the importance of cognitive and emotional appraisals of messages (Dunlop et al., 2010; Nabi, 1999; Petty & Cacioppo, 1986; Witte, 1992). However, depending on an individual's processing style, how a person feels about a message may be more influential in decision-making than what they think about the message.

For this reason, we consider the Need for Cognition and the Need for Affect as potential moderators of message effectiveness. Need for Cognition indexes an individual's intrinsic motivation to enjoy and engage in thinking (Cacioppo & Petty, 1982). Individuals with a high Need For Cognition may be more influenced by messages with strong arguments or that present complex ideas and issues. Need For Cognition is also a predictor of the "route of persuasion" according to influential models of persuasion. Need for Affect, on the other hand, captures an individual's tendency to approach, and not avoid, intense affective experiences (Maio & Esses, 2001). Individuals who score highly on the Need For Affect scale may be more responsive to messages that elicit strong emotional responses.

Evidence suggests that messages are more persuasive when congruent with individual processing styles (Haddock et al., 2008). Therefore, in this chapter, we investigate whether Need For Cognition and Need For Affect moderate the influence of nudge messages on self-reports. In fact, in this chapter, we show subjects with a higher Need For Affect, report to be less adherent in NTF-self-reports but not TF-self-reports when exposed to the Health and Burden-NHS messages relative to the control. While this moderating effect is not in the direction of nudge bias, it does shed light on a potential determinant of heterogeneity in the effects of nudge messages directly on self-reports.

Few studies have examined the role of cognitive and emotional appraisals of persuasive messages in mediating relationships between individual processing styles and message effectiveness. Therefore, in this chapter, we also investigate whether the effect of information processing style on self-reports is mediated by affective and cognitive appraisals of the message. We find that the effect of Need For Affect on NTF-self-reports

(but not TF-self-reports) is partially mediated by cognitive appraisals of the message.

Routes of persuasion

Models of persuasion, such as the Elaboration Likelihood Model (ELM) (Petty & Cacioppo, 1986) or the Heuristic-Systematic model (Chaiken, 1987)provide a framework to predict message persuasiveness by distinguishing between two routes of persuasion: the peripheral (heuristic) and central (systematic) route. The central route involves more effortful consideration of message content, and factors such as the strength of the argument are more predictive of message effectiveness. On the other hand, persuasion via a peripheral route relies more heavily on heuristic thinking (e.g., how the message makes you feel, whether you like the message etc.). Individuals are more likely to process information via the central route when motivated to engage in effortful thinking; where motivation is an increasing function of the Need For Cognition and personal involvement, defined as the degree to which a person cares about improving their behaviour to improve their outcomes (Braverman, 2008; Kruglanski et al., 2006).

Literature examining the relationship between effortful thinking and the effectiveness of nudge messages is mixed. Some evidence suggests framing effects are attenuated when messages are processed via the central route (McElroy & Seta, 2003; Sieck & Yates, 1997; Simon et al., 2004; Takemura, 1992, 1994). A nudge, in most characterisations, is defined by its ability to target the subconscious routines and biases present in human decision-making and behaviour whilst maintaining personal autonomy (the libertarian in "libertarian paternalism"). In other words, nudge works on System 1 processes (peripheral route) while leaving System 2 unengaged (central route) (Banerjee & John, 2021; Schmidt & Engelen, 2020). Under this assumption, we expect nudges to influence behaviour only when processed via the peripheral route. In other words, subjects with low Personal Involvement and low Need For Cognition should be more likely to overstate adherence in nudge message conditions.

On the other hand, it may be that for a nudge message to move behaviour, it must first be processed more centrally. According to Sunstein (2016), nudge messages, or in-

formation disclosures, *can* work via System 1 (by inculcating an emotional response) or System 2 processes (by informing deliberative processes). Suppose nudge messages are to influence behaviour via System 2. In that case, the information in the message must first be encoded as decision-relevant to be considered at the point-of-decision (Mertens et al., 2022). In other words, for a nudge to influence behaviour, individuals must first engage in effortful cognitive processing. Research has also shown that people are more likely to pay attention to and use an intervention when personal involvement is high. For example, nutrition labels are more frequently used by consumers concerned about their diet and overall health than consumers who do not share those concerns (Campos et al., 2011). Based on this line of thinking, we might expect high personal involvement (a determinant of processing information via the central route) to be predictive of intervention effectiveness. Therefore, subjects with high personal involvement and a high Need For Cognition may be more likely to overstate adherence when exposed to a nudge message. We do find evidence to suggest PI scores slightly moderate nudge bias in TF-self-reports for the £-NHS message condition. The results suggest that for those with higher PI scores, TF-self-reports were nudge biased by the £-NHS message.

In what follows, we investigate the following subject-level characteristics as potential moderators of nudge bias: Impression Management (IM), Need for Affect (NFA), Need for Cognition (NFC), and Personal Involvement (PI) with the issue of improving medication adherence. Then, based on these results, a battery of measures designed to test how the messages made respondents feel and what respondents thought of the messages (i.e., perceived message quality) are probed as potential mediators through which subject-level traits may influence nudge bias.

2.2 Empirical strategy

We begin here by outlining our empirical strategy, model specification, and hypotheses. In Section 2.3, we provide more detail about how moderators and mediators are assessed in Study 3.

On a conceptual level, moderators and mediators are third variables that describe distinct mechanisms through which an independent variable (IV) influences the dependent variable (DV). Baron & Kenny (1986) define a moderator as a third variable that acts to shrink or enhance the relationship between the independent and dependent variables. In contrast, a mediator is a third variable that provides a generative mechanism that carries forward the effect of the independent variable onto the dependent variable. In other words, moderators can tell us *for whom* a nudge biases self-reports, and a mediator can describe *how* a nudge influences self-reports.

As we do not expect the nudges (IV) to influence how respondents respond to standardised personality scale questionnaires, trait variables (IM, PI, NFC, and NFA) are investigated as moderators. On the other hand, affective and cognitive reactions to the messages will be, by definition, a function of the nudge message and therefore are investigated as mediators.

2.2.1 Moderation analysis

To investigate moderation effects, we use simple slopes analysis on moderated regressions. The subject characteristics investigated as moderators are IM, PI, NFC, and NFA. All hypothesised moderator scores are standardised to aid interpretation and minimise multicollinearity.[5] To test our moderation hypotheses, we specify a multiple moderator model to include a two-way interaction term for each moderator and moderated relationship.

Model 1 is constructed to test the moderating effects of IM, PI, NFC, and NFA scores on the relationship between message condition and self-reported adherence. As shown in Figure 2.1, Model 1 tests the hypotheses that the effect of observing a nudge message on self-reported adherence is, in turn, influenced by a subject's IM, PI, NFC, and NFA score. For each subject i, let $Y_{j,i}$ denote the level of self-reported adherence as measured by the MARS-5; where $j = (0, 1)$ indicates whether self-reports are time-

[5]Individual moderator scores were standardised by first subtracting the sample mean score, then dividing this number by the sample score standard deviation.

framed ($j = 1$), or not ($j = 0$). As such, the effects on time-framed ($Y_{1,i}$) and non-time-framed ($Y_{0,i}$) self-reports are estimated separately. T_i is a 1 x 4 vector of treatment message condition dummies. Therefore, γ_{ji} represents a vector of coefficients (1 x 4) denoting the estimated interaction coefficients between moderator variables and the message condition group. Finally, Z_i is a vector of individual-level demographic and medication regime controls.

$$Y_{j,i} = \alpha_{j,i} + \beta_{j,i}T_i + \delta_{1,j,i}IM_i + \delta_{2,j,i}PI_i + \delta_{3,j,i}NFC_i + \delta_{4,j,i}NFA_i + \boldsymbol{\gamma_{1,j,i}}T_i \times \boldsymbol{IM_i} + \boldsymbol{\gamma_{2,j,i}}T_i \times \boldsymbol{PI_i} + \boldsymbol{\gamma_{3,j,i}}T_i \times \boldsymbol{NFC_i} + \boldsymbol{\gamma_{4,j,i}}T_i \times \boldsymbol{NFA_i} + Z_i + \epsilon_{j,i} \quad (1)$$

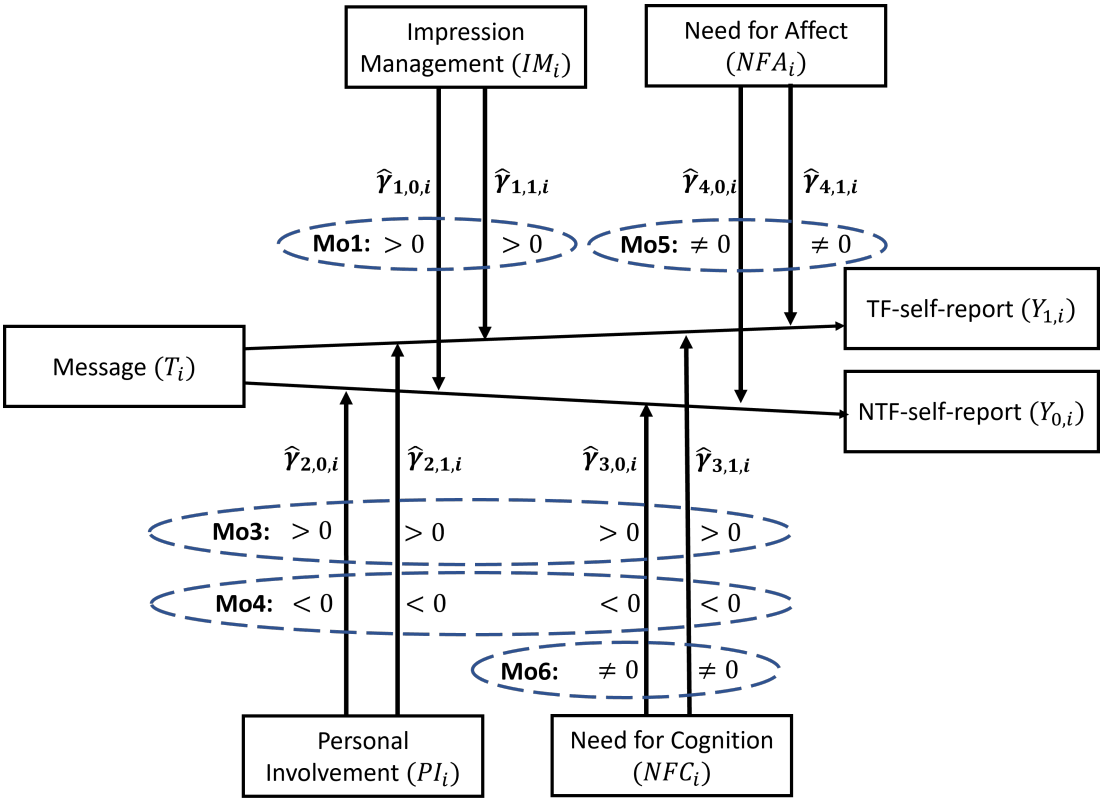


Figure 2.1: Moderation hypotheses tested using Model 1

Figure 2.1 also outlines the moderation hypotheses tested in Model 1. Suppose

nudge bias in self-reports of adherence is driven by social desirability bias. In that case, the self-reports of subjects with higher IM scores should be more sensitive to the presence of a nudge message. Therefore, we hypothesise that the effect of the message on self-reports is increasing in IM score: Mo1: $\gamma_{1,j,i} > 0$.

Models of persuasion provide a dual-system framework to explain how message exposure can lead to changes in attitudes, beliefs, and behaviour. To test whether nudge bias is a by-product of message effectiveness, we test whether predictors of the route of persuasion (PI and NFC) significantly moderate nudge bias in self-reports. Specifically, we test two opposing hypotheses: that predictors of central processing are associated with nudge bias in self-reports (Mo3: $\gamma_{2,j,i} + \gamma_{3,j,i} > 0$), and those predictors of peripheral processing are associated with nudge bias in self-reports (Mo4: $\gamma_{2,j,i} + \gamma_{3,j,i} < 0$).

Additionally, drawing upon existing evidence of heterogeneity in message effectiveness across individual processing styles, we experimentally investigate the moderating effect of NFA and NFC on nudge bias in self-reports of adherence. We do not formulate a prior regarding which nudge message is more congruent with an NFC or NFA orientation, therefore we test the following two-sided hypotheses: that NFA influences the presence of nudge bias in self-reports (Mo5: $\gamma_{4,j,i} \neq 0$), and NFC influences the presence of nudge bias in self-reports (Mo6: $\gamma_{3,j,i} \neq 0$).

Model 2 is an augmented version of Model 1, including a 3-way interaction term between a time-framing dummy variable (j_i), IM score, and message treatment condition. As shown in Figure 2.2, Model 2 tests whether the effect of IM scores on the relationship between a message and self-reported adherence is influenced by whether self-reports were time-framed. Here, π_i is a 4 x 1 vector of coefficients (one for each message treatment condition) on the 3-way interaction. The significance of this 3-way interaction indicates whether the moderating effect of IM scores on the relationship between message condition and self-reports differs depending on whether self-reports are time-framed or not.

$$y_i = \mu_i + \rho_i T_i + \sigma_{1,i} IM_i + \omega_i j_i + \tau_i T_i \times j_i + \theta_{1,i} T_i \times IM_i + \boldsymbol{\pi_i} T_i \times \boldsymbol{IM_i} \times \boldsymbol{j_i}$$
$$+ \sigma_{2,i} PI_i + \sigma_{3,i} NFC_i + \sigma_{4,i} NFA_i + \theta_{2,i} T_i \times PI_i + \theta_{3,i} T_i \times NFC_i + \theta_{4,i} T_i \times NFA_i$$
$$+ Z_i + \epsilon_{j,i} \quad (2)$$

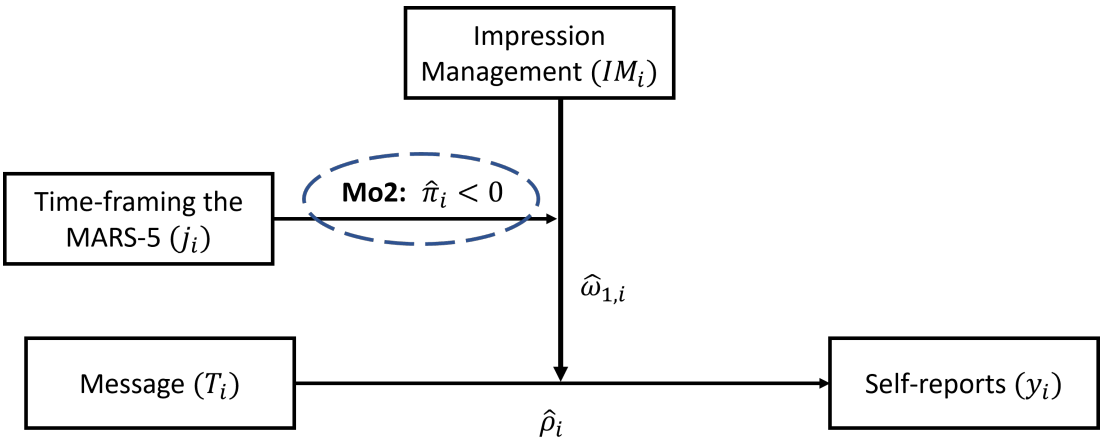


Figure 2.2: Moderation hypotheses tested using Model 2

If selective recall is a vehicle to consciously distort self-reports towards the more socially desirable response, and NTF-self-reports allow subjects to be more selective in recalling past behaviour, then time-framing should reduce social desirability bias before it can occur. Therefore, as shown in Figure 2.2, this leads us to our secondary hypothesis about the moderating effect of IM scores on nudge bias: time-framing the MARS-5 reduces the influence of IM score on the effect of a nudge message on self-reported adherence: Mo2: $\pi_i < 0$.

2.2.2 Mediation analysis

Structural equation modelling is used to investigate potential mediators of nudge bias in self-reports (see Figure 2.3) (Gunzler et al., 2013). The primary hypothesis is whether the effect of the independent variable (IV_i) on the dependent variable (DV_i) is mediated

by a change in the mediating variable (M_i). C is the direct effect, AB is the indirect effect, and the total effect is given by $C' + AB$.

The mediation proportion is defined as the percentage change in regression coefficients when we include the mediating variable in the model (e.g., indirect effect (AB) / total effect ($C' + AB$)). Mediation becomes relevant when we reject the null hypothesis that IV_i has no direct effect on DV_i ($H_0 : C = 0$). If we fail to reject this hypothesis, then the IV_i and DV_i are unrelated, and we do not consider potential mediators. However, if we reject this null hypothesis, assessing partial mediation via the direct, indirect, and total effects may contribute to our understanding of the determinants of nudge bias.

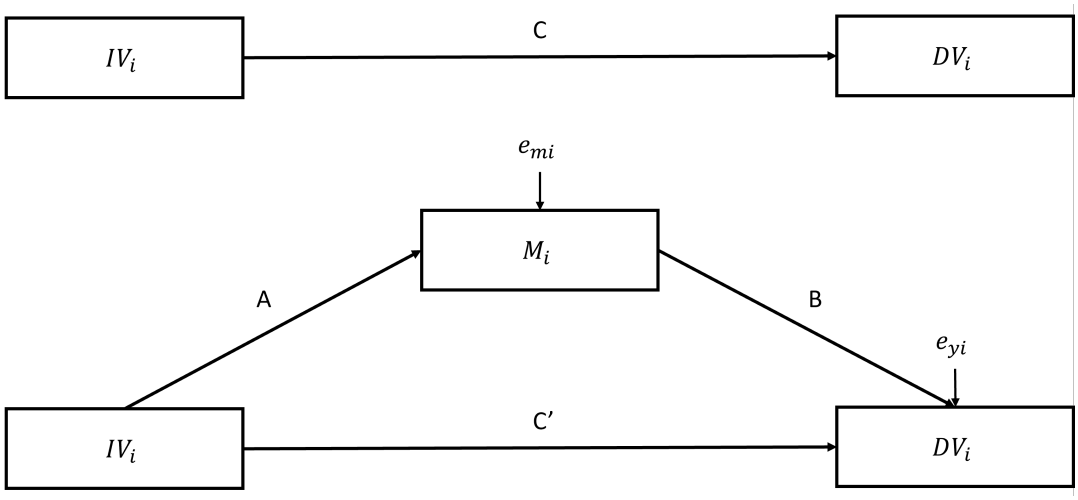


Figure 2.3: Mediation analysis framework

2.3 Design

Core experimental design and procedures for the experiment are presented in Chapter 1.[6] In this section, we present aspects of the experimental design relevant to the investigation of nudge bias mechanisms. Study 3 was conducted in 2 stages. First, participants completed a pre-screening survey to evaluate participant eligibility and measure some potential moderators of message persuasion. Then, subjects who passed

[6] see Section 1.2.1 for general design details, and Section 1.4.1 for Study 3 specific design details.

the eligibility screening (those who reported taking medication for a long-term illness) were invited to complete the main experiment.

2.3.1 Moderator and Mediator measures

During pre-screening, subjects completed a series of measures to index NFC, NFA and PI.[7]

Personal involvement (PI) was measured using two items: (i) How important is it for you to improve how you are taking your medication? and (ii) How happy are you with your health at the moment as it relates to your chronic condition? (1 "Not at all" to 7 "Completely"). A PI score was calculated as the average of the two responses (with the second item reverse coded). After reverse coding, the correlation between items was small but positive ($\rho = 0.2038, p < 0.001$). High personal involvement was defined as a PI score greater than or equal to the sample median of 4.5.

Need for Cognition (NFC) was assessed using a 6-item NFC scale (Lins de Holanda Coelho et al., 2020). Individual NFC scores are calculated as the mean response to all items on a 7-point scale from 1 "Not at all" to 7 "Completely", with negatively keyed items reverse coded. NFC scores range from 1 to 7, where a higher score indicates a greater preference for engaging in cognitive processing.

Need for Affect (NFA) was assessed using the 10-item scale Need for Affect Questionnaire (NAQ-S) (Appel et al., 2012). Individual NFA scores are calculated as the mean response to all items on a 7-point scale from 1 "Not at all" to 7 "Completely", with negatively keyed items reverse coded. NFA scores range between 1 and 7, where a high score indicates a high preference to approach emotion, and a low score, to avoid emotion.

NFA-NFC Orientation was assessed as the difference between NFA and NFC score (Aquino et al., 2020). The average NFA score in the sample was 4.60 ($SD = 0.96$), and the average NFC score was 4.63 ($SD = 1.14$). NFA and NFC scores were slightly

[7] For exact question wording and experimental instructions in the pre-screening stage, see Appendix A.1.5.

positively correlated ($\rho = 0.2038, p < 0.001$). A higher orientation score indicates higher reliance on affect as information. Conversely, a lower score indicates a higher reliance on cognition. The average orientation score in the sample is -0.03 ($SD = 1.33$). Subjects are coded as NFA-oriented if the orientation score is positive (n = 1,019) and NFC-oriented if the orientation score is negative (n = 1,042).[8]

In the main experiment, after subjects were shown a message and asked to complete a self-report of medication adherence, IM scores, along with affective and cognitive reactions to the message, were assessed:

Impression management (IM) is assessed via the Balanced Inventory of Desirable Responding scale (BIDR-16) (Hart et al., 2015; Paulhus, 1988; Stöber et al., 2002). Responses to the BIDR are used to calculate two subscales: impression management (IM) and socially desirable enhancement (SDE) ($\rho = 0.3452, p < 0.001$). Scores range from 1 to 7.

Affective and cognitive reaction to the message

After subjects completed self-reports of adherence in the main experiment, they were asked to evaluate cognitive and emotional responses to the message.[9] Subjects in a nudge message condition completed both an affective and cognitive evaluation of their message. The order within which subjects completed the affective and cognitive appraisal was determined randomly to control for potential spillover effects of one form of evaluation on the other. Subjects in the control group only completed the affective evaluation.

The Discrete Emotions Questionnaire (DEQ) (Harmon-Jones et al., 2016) was used to assess affective response when reading the message. The scale was implemented as follows: all subjects (including the control group) were shown the message again and provided with a list of 32 emotions. Subjects were asked to select all the emotions they experienced to any degree while reading the message.[10] The scale assesses affective

[8] No subject in the sample had an orientation score of exactly 0.

[9] See Appendix B.1.1 for affective and cognitive questions as participants saw them.

[10] Respondents were also asked to rate the degree with which the emotions were felt on a scale from 1

response in terms of 8 emotions (Anger, Disgust, Fear, Anxiety, Sadness, Desire, Relaxation, and Happiness) via 8 4-item subscales. For each emotion, a binary score is created to indicate whether the emotion was felt or not.

Discrete emotions can be categorised along multiple dimensions, including valence (positive/negative), arousal (high/low), and motivational direction (approach/avoid). Anger is often regarded as a negative, high-arousal emotion associated with approach motivation (Berkowitz & Harmon-Jones, 2004; Carver & Harmon-Jones, 2009; Peterson & Harmon-Jones, 2012). Disgust and Fear are high-arousal negative emotions associated with avoidance motivation (Harmon-Jones et al., 2016). Anxiety has a lot of overlap with Fear (both being high arousal and negative emotions), however, it is instead associated with behavioural conflict (both approach and avoidant motivation). Sadness is a negative, low arousal emotion, mostly associated with approach (Panksepp, 2004). Happiness is a positive emotion associated with the approach motivational system, and Relaxation and Desire are positive emotions associated with low and high approach respectively (Harmon-Jones et al., 2016).

Moving to cognitive evaluations of nudges, subjects in a nudge message condition were asked to read a series of statements about the message and indicate the degree to which they agree on a scale from 1 "Not at all" to 7 "Completely". Responses were then used to create the following scores:

Appraisal (Kim et al., 2021) is the mean response to 3 items: (i) the message is interesting, (ii) the message is worth sharing, (iii) I like the message (Cronbach's alpha = 0.7484).

Engagement (Godinho et al., 2016; Vidrine et al., 2007) is the mean response to 4 items: (i) the message is relevant to me, (ii) the message makes me think about my actions, (iii) the message grabs my attention, (iv) the message is memorable (Cronbach's alpha = 0.7487).

Quality (Braverman, 2008; Godinho et al., 2016) is the mean response to 2 items:

"Slightly" to 4 "Very much". The sum of responses across four items can be used to create an individual affective score for each of the 8 emotions ranging from 0 to 16.

(i) the message is credible, (ii) the message is believable (Cronbach's alpha = 0.7860).

Intentions (Braverman, 2008) is the mean response to 2 items: (i) the message makes me want to improve my adherence, (ii) the message will help me to improve my adherence (Cronbach's alpha = 0.8415).

Perception of adherence as the desirable behaviour (desirability) is the mean response to 3 items: (i) the message reminds me what I ought to do, (ii) the message reminds me of something I want to do, (iii) the message is trying to make me do something good (Cronbach's alpha = 0. 7189).

Psychological reactance is a process that occurs when a subject perceives their autonomy to be threatened (Brehm & Brehm, 2013; Reynolds-Tylus, 2019; Torrance & Brehm, 1968). Reactance to persuasive messages can lead to interventions backfiring (i.e., people doing the opposite of what the message is encouraging them to do) in an attempt to re-establish threatened or lost freedom (Byrne & Hart, 2009). The brief-Reactance to Health Warnings Scale (brief-RWHS) was used to assess reactance to the message (Hall et al., 2017). The scale consists of 3 items: (i) This message is trying to manipulate me, (ii) This message is overblown, and (iii) This message annoys me. The mean response across items was used as a score of reactance to the message (Cronbach's alpha = 0. 7551).

2.3.2 Outcome measures

The main outcome variable is self-reports of medicines adherence measured using the 5-item Medication Adherence Report Scale (MARS-5) (Chan et al., 2020). The MARS-5 assesses how adherent subjects are based on the frequency (on a 5-point Likert scale from "Never" to "Always") with which they diverge from their prescribed regime. The items are designed to measure intentional non-adherent behaviour (change dose, stop medication, skip a dose, use less than prescribed) and unintentional non-adherent behaviour (forget dose) (Horne & Weinman, 2002). In Study 3, subjects either completed the MARS-5 with time-framed recall instructions (TF) or no recall instructions (NTF) (see Appendix A.1.2).

The continuous scoring method is used to assess the level of self-reported adherence. Responses to each MARS-5 item are coded as follows: 5 = Never, 4 = Sometimes, 3 = About half the time, 2 = Most of the time, 1 = Always. Continuous scores are calculated as the sum of responses to the MARS-5, and therefore are bound between 5 (perfect non-adherence) and 25 (perfect adherence). From this point onwards, we will refer to the continuous adherence score from responses to the NTF MARS-5 as the NTF-self-report, and the continuous adherence score calculated from responses to the TF MARS-5 as the TF-self-report.

2.3.3 Screening, sampling, & randomisation

All participants were recruited via Prolific. Participation in the experiment was conditional on passing Prolific pre-screening criteria: (i) live in the UK, and (ii) be diagnosed with a chronic illness. In addition, we also included a question at the top of the experiment to ensure participants were currently taking medication for a long-term illness.

After subjects completed the pre-screening stage, eligible respondents were categorised into one of four strata based on the crosscut of PI and orientation scores: 2 (High PI, Low PI) x 2 (NFC oriented, NFA oriented). PI and orientation scores were orthogonal ($\rho = -0.008, p = 0.8087$). Then, upon entry into the main experiment, subjects were block randomised evenly into a message and measurement condition group by strata.

2.4 Results

Sample characteristics differ somewhat between strata in Study 3. However, tests for joint orthogonality confirm that respondent characteristics are balanced across message ($p = 0.6717$) and time-frame self-report conditions ($p = 0.0798$).[11]

[11] For more information about the number of respondents recruited, randomised, and analysed by strata see Appendix B.2.1. For a more detailed description of sample characteristics in Study 3 by strata, see Appendix B.2.2.

Table 2.1 summarises the outcome, moderator, and mediator variables and their correlations with self-reports of adherence.[12]

[12]See Appendix B.2.3 for a summary of correlations between moderator and mediator variables.

			Correlation		F-test/chi2 test	
					Message	Time-frame
	M	SD	(1)	(2)	p	p
Outcome variables						
(1) NTF-self-reports	22.54	2.78	1.000	.	0.2759	.
(2) TF-self-reports	23.11	2.77	.	1.000	0.0156	.
Subject Characteristics						
NFC/NFA orientation	-0.02	1.31	0.0483	0.0340	0.9020	0.9798
Need for Cognition (NFC)	4.63	1.12	0.0157	0.0060	0.8637	0.7792
Need for Affect (NFA)	4.61	0.96	0.0852*	0.0533	0.8539	0.7701
Personal Involvement	4.18	1.41	-0.1264*	-0.1431*	0.4993	0.9356
Importance of improving behaviour	4.50	1.95	-0.1212*	-0.1221*	0.6579	0.9146
Happy with current outcome	4.15	1.71	0.0697	0.0975*	0.5557	0.9912
Impression management	4.49	1.00	0.1390*	0.1270*	0.4791	0.2309
Message Characteristics						
Appraisal†	4.75	1.36	0.1692*	0.1437*	0.0360	0.7349
Engagement†	4.78	1.38	0.0061	0.0315	0.0010	0.8325
Quality†	5.64	1.20	0.1926*	0.2043*	0.0000	0.2615
Intentions†	4.12	1.89	-0.0487	0.0311	0.0133	0.4740
Reactance†	2.35	1.29	-0.1508*	-0.2234*	0.0329	0.3285
Adherence desirability†	4.62	1.58	-0.0571	-0.0009	0.0000	0.0130
% Anger	0.18	0.39	-0.0176	-0.1061*	0.0000	0.5223
% Anxiety	0.38	0.49	-0.1743*	-0.1998*	0.0009	0.4854
% Disgust	0.09	0.28	-0.0158	-0.1232*	0.0000	0.0102
% Fear	0.14	0.34	-0.1347*	-0.1261*	0.0010	0.8537
% Sadness	0.25	0.43	-0.1038*	-0.1481*	0.0000	0.8099
% Desire	0.09	0.29	-0.0941*	-0.0861*	0.0000	0.5414
% Relaxed	0.38	0.48	0.0421	0.0483	0.0000	0.1507
% Happiness	0.19	0.39	0.0357	0.0513	0.0000	0.1624
% Approach	0.59	0.49	-0.0950*	-0.1279*	0.0000	0.3610
% Avoidance	0.41	0.49	-0.1559*	-0.1974*	0.0000	0.2936
% Positive	0.45	0.50	0.0331	0.0257	0.0000	0.4915
% Negative	0.49	0.50	-0.1472*	-0.2000*	0.0000	0.7082
% High Arousal	0.55	0.50	-0.0793	-0.1214*	0.0000	0.3651
% Low Arousal	0.56	0.50	-0.0119	-0.0316	0.0000	0.2785

Note: n = 2061 (†: n = 825). Pearson correlation coefficients are displayed in column (1) for associations with NTF-self-reports and (2) for TF-self-reports. Significance level indicated: * p <0.001. Approach: Anger, Anxiety, Sadness, Desire, and Happiness. Avoidant: Anxiety, Disgust, and Fear. Positive: Desire, Relaxation, and Happiness. Negative: Anger, Anxiety, Disgust, Fear and Sadness. High Arousal: Anger, Anxiety, Disgust, Fear, Desire, and Happiness. Low Arousal: Sadness and Relaxation.

Table 2.1: Study 3 summary of outcome, moderator, and mediator variables

Impression management: High IM scores indicate a greater proclivity to respond in a socially desirable manner.[13] On average, IM scores were positively correlated with both NTF- ($\rho = 0.1390, p < 0.001$) and TF-self-reports ($\rho = 0.1270, p < 0.001$). The results of Model 1 are summarised in Table 2.2. These regression results show that independent of the message condition, IM scores were positively associated with TF-self-reports of adherence ($\beta = 0.388, p < 0.001$) but not NTF-self-reports.

The estimated moderating relationships between IM score and message condition were not statistically significant. Only the interaction term between being in the £-NHS condition and IM scores for TF-self-reports approaches significance; however, it does so in the opposite direction as the one hypothesised (£-NHS x IM: $\beta = -0.451, p = 0.095$). As such, we find no evidence to support Mo1.

Simple slopes analysis is summarised in Appendix B.2.4). For NTF-self-reports, we find that IM did not influence the message condition's effect on self-reports. Higher IM scores only significantly increased TF-self-reports in the control condition.[14]

Finally, as shown in Table 2.3, we find no evidence to support Mo2: time-framing the MARS-5 did not decrease the effect of IM scores on the relationship between message condition and self-report (nudge bias).[15]

[13] IM scores were non-normally distributed ($Adj.\chi 2(2) = 7.82, p = 0.0201$).

[14] See Table B.3.

[15] Notably, when Model 2 is estimated with subject-level controls (particularly age), the effect of IM scores on self-reports becomes statistically insignificant. Indeed, IM scores were found to have a small positive correlation with age ($\rho = 0.2214, p < 0.001$).

	(1) TF-self-reports		(2) NTF-self-reports	
	β	*(se)*	*β*	*(se)*
Message condition				
Health	**0.555****	(0.254)	0.186	(0.275)
£-NHS	**0.605****	(0.279)	0.443*	(0.259)
Burden-NHS	0.374	(0.266)	0.182	(0.272)
Descriptive norm	-0.454*	(0.263)	0.147	(0.269)
IM	**0.388*****	(0.104)	0.113	(0.106)
Health x IM	-0.288	(0.251)	0.175	(0.285)
£-NHS x IM	-0.451*	(0.270)	-0.147	(0.281)
Burden-NHS x IM	-0.175	(0.263)	-0.169	(0.255)
Norm x IM	-0.057	(0.281)	0.153	(0.263)
PI	**-0.604*****	(0.102)	**-0.369*****	(0.107)
Health x PI	0.419	(0.270)	-0.092	(0.279)
£-NHS x PI	0.513*	(0.282)	-0.121	(0.267)
Burden-NHS x PI	0.403	(0.288)	0.310	(0.257)
Norm x PI	0.240	(0.263)	0.109	(0.265)
NFC	-0.116	(0.105)	-0.047	(0.105)
Health x NFC	0.034	(0.298)	-0.301	(0.289)
£-NHS x NFC	0.217	(0.246)	-0.126	(0.265)
Burden-NHS x NFC	-0.275	(0.303)	0.254	(0.271)
Norm x NFC	-0.337	(0.246)	-0.155	(0.311)
NFA	0.080	(0.106)	**0.251****	(0.102)
Health x NFA	-0.124	(0.253)	**-0.641****	(0.314)
£-NHS x NFA	-0.363	(0.297)	-0.095	(0.270)
Burden-NHS x NFA	0.116	(0.264)	-0.453*	(0.272)
Norm x NFA	0.096	(0.260)	0.005	(0.301)
Constant	**18.921*****	(0.524)	**18.547*****	(0.612)
Controls	*YES*		*YES*	
R^2	0.196		0.198	
F-stat	6.622		6.697	
N	1016		1013	

Note: Significance indicated: * p < 0.1, ** p < 0.05, *** p <0.01. Standardised scores were used for moderator variables. Controls include demographics (age, education) and medication regime (recency, subjective complexity, importance) characteristics. Baseline group: Message condition = control group.

Table 2.2: Model 1 regression results

	(1)		(2)	
	Self-reports			
	β	*(se)*	β	*(se)*
Message condition				
Health	0.185	(0.301)	0.141	(0.272)
£-NHS	**0.640****	(0.281)	0.436*	(0.256)
Burden-NHS	0.373	(0.296)	0.196	(0.269)
Descriptive norm	0.143	(0.292)	0.175	(0.266)
IM	**0.309*****	(0.113)	0.147	(0.103)
Health x IM	0.248	(0.305)	0.114	(0.275)
£-NHS x IM	-0.042	(0.302)	-0.148	(0.275)
Burden-NHS x IM	-0.142	(0.277)	-0.139	(0.251)
Norm x IM	0.158	(0.284)	0.130	(0.258)
Time-frame (TF)	**0.492*****	(0.155)	**0.490*****	(0.141)
Health x TF	0.530	(0.413)	0.404	(0.373)
£-NHS x TF	0.031	(0.412)	0.170	(0.377)
Burden-NHS x TF	0.111	(0.416)	0.243	(0.379)
Norm x TF	-0.312	(0.412)	-0.637*	(0.375)
TF x IM	0.198	(0.158)	0.222	(0.143)
Health x TF x IM	-0.486	(0.401)	-0.337	(0.365)
£-NHS x TF x IM	-0.347	(0.418)	-0.315	(0.382)
Burden-NHS x TF x IM	0.033	(0.397)	-0.056	(0.361)
Norm x TF x IM	-0.285	(0.408)	-0.226	(0.371)
Constant	**22.418*****	(0.109)	**18.493*****	(0.398)
Add. Moderators	YES		YES	
Controls	NO		YES	
R^2	0.067		0.197	
F-stat	4.304		10.571	
N	2061		2029	

Note: Significance indicated: * p < 0.1, ** p < 0.05, *** p < 0.01. Standardised scores were used for moderator variables. Additional moderators and their interactions with message conditions are included in both models: PI, NFC, and NFA. Controls include demographics (age, education) and medication regime (recency, subjective complexity, importance) characteristics. Baseline group: Message condition = control group; Time-frame = NTF-self-reports.

Table 2.3: Model 2 regression results

2.4.1 Moderation

NFC and PI: PI was negatively correlated with TF- ($\rho = -0.1431, p < 0.001$) and NTF- self-reports ($\rho = -0.1264, p < 0.001$). In other words, respondents who ascribed a higher degree of importance to improving their adherence were, on average, also the ones who reported being less adherent to their medication.[16] Average NFC score was not correlated with self-reports of adherence.[17] The regression analysis summarised in Table 2.2 shows that subjects with higher PI scores, on average, reported a lower degree of adherence ($\beta = -0.604 p < 0.001$). In contrast, the NFC score was orthogonal to self-reports of adherence.

A closer examination of the interaction terms between PI score and message conditions reveals that PI score positively moderated TF-self-reports in the £-NHS condition in the direction proposed by Mo3 ($\beta = 0.513, p = 0.069$) (see Table B.4 in Appendix B.2.4). This suggests nudge bias in TF-self-reports in the £-NHS condition was partly driven by subjects with higher PI scores. However, no such effect was observed in NTF-self-reports or for NFC scores.

NFC and NFA: While NFC does not moderate the effect of message condition on self-reports, NFA appears to negatively moderate NTF-self-reports in the Health message condition ($\beta = -0.641, p = 0.042$) and, to a lesser degree, the Burden-NHS message condition ($\beta = -0.453, p = 0.097$) (see Table 2.2).[18] While NFA was associated with higher average self-reports of adherence ($\beta = 0.251, p = 0.014$) on average, for subjects in the Health and Burden-NHS message conditions, higher NFA scores were associated with lower self-reported adherence. This effect is detected in NTF-self-reports but not in TF-self-reports.

[16] PI scores are non-normally distributed ($Adj.\chi 2(2) = 38.46, p < 0.0001$).
[17] NFC scores are also non-normally distributed ($Adj.\chi 2(2) = 30.05, p < 0.0001$).
[18] NFA scores were non-normally distributed ($Adj.\chi 2(2) = 10.64, p = 0.0049$).

2.4.2 Affective and cognitive evaluations of messages

In Table 2.1, we find that some cognitive and affective appraisals of the messages were associated with self-reports of adherence. More specifically, greater self-reported adherence in nudge message conditions was associated with higher message appraisal ($\rho_{NTF} = 0.1692, \rho_{TF} = 0.1437$) higher message quality ($\rho_{NTF} = 0.1926, \rho_{TF} = 0.2043$), and lower reactance ($\rho_{NTF} = -0.1508, \rho_{TF} = -0.2234$). In terms of affect, messages that led to feelings of anger, anxiety, disgust, fear, sadness, and desire were associated with lower self-reports of adherence.[19] The emotions associated with lower self-reports of adherence were negative in valance and associated with both approach and avoidance motivation.

Affect scores were more closely associated with TF-self-reports of adherence than NTF-self-reports. Notably, anger, disgust, and high arousal emotions were associated with TF-self-reports, but not NTF-self-reports.[20]

[19] We cannot be sure about the direction of causality here; however, it seems likely that subjects with lower adherence to medicines (who report as much) were more likely to report negative and high arousal emotions when reading the message.

[20] For comparisons in the affective responses for each message conditions, between subjects who completed the TF-and NTF-self-report, see Appendix B.2.5.

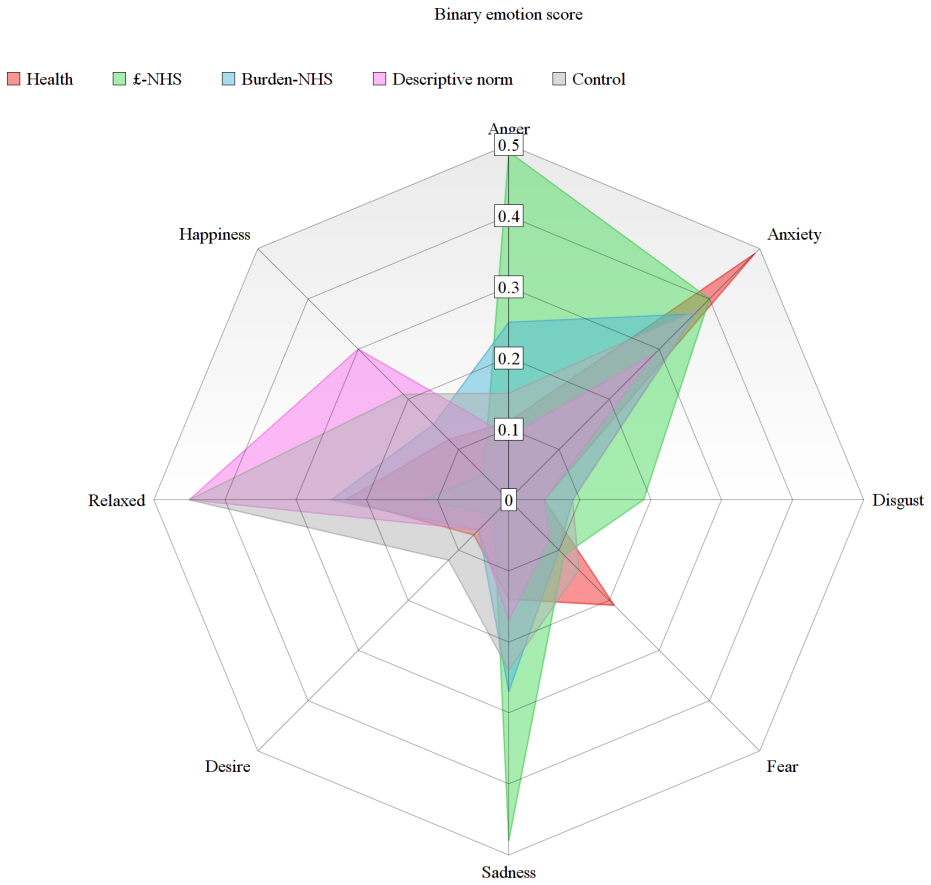


Figure 2.4: Proportion of sample reporting each emotional response across message conditions

Figure 2.4 and Table 2.4 show affective reactions elicited across message conditions in Study 3. The health message was associated with anxiety (49%) and fear (21%). The £-NHS message was the most likely to provoke feelings of anger (49%), disgust (19%) and sadness (48%). The Burden-NHS message created an average emotional response between the Health and £-burden messages. On the other hand, the descriptive norm message elicited the most positive affective reaction out of all nudge messages (e.g., happiness was reported by 30% of the subjects in this condition.)

Subjects in the control group did not answer the cognitive evaluation items. Therefore, the descriptive norm group (as this is the only group where we do not observe

nudge bias) is used as a baseline to assess differences in cognitive appraisal. As shown in Table 2.4, the £-NHS message was rated as having the highest appraisal score. The Health message was given the highest engagement, message quality, and adherence desirability scores. Finally, the Burden-NHS message had the highest average reactance score.

	Control		Health		£-NHS		Burden-NHS		Norm	
	M	SD	M	SD	M	SD	M	SD	M	SD
Outcome variables										
NTF-self-reports	22.43	2.85	22.52	3.05	**23.03****	2.19	22.78	2.45	22.51	2.87
TF-self-reports	22.94	2.91	**23.67****	1.79	23.53*	2.08	**23.46*****	2.64	22.79	3.36
Evaluation										
Appraisal	.	.	4.80*	1.27	**4.93*****	1.19	4.72	1.57	4.55	1.35
Engagement	.	.	**5.03*****	1.25	**4.86*****	1.33	4.71	1.57	4.50	1.33
Quality	.	.	**5.95*****	0.99	**5.71*****	1.2	**5.63*****	1.28	5.26	1.21
Intentions	.	.	**4.26*****	1.78	**4.20****	1.9	**4.27*****	1.96	3.75	1.89
Reactance	.	.	2.33	1.31	2.33	1.28	**2.55****	1.44	2.17	1.09
Desirability	.	.	**5.07*****	1.39	4.28	1.61	4.71	1.53	4.41	1.69
% Anger	0.15	0.35	0.11	0.31	**0.49*****	0.50	**0.25*****	0.43	**0.09****	0.28
% Anxiety	0.38	0.48	**0.49*****	0.50	0.40	0.49	0.37	0.48	**0.29****	0.45
% Disgust	0.09	0.28	0.05*	0.21	**0.19*****	0.39	0.09	0.28	0.05*	0.22
% Fear	0.14	0.35	**0.21****	0.41	0.11	0.31	0.10*	0.30	**0.08****	0.28
% Sadness	0.24	0.43	**0.14*****	0.35	**0.48*****	0.50	0.27	0.45	**0.17****	0.38
% Desire	0.12	0.32	**0.07****	0.25	**0.03*****	0.18	**0.06****	0.24	**0.06****	0.24
% Relaxed	0.45	0.50	**0.23*****	0.42	**0.12*****	0.33	**0.25*****	0.43	0.45	0.50
% Happiness	0.21	0.41	**0.12*****	0.32	**0.05*****	0.22	**0.15****	0.36	**0.30*****	0.46
% Approach	0.32	0.47	**0.23*****	0.42	**0.54*****	0.50	0.37	0.48	0.34	0.48
% Avoidance	0.39	0.49	**0.51*****	0.50	**0.49*****	0.50	0.41	0.49	**0.31****	0.46
% Positive	0.54	0.50	**0.30*****	0.46	**0.15*****	0.36	**0.30*****	0.46	0.55	0.50
% Negative	0.45	0.50	**0.55*****	0.50	**0.78*****	0.42	**0.57*****	0.50	**0.35*****	0.48
% High Arousal	0.51	0.50	**0.59****	0.49	**0.71*****	0.45	**0.62*****	0.49	0.52	0.50
% Low Arousal	0.60	0.49	**0.35*****	0.48	0.58	0.49	**0.49*****	0.50	0.54*	0.50

Note: Sample size for control group = 1,236, health message = 208, £-NHS = 207, burden-NHS = 204, norms = 206. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1. Significance is indicated for non-parametric tests (Wilcoxon ranksum test) for continuous variables and Chi2 tests for dichotomous variables with the control group as the baseline. The descriptive norm message group is used as a baseline where the variable is not measured in the control group. Approach indexes whether anger, anxiety, desire, happiness, or sadness were experienced. Avoidance indexes whether anxiety, fear, or disgust was experienced. Positive indexes, whether happiness, relaxation, or desire, were experienced. Negative indexes, whether anxiety, disgust, sadness, fear, or anger, were experienced. High arousal indexes whether fear, anger, disgust, anxiety, desire, or happiness were experienced. Low arousal indexes whether relaxation or sadness was experienced.

Table 2.4: Average outcome and mediator variable scores across message group

2.4.3 Mediation

PI score → NTF-self-reports

Subjects with high PI scores report lower adherence than those with low PI scores in the NTF-self-report. PI score was also correlated with affective and cognitive reactions to the message. Those with high PI scores were more likely to report emotional reactions (anger, anxiety, disgust, fear, sadness, desire) and higher message appraisal ($\rho = 0.0998$), engagement ($\rho = 0.1705$), and intentions ($\rho = 0.1837$) scores.

The results of mediation analysis between PI scores and NTF-self-reports are summarised in Table 2.5 and Figure 2.5. The results show intention and appraisal score mediated 33% of the relationship between PI and self-reports of NTF-adherence. In other words, subjects with high PI reported being more adherent (in the NTF-self-report) when they liked the message (appraisal) and less adherent (in the NTF-self-report) when the message had a positive influence on their intentions to be adherent. However, based on the results of this experiment, we cannot be sure of the causal direction of these relationships. While subjects with higher PI scores also reported lower adherence in the TF-self-report, reactions to the messages did not mediate the effect of PI on TF-self-reports.

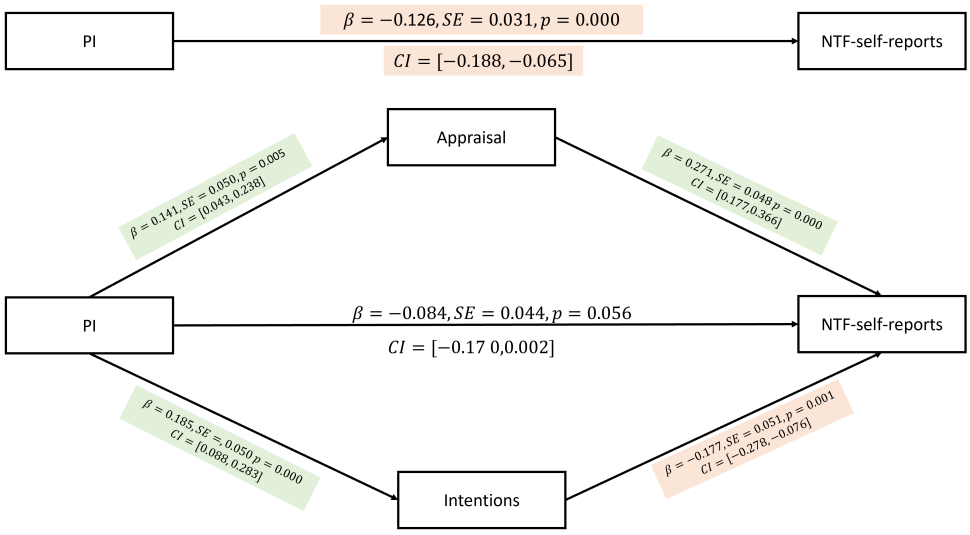


Figure 2.5: The relationship between PI score and NTF-self-reports as mediated by appraisal and intentions scores

	X	M	Y	β	SE	p	95% CI	
C	PI		NTF-self-reports	-0.126	0.031	0.000	-0.188	-0.065
A	PI	Appraisal	NTF-self-reports	0.141	0.050	0.005	0.043	0.238
A	PI	Intentions	NTF-self-reports	0.185	0.050	0.000	0.088	0.283
B	PI	Appraisal	NTF-self-reports	0.278	0.052	0.000	0.177	0.380
B	PI	Intentions	NTF-self-reports	-0.181	0.054	0.001	-0.288	-0.075
C'	PI	Appraisal & Intentions	NTF-self-reports	-0.086	0.045	0.057	-0.174	0.002

Note: Standardized coefficients estimated using maximum likelihood estimator with robust standard errors. Error variance for NTF-adherence = 0.937, appraisal score = 0.980, intentions score = 0.966. Covariance between appraisal and intentions = 0.522 ($p < 0.001$)

Table 2.5: The relationship between PI score and NTF-self-reports as mediated by appraisal and intentions scores

NFA score → NTF-self-reports

NFA score was correlated with a selection of cognitive reactions to the message. Namely, message appraisal ($\rho = 0.0966$), engagement ($\rho = 0.1241$), quality ($\rho =$

0.0947), adherence social desirability ($\rho = 0.0911$), and reactance ($\rho = -0.0919$). However, in mediation analysis, the effect of NFA on NTF-self-reports was mediated only by message quality score. The mediation analysis results are summarised in Figure 2.6 and Table 2.6. Message quality mediated 32% of the relationship between NFA and NTF-self-reports. This suggests that the positive relationship between NFA and NTF-self-reports is partially driven by subjects with higher NFA thinking that the message is high quality (believable and credible).

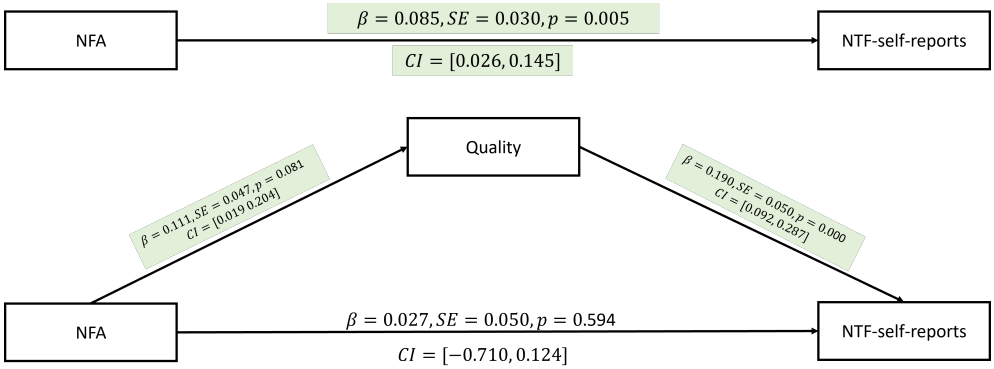


Figure 2.6: The relationship between NFA score and NTF-self-reports as mediated by quality score

	X	M	Y	β	SE	p	95% CI	
C	NFA		NTF-self-reports	0.085	0.030	0.005	0.026	0.145
A	NFA	Quality	NTF-self-reports	0.111	0.047	0.018	0.019	0.204
B	NFA	Quality	NTF-self-reports	0.190	0.050	0.000	0.092	0.287
C'	NFA	Quality	NTF-self-reports	0.027	0.050	0.594	-0.710	0.124

Note: Standardized coefficients estimated using maximum likelihood estimator with robust standard errors. Error variance for NTF-adherence = 0.962, quality = 0.988.

Table 2.6: The relationship between NFA score and NTF-self-reports as mediated by appraisal and intentions scores

2.5 Discussion & Conclusion

This chapter investigates the mechanisms driving nudge bias self-reports of adherence. Using data from the experiment in Study 3, we experimentally examine two possible channels through which nudge messages may directly influence self-reports when behaviour cannot yet change. Ultimately, we find that neither measures of social desirability bias nor indicators of nudge effectiveness can fully explain why subjects in nudge message conditions report greater adherence than subjects in the control. However, we find evidence to suggest that TF-self-reports offer a more accurate measure of actual adherence relative to NTF-self-reports. Finally, we conclude with recommendations for policy-makers interested in controlling for nudge bias in evaluating and analysing nudge message effects estimated with self-report data.

In terms of support for TF-self-reports providing a more accurate measure of patient adherence, we make three observations.

First, we find a stronger relationship between PI and lower self-reported adherence when self-reports are time-framed ($\beta = -0.604, p < 0.001, CI = [-0.805, -0.404]$), relative to when they are not time-framed ($\beta = -0.369, p = 0.001, CI = [-0.580, -0.159]$). In general, we might expect PI and actual adherence to be correlated. For example, all else held constant, subjects who report having a greater desire to improve their adherence and being less satisfied with their current health (PI) may do so directly because they are less adherent to their medications. Under this assumption, our results indicate that TF-self-reports provide a more accurate measure of actual behaviour. Additionally, for subjects in a nudge message condition, the effect of the PI score on NTF-self-reports was mediated by subjective cognitive judgements about the quality of the message and the effect of the message on intentions to be adherent. In contrast, no such mediation effect was detected for TF-self-reports.

Second, we find evidence that NFA, a trait associated with a subject's tendency to approach rather than avoid strong affective experiences, influences NTF-self-reports. Unlike PI, however, we would not expect NFA to be strongly associated with actual

medication-taking behaviour. Despite this, we find a significant relationship between NTF-self-reports and NFA in the regression analysis. The relationship between NFA and NTF-self-reports differs significantly between nudge message conditions. Notably, for subjects with high NFA, the effect of being in the health (and to a lesser extent the burden-NHS message) had a negative effect on NTF-self-reports.[21] In other words, the level of adherence subjects report when seeing the health message depends on NFA score only when no time frame is specified.

Third, we find that emotional reactions (particularly the high-arousal negative emotions, anger and disgust) to nudge messages are more strongly associated with TF-self-reports than NTF-self-reports. In general, we might expect subjects with higher levels of non-adherence to experience a greater negative emotional response when reading the nudge messages (i.e., a message about the cost of non-adherence to the NHS is more likely to make a reader feel bad if they are non-adherent to their medication). Under this assumption, a stronger relationship between negative emotions and the level of non-adherence reported is consistent with our interpretation that TF-self-reports provide a more accurate measure of actual adherence. It should be noted that controlling for affective reactions to the messages in the regression analysis has no effect on the relationship between message condition and self-reports. Therefore, nudge bias is not associated with the strength and type of affective reaction to a message in our experiment.

We, therefore, conclude, that while TF-self-reports may provide a more accurate - and less noisy - measure of actual adherence behaviour, they are not nudge-robust. Indeed, we find nudge bias in TF-self-reports was slightly moderated by the degree of personal involvement in the issue of improving medication adherence. Therefore, policymakers and researchers interested in evaluating the effects of nudge messages on medication-taking behaviour should use subjective and objective measures where possible. However, when self-reports are the most feasible measure, three recommend-

[21] Alternatively, the interaction effect could be interpreted as the message condition moderating the relationship between NFA and NTF-self-reports.

ations can be made based on the results of Chapters 1 and 2: (1) a suitable recall period should be specified at the point of self-report, (2) social desirability bias should be measured and controlled for using a suitable scale, and (3) to improve confidence in the estimated effect of a nudge on behaviour, the direct effect of the nudge on self-reports when behaviour cannot yet change (i.e., nudge bias) should be estimated using an independent, comparable subject pool. Therefore, while nudges may lead individuals to overstate adherence, we submit that the effect of nudge bias on self-reports can be estimated and controlled for in the analysis, in a similar way to social desirability bias.

To this end, our contribution is threefold. First, we identify a replicable weakness in self-reports as a nudge message evaluative measure. Second, our novel experimental design provides a framework for policy-makers to estimate nudge bias and increase confidence in the robustness of behavioural evaluations based on self-reports, and for researchers, to further study the psychological processes underlying nudge bias. Third, we have shown that nudge bias cannot be explained by social desirability bias (as measured via Impression Management) or individual heterogeneity in terms of the route of persuasion and NFA and NFC processing styles.

Part II

Chapter 3

Communicating the move to individualised donor selection policy: altruism and safety frames

Abstract[1]

Background: In recent history, men-who-have-sex-with-men (MSM) have been deferred from donating blood. However, recent evidence supports the adoption of donor screening based on individual sexual behavior over population-based criteria. We explore how best to frame communications about such a change to minimize negative consequences (e.g., reduced number of donors). We examine the effectiveness of risk (safety vs reduced risk), and altruism (donor vs recipient) frames with respect to considerations to donate (approach) or feel deterred (avoid), and mechanisms linked to under-reporting sexual behavior (e.g., embarrassment).

Study Design and Methods: We conduct a 2 (risk frame: reduced risk vs. safety) by 3 (altruism frame: donor, vs. recipient vs. both) between-subjects online experiment (n = 2677). The main outcomes were intentions to donate and feelings of being deterred (both self and others) from donating. We also assess the extent to which forgetting, embarrassment/shame and question irrelevance were perceived to be associated with under-reporting sexual behaviour.

Results: Safety frames or frames focused on the recipient resulted in people being less likely to feel deterred from donating. People from ethnic minorities were more likely, regardless of the frame, to be deterred. Forgetting, embarrassment/shame and perceived irrelevance of reporting behaviours were all associated with higher avoidance, and using smartphones was perceived as an acceptable memory aid.

Discussion: Transfusion services moving to an individualized policy should frame donor selection in terms of safety and/or the recipient, explore sensitivities in ethnic minority communities, and consider ways to normalize reporting sexual behaviour and the use of smartphones as a memory aid.

[1] This is joint work with Eamonn Ferguson (*School of Psychology, University of Nottingham; NIHR Blood and Transplant Research Unit*), Claire Lawrence (*LawrencePsychAdvisory*), Chris Starmer, Abigail Barr (*School of Economics, University of Nottingham*), Katy Davison, Claire Reynolds & Su Brailsford (*NHS Blood and Transplant/UK Health Security Agency Epidemiology Unit*). At the time of submitting this thesis, a version of the paper titled "*Communicating the move to individualized donor selection policy: Framing messages based on altruism and safety*" has been accepted into Transfusion which is an applied health journal. This work was funded by a grant from the UKFORUM to Eamonn Ferguson & Claire Lawrence. The views expressed in this paper are those of the authors and do not reflect any of the organizations or funder related to this paper.

3.1 Introduction

Internationally blood services have adopted population-based screening policies for men-who-have-sex-with-men (MSM), either permanently or temporarily deferring them (Fisayo, 2021; Goldman et al., 2018; Haire et al., 2018). However, accumulated evidence (Aubé et al., 2021; Caffrey et al., 2022; Davison et al., 2021; Germain, 2016) and improved Nucleic Acid Testing (Busch et al., 2005) indicate such policies require review to ensure that they are justifiable, fair, and equitable (Caruso et al., 2019; Karamitros et al., 2017; Kesby & Sothern, 2014). Instead, selecting people with high-risk sexual behaviour has been recommended (O'Brien et al., 2021; Sanquin, 2021; Wentz et al., 2019). In 2020, the FAIR (For the Assessment of Individual Risk) project delivered recommendations that the United Kingdom (UK) blood services supporting the replacement of time-based MSM deferrals with an individualized assessment of all donors on sexual behaviour and sexually transmitted infection history (FAIR Steering Group, 2020). This paper explores a resulting challenge of how best to frame communications about such a policy change to minimize potential negative consequences (e.g., putting people off donating).

Approach-Avoidance Framework

We adopt an approach-avoidance framework to assess blood donation motivation (Bach et al., 2014; Carver, 2006; Corr, 2013; Elliot, 1999; Gray & McNaughton, 2000; Hull, 1952; Loh et al., 2017; Mowrer, 2009; Sherman et al., 2006; Wright et al., 2013, 2012). Animal models of human anxiety often invoke a conflict to balance the desire to seek reward with the impulse to avoid harm (Bach et al., 2014; Loh et al., 2017). Under a reflexive "Pavlovian" approach-avoidance framework of motivation, behaviour (i.e., the act of donating blood) is associated with affective states that make it something people want to approach (attain) or avoid (Carver, 2006; Elliot, 1999; Gray & McNaughton, 2000; Hull, 1952; Wright et al., 2013, 2012). Motivation to approach behaviour is associated with positive affect (e.g., happiness) and avoidance with negative affect (e.g.,

shame) (Carver, 2006; Elliot, 1999; Gray & McNaughton, 2000; Hull, 1952). An individualized screening policy, based on sexual behaviour, may influence considerations of both approach (intentions to donate blood) and avoidance (put-off donating blood) (Ferguson & Chandler, 2005). Therefore, the question is: how best to frame communications to minimize avoidance and/or increase approach?

Framing the move to individualized screening approach: Risk and altruism

Framing risk: Under the precautionary principle, any change to donation policy must not increase potential harm (De Kort et al., 2016; Kramer et al., 2017; Mikkelsen et al., 2021; Timmermann, 2017). Here, potential harm equates to the probability (r) that an infectious donation is made within a non-detectable window period (the viral residual risk). Therefore, the expected value of a transfusion of donated blood is $(1-r)G+rB$, where $0 < r < 1$, G is the net benefit generated by a transfusion received safely, and B is the net cost generated by a transfusion leading to an infection. Under the precautionary principle, the change to an individualized screening policy must reduce, or at least not increase, viral residual risk ($r' \leq r$, where r' is the probability of an infectious donation after the change). The empirical question that we investigate is whether communications informing individuals about the effects of the policy change ($r' \leq r$) should make salient the higher likelihood of a good outcome (increased recipient safety: $(1-r')G \geq (1-r)G$) or the lower probability of a bad one (reduced recipient risk: $r'B \leq rB$).

Theory and evidence show that losses loom larger than equivalent gains, implying that focusing on reducing risk will be more effective than enhancing perceptions of safety (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991). Indeed, blood services have historically focused on risk (ABO, 2022). However, the way people process risk is influenced by heuristics and emotions (Loewenstein et al., 2001; Slovic et al., 2007; Tversky & Kahneman, 1974). For example, people have a systematic preference for certainty over uncertainty (Tversky & Kahneman, 1974). As, in common parlance, risk signals uncertainty, a message framed around risk will invoke uncertainty and,

thereby, induce avoidance. In terms of emotional processes, the risk-as-feelings hypothesis (Loewenstein et al., 2001) and the affect heuristic (Slovic et al., 2007) emphasize the key role of affective processes in risk perception. When an event (in this case, the policy) is viewed positively, then communicating gains (i.e., increased safety) will enhance the perceived benefits of the event. Evidence suggests people view the FAIR policy change as positive (FAIR Steering Group, 2020). Therefore, safety-based frames should be more effective than risk-based frames.

Altruism frames: Increasing focus on the well-being of others ('altruistic framing') increases the uptake of cooperative behaviours (Betsch et al., 2013, 2017; Brewer et al., 2017; Chapman et al., 2012; Isler et al., 2020; Li et al., 2016; Shim et al., 2012). Therefore, altruistically framed messages (e.g., focusing on the recipient) should motivate the approach towards the decision to donate.

Therefore, we expect altruistic and safety-based frames to be more effective than non-altruistic frames and non-safety-based frames at motivating approach consideration towards donating.

Reporting Sexual Behaviour: Approach-Avoidance Mechanisms

Primarily, such a policy should not put people off donating. However, when donating, donors must also be compliant with the selection criteria (Lau et al., 2021; O'Brien et al., 2019; Wentz et al., 2019). Thus, as a secondary aim, we explore the awareness of non-compliance factors in potential donors after the policy change. Understanding the salient non-compliance mechanisms may inform targeted strategies to reduce non-compliance. We examine three factors that may influence non-compliance. First, the anticipated shame/embarrassment of being asked about sexual behaviour (Haidt, 2003, 2007; Tangney et al., 2007). Second, non-compliance owing to error, namely that sexual behaviour was forgotten (McAuliffe et al., 2007). Finally, the perception that the questions are irrelevant because blood is tested (Cutts et al., 2021).

3.2 Methods

Sampling procedure

Stratified random sampling was employed (Figure 3.1), through Prolific, to over-sample LGBQ+ and ethnic minority populations.[2] Initially, a representative sample (age, gender, and ethnicity) of the UK population (n = 1495) was recruited, followed by additional samples of UK participants exclusively from ethnic minorities (n = 707) and LGBQ+ (n = 703) communities. All data was collected in February 2021.

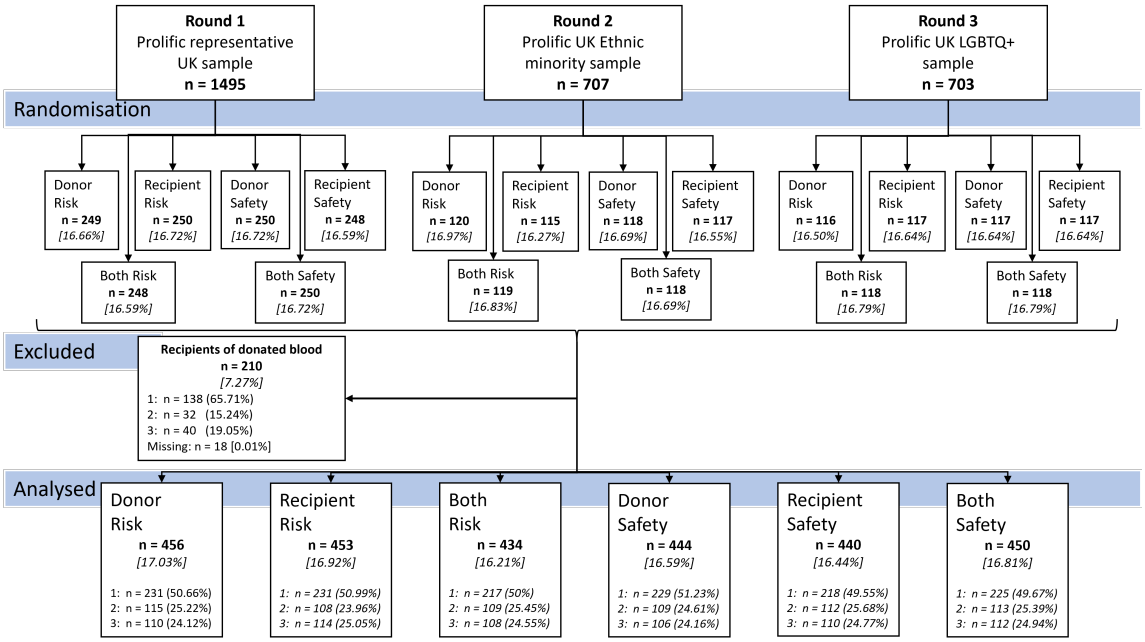


Figure 3.1: Sampling strategy

[2]The initialism LGBQ+ stands for Lesbian, Gay, Bisexual, Queer or Questioning, where the '+' represents those who are part of the community, but for whom LGBQ does not accurately capture or reflect their identity. We ask subjects about sexual orientation, but not gender identity beyond the binary. Therefore, the 'T' for Transgender identity that normally would be specified in the initialism is not present. This is not to say subjects coded as LGBQ+ were not transgender, just that we do not collect this data.

Design

Participants were randomly assigned to one of the six conditions formed by crossing the 2 risk frames (risk vs safety) with the 3 altruism frames (donor, the recipient or both) (Figure 3.2).[3]

[3]For full instructions see Appendix C.1.1.

The rules on blood donation in the UK will change in the summer of 2021. Previously, men-who-have-sex-with-men (MSM) were not able to donate blood unless they had abstained from sexual behaviour for 3 months. The new policy will mean that all potential donors, regardless of age, sex, sexuality, or ethnicity, will complete the same **[1]** before they donate.

The **[1]** covers a wide range of questions of health and travel to ensure that **[2]**. The **[1]** will now include several new questions on sexual behaviours. These questions are asked to further ensure that **[3]**.

Thus, donation decisions are based on donor's behaviour, not their sexuality.

[1]	REDUCE RISK	PROMOTE SAFETY
DONOR	'Donor Health Check'	'Donor Health Check'
RECIPIENT	'Recipient Risk Check'	'Recipient Safety Check'
DONOR-RECIPIENT	'Donor Health and Recipient Risk Check'	'Donor Health and Recipient Safety Check'

[2]	REDUCE RISK	PROMOTE SAFETY
DONOR	the donor is healthy to donate blood	the donor is healthy to donate blood
RECIPIENT	the blood is at low risk for the recipient to receive	the blood is safe for the recipient to receive
DONOR-RECIPIENT	both the donor is healthy to donate blood and blood is at low risk for the recipients to receive.	both the donor is healthy to donate blood and the blood is safe for recipients to receive

[3]	REDUCE RISK	PROMOTE SAFETY
DONOR	donors who are at low risk of infection can donate blood.	donors are safe to give blood.
RECIPIENT	any risk of infection to the recipients of blood is low.	recipients receive blood that is safe for use.
DONOR-RECIPIENT	donors who are at low risk of infection can donate blood so that any risk of infection to the recipients of blood is low.	donors are safe to give blood such that recipients receive blood that is safe for use.

Figure 3.2: Structure of the framed communications

Measures

Pre-manipulation measures

Demographics: We recorded age, gender identity (Female/ Male/ Gender non-conforming/ Other/ Prefer not to say), ethnicity across 18 ONS categories and sexual identity (Appendix C.2.1).

Blood donation history: Participants were asked whether they had ever donated blood, and if yes, whether this was in the UK, and the time since their last donation (less than a month ago/ 2 to 12 months ago/ 12 months to 2 years ago/ longer than 2 years ago/ cannot remember). Respondents were coded as non-donors, lapsed donors (blood donors who had not donated in the last 2 years), and current donors (donors who had donated within the last 2 years). Participants were also asked if they had ever been a recipient of blood or its components (Yes/No).

Post-manipulation measures

After reading the communication they had been assigned, participants answered the following questions.

Manipulation check: The focus of the statement was assessed with the question: "Who is the focus of the statement?" (from 0 = "the donor only", to 10 = "the recipient only"). The salience of the donor relative to the recipient in the statement was also assessed: "To what extent does the statement make you think about the patients who receive blood?" (from 1 = "Not at all" to 7 = "Completely").

Primary outcomes

Approach (intentions) and Avoidance (deterrence): Approach was assessed by the sum of two yes/no items: (i) Do you plan to donate blood in the near future? and (ii) Would you be willing to donate blood? Avoidance was assessed using two items: (i) To what extent would the statement put you off donating blood (self-deter) and (ii) To what extent do you think the statement would put others off from donating blood? (other-deter) (from 1 = "Not at all" to 7 = "Completely"). These indices of approach and avoidance were normalized between 0 and 1 (Appendix C.2.2).

Approach-Avoidance Index (AAI): A strength of approach-avoidance index (AAI)

was constructed ranging from -1, strong motivation towards avoidance, 0, equal approach and avoidance, and 1, strong motivation towards approach. Two AAIs were constructed: (i) self-AAI based on the normalized approach index minus the normalized self-deter index and (ii) normative-AAI based on the normalized approach index minus the normalized sum of the self-deter and other-deter indices. Consistent with the normative principle that people are less likely to act if they think others will not act, we sum self- and other-deter as they are highly correlated ($\rho = 0.508, p < 0.001$) (Cutts et al., 2021). (See Appendix C.2.2 for the formulae used).

Mechanisms driving possible under-reporting: To assess the salience of non-compliance factors, indirect questioning is employed to reduce socially desirable responding. Indirect questioning is often used to mitigate social desirability bias (Fisher, 1993). The main idea is that respondents may find it easier to express their opinions on a sensitive issue in response to an indirect question (Fisher & Tellis, 1998; Jo et al., 1997). Therefore, all participants saw the following stem: "To what extent do you think each of the following factors influences how accurately people report on their sexual behaviour over the last 3 months?" (1 = "Not at all" to 7 = "Completely"): (i) they had forgotten aspects of their previous sexual behaviour, (ii) feeling embarrassed to report on their sexual behaviour, (iii) feeling ashamed to report on their sexual behaviour and (iv) feeling that the questions are not relevant as all blood is tested anyway and so decide not to report their sexual behaviour. A negative emotions score was calculated as the average response of feeling embarrassed and ashamed ($\rho = 0.794, p < 0.001$).

Secondary outcomes

Also, as secondary outcomes, we assessed beliefs about the blood supply (safety), the screening process (fairness and equality) and the use of smartphones as a memory aid (Appendix C.2.3).

Statistical analysis strategy

All analyses were conducted in Stata 17 and SPSS 27. All p-values are two-tailed. Seven percent (n = 210) of the sample reported that they had received blood and were

excluded from the analysis. The results were not sensitive to the exclusion of recipients.

3.3 Results

Sample characteristics

Sample characteristics are summarised in Table 3.1. For the regression analysis, a single category, LGBQ+, was created encompassing Lesbian, Gay, Bisexual, Queer, Pansexual, Bi-curious and Asexual: n = 788. Balance tests confirm randomization (Appendix C.2.4).

Examining Table 3.2, intentions to donate blood were high ($M = 1.42, SD = 0.68$: normalized $M = 0.71$), and self-deter ($M = 1.99, SD = 1.46$, normalised $M = 0.17$) and normative-deter ($M = 3.01, SD = 1.58$, normalised $M = 0.33$) were low. AAI scores were positive indicating feelings of approach towards donation dominated (self-AAI: $M = 0.55, SD = 0.45$; normative-AAI: $M = 0.46, SD = 0.43$). Perceived safety ($M = 11.21, SD = 2.68$) and fairness ($M = 24.27, SD = 3.30$) were high and significantly associated with greater approach and lower avoidance. Anticipated negative emotions (shame/embarrassment) were the most likely to be seen to influence the under-reporting of sexual behaviour ($M = 5.36, SD = 1.23$), followed by the perceived irrelevance ($M = 4.98, SD = 1.57$), and forgetting ($M = 4.03, SD = 1.58$). Awareness of all three mechanisms of under-reporting was weakly positively correlated with both self- and other-avoidance. Awareness of forgetting as a mechanism of under-reporting was positively associated with the appreciation of smartphones as an effective memory aid.

	N	%	*Age*	*%Male*	*%Asian*	*%Black*	*%Mixed*	*%White*
Blood donation history								
Non-donor	1755	65.56	34.74	0.40	0.09	0.19	0.08	0.62
Lapsed donor	600	22.41	47.44	0.44	0.04	0.11	0.04	0.78
Current donor	315	11.77	36.43	0.43	0.08	0.09	0.05	0.76
Prefer not to say	7	0.26	44.43	0.43	0.13	0.25	0.13	0.31
Sexual orientation								
Asexual	50	1.87	35.06	0.32	0.04	0.12		0.84
Bisexual	366	13.67	29.26	0.21	0.07	0.06	0.02	0.83
Gay	118	4.41	36.14	0.91	0.07	0.05	0.01	0.87
Heterosexual/straight	1882	70.30	40.63	0.46	0.08	0.20	0.08	0.61
Lesbian	93	3.47	30.68		0.05		0.01	0.94
Queer	31	1.16	29.52	0.13	0.03	0.06	0.03	0.87
Pansexual	48	1.79	26.06	0.17	0.06	0.08	0.04	0.79
Bi-curious	31	1.16	27.23	0.32	0.16	0.16	0.06	0.55
Prefer not to say	58	2.17	35.55	0.24	0.16	0.14	0.04	0.63
Total	**2677**	**100**	**37.80**	**0.42**	**0.16**	**0.07**	**0.08**	**0.68**

Note: Non-donor = never donated, Lapsed = donated more than 2 years ago, Current = donated within the last 2 years. Asexual = people who self-identify as asexual, Bisexual = people who self-identify as bisexual, Gay = people who self-identify as gay, Straight = people who self-identify as straight, Lesbian = people who self-identify as lesbian, Queer = people who self-identify as queer, Pansexual = people who self-identify as pansexual, Bi-curious = people who self-identify as bi-curious. Asian: people from Asian ethnic communities, Black: people from Black ethnic communities, Mixed: people from Mixed ethnic communities, White: people from White ethnic communities.

Table 3.1: Analysed sample characteristics

	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Approach (intentions)	1.42	0.68	1.00									
2. Self-avoidance (self-deter)	1.99	1.46	−0.17 ***	1.00								
3. Other-avoidance (other-deter)	3.01	1.58	−0.07 ***	0.51***	1.00							
4. Normative-AAI	0.46	0.43	0.86***	−0.57 ***	−0.50 ***	1.00						
5. Safety	11.21	2.68	0.24***	−0.36 ***	−0.19 ***	0.35***	1.00					
6. Fairness	24.27	3.30	0.13***	−0.39 ***	−0.33 ***	0.30***	0.44***	1.00				
7. Forgetting	4.03	1.58	0.01	0.10***	0.11***	−0.05 **	−0.08 ***	0.00	1.00			
8. Anticipated negative emotion	5.36	1.23	−0.02	0.08***	0.13***	−0.08 ***	−0.16 ***	−0.06 *	0.24***	1.00		
9. Perceived irrelevance	4.98	1.57	−0.05 **	0.11***	0.11***	−0.10 ***	−0.16 ***	−0.14 ***	0.19***	0.42***	1.00	
10. Effectiveness of smartphones	3.49	1.74	0.11***	0.00	0.01	0.09***	0.13***	0.14***	0.10***	−0.08 ***	−0.12 ***	1.00

Note: Table displays pairwise Pearson correlation coefficients * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.2: Matrix summarising outcome variable sample means, standard deviations, and pairwise correlation coefficients with significance

Manipulation checks for effects of frame and focus: patient or donor focus

The perceived patient focus was higher in the altruism frames, particularly for the combined donor-recipient message (Appendix C.2.5).

Frames and approach-avoidance considerations

Analysis of normalized approach (intentions) (Table 3.3, Model 1) indicates no framing effects. However, a higher approach was observed for lapsed and current donors and younger participants. A lower approach was observed in LGBQ+ respondents.

Exposure to a safety frame, compared to a risk frame, or a recipient frame compared to a donor frame when the risk frame is applied reduced (i) self-avoidance (Model 2), (ii) other-avoidance (Model 3) and (iii) normative-avoidance (Model 4). There was also a significant interaction between risk and altruism frames on self-avoidance (Model 2), other-avoidance (Model 3) and normative-AAI (Model 4). Examining the margins for these interactions indicates that the highest self-avoidance occurred for a combination of risk and donor frames (Appendix C.2.6).

Self-avoidance was greater for men and people from Asian, Black, and Mixed ethnic communities relative to people from White communities and lower for people from LGBQ+ communities and both lapsed and current donors (Model 2). The demographic differences for other-avoidance are similar (Model 3). Except we observe insignificant effects of gender, being from Mixed ethnicity, and being a current donor.

Analysis of self-AAI and normative-AAI scores (Models 5 & 6) indicate that exposure to a recipient-frame, compared to a donor-frame, reduced avoidance relative to approach.

	(1) Normalised approach		(2) Normalised self-avoidance		(3) Normalised other-avoidance		(4) Normalised normative-avoidance		(5) Self-AAI		(6) Normative-AAI	
Risk												
Safety	0.003	[-0.040,0.045]	**-0.044****	[-0.075,-0.012]	**-0.048****	[-0.083,-0.013]	**-0.046****	[-0.074,-0.017]	0.046	[-0.011,0.103]	0.049	[-0.006,0.103]
Altruism												
Recipient	0.023	[-0.020,0.065]	**-0.037***	[-0.068,-0.006]	**-0.035***	[-0.070,-0.001]	**-0.036***	[-0.064,-0.008]	**0.059***	[0.003,0.116]	**0.058***	[0.004,0.112]
Both	0.011	[-0.032,0.054]	−0.021	[-0.052,0.011]	−0.025	[-0.060,0.010]	−0.023	[-0.051,0.006]	0.032	[-0.025,0.090]	0.034	[-0.021,0.089]
Risk ×Altruism												
Safety × Recipient	0.000	[-0.060,0.061]	**0.047***	[0.003,0.092]	**0.068****	[0.019,0.117]	**0.058****	[0.017,0.098]	−0.047	[-0.128,0.034]	−0.057	[-0.134,0.021]
Safety × Both	−0.023	[-0.083,0.038]	0.034	[-0.011,0.078]	0.049	[-0.000,0.099]	**0.041***	[0.001,0.082]	−0.056	[-0.137,0.025]	−0.064	[-0.142,0.014]
Age	**-0.007*****	[-0.008,-0.006]	0.000	[-0.000,0.001]	0.000	[-0.001,0.001]	0.000	[-0.001,0.001]	**-0.007*****	[-0.008,-0.006]	**-0.007*****	[-0.008,-0.005]
Male	−0.016	[-0.041,0.009]	**0.042*****	[0.023,0.061]	0.011	[-0.010,0.031]	**0.026****	[0.009,0.043]	**-0.058*****	[-0.092,-0.024]	**-0.042***	[-0.075,-0.010]
LGBQ+	**-0.037***	[-0.067,-0.006]	**-0.025***	[-0.048,-0.003]	**-0.026***	[-0.051,-0.001]	**-0.026***	[-0.046,-0.005]	−0.011	[-0.052,0.030]	−0.010	[-0.050,0.029]
Ethnicity												
Asian	−0.020	[-0.057,0.016]	**0.129*****	[0.102,0.157]	**0.084*****	[0.054,0.114]	**0.107*****	[0.082,0.132]	**-0.150*****	[-0.199,-0.100]	**-0.127*****	[-0.174,-0.080]
Black	0.003	[-0.048,0.055]	**0.082*****	[0.044,0.120]	**0.062****	[0.020,0.105]	**0.072*****	[0.037,0.107]	**-0.078***	[-0.148,-0.008]	**-0.068***	[-0.135,-0.002]
Mixed	**-0.053***	[-0.100,-0.005]	**0.061*****	[0.027,0.096]	0.011	[-0.027,0.050]	**0.036***	[0.005,0.068]	**-0.114*****	[-0.177,-0.050]	**-0.088****	[-0.149,-0.028]
Donor status												
Lapsed donor	**0.137*****	[0.105,0.169]	**-0.069*****	[-0.093,-0.046]	**-0.034***	[-0.060,-0.007]	**-0.052*****	[-0.073,-0.030]	**0.207*****	[0.164,0.250]	**0.189*****	[0.148,0.230]
Current donor	**0.301*****	[0.262,0.340]	**-0.043****	[-0.072,-0.014]	−0.003	[-0.035,0.029]	−0.023	[-0.049,0.003]	**0.344*****	[0.292,0.397]	**0.324*****	[0.274,0.375]
Constant	0.003	[-0.040,0.045]	−0.044**	[-0.075,-0.012]	−0.048**	[-0.083,-0.013]	0.256***	[0.219,0.292]	0.756***	[0.682,0.829]	0.656***	[0.586,0.727]
R^2	0.145		0.079		0.029		0.064		0.121		0.115	
N	2552		2552		2552		2551		2551		2550	

Note: * p < 0.05, ** p < 0.01, *** p < 0.001. Baseline groups: Risk = Risk frame; Altruism = donor frame; Ethnicity = people from white ethnic backgrounds; Donor status = non-donors. Variable definitions: Self-AAI = approach-avoidance index based on normalized approach minus normalized self-avoidance; Normative-AAI = approach-avoidance index based on normalized approach minus normalized self plus other avoidance. Coefficients are unstandardized.

Table 3.3: OLS regression for normative intentions, normative-self-deferral, normative-other-deferral, normative self-other deferral, self-approach-avoidance score, and approach-avoidance score

Awareness of mechanisms of non-compliance

There were no framing effects on awareness of mechanisms that may lead to under-reporting of sexual behaviour (Appendix C.2.7). However, relative to women, men reported lower awareness of all mechanisms, younger respondents reported greater awareness of forgetting and negative emotions, and current donors reported lower awareness of negative emotions as mechanisms leading to under-reporting.

Relative to those from White communities, people from Black communities reported greater awareness of all three mechanisms, and people from Asian communities reported greater awareness of forgetting leading to under-reporting.

Effects of frames on perceived safety, fairness and equality

There were no significant framing effects on perceived safety or fairness. However, perceived safety was lower for older, Asian, Black and Mixed ethnicity participants and higher in lapsed and current donors and LGBQ+ respondents (Appendix C.2.8).

3.4 Discussion

We have explored how best to frame a policy change involving a shift towards individualized risk assessment of donors' sexual behaviour and infection history. Frames focused on increasing safety (rather than reducing risk) and/or the recipient (rather than the donor) decreased participants' likelihood of being put off (avoiding) donating. Furthermore, we show for the first time that people are aware of mechanisms associated with the under-reporting of sexual behaviour. Theoretical and practical implications are discussed below.

Altruism and safety frames

Our findings support and are consistent with growing evidence suggesting altruism frames enhance health behaviors requiring cooperation (Brewer et al., 2017; Chapman et al., 2012). We argue that safety-based frames would be most effective in commu-

nication as the policy change is viewed positively, and our findings support this (FAIR Steering Group, 2020).

One aim of the paper was to identify a frame that will reduce the extent to which potential donors feel put off from donating. Consistent with evidence that cooperative behaviour is associated with normative considerations (Fischbacher et al., 2001) we find both the degree respondents report to personally, and that others would be put off is reduced by altruism and safety-based frames. The fact that both feelings of self and other avoidance respond to framing effects could be evidence of a desire for consistency (both in terms of response bias and normative conformity), or heuristic thinking (i.e., assuming others will think as you do). On the other hand, feelings that others may be put off from donating increase the perceived cost (or reduced benefit) of donating. When deciding to donate, evidence suggests donors will weigh up the costs and benefits of donating (Ferguson & Chandler, 2005). As such, a safety-based and altruism frame may reduce feelings of self-avoidance (and therefore the perceived net cost of donating) via the anticipated reduced negative reaction of others.

Therefore, we argue that safety-based and recipient-focused frames may have a wider positive social impact and these findings suggest one novel route through which safety and altruism frames influence personal behaviour by influencing normative expectancies.

Donor status, ethnicity and avoidance

Participants who were current donors, lapsed donors, or LGBQ+ reported lower avoidance. Thus, the number of current active donors should not reduce under this policy. However, people from Asian, Black and Mixed ethnic communities were more likely to be put off. People from ethnic minority communities are, in general, less likely to donate (Boulware et al., 2002; Ferguson et al., 2022; Josephson et al., 2007; Shaz et al., 2008). Therefore, it is of concern that the policy change is linked to greater donation avoidance in these communities.

Awareness of mechanisms for under-reporting of sexual behaviour

People were aware of all three mechanisms linked to under-reporting of sexual behaviour: (i) feeling embarrassed (Tangney et al., 2007), (ii) under-reporting in error (forgetting) (McAuliffe et al., 2007) or (iii) answering perceived irrelevant questions(Cutts et al., 2021). Embarrassment was rated the most likely of the mechanisms to lead to non-compliance, followed by forgetting and irrelevance of the questions. Thus, non-compliance interventions that normalize reporting sexual behaviours - based on descriptive norms (e.g., these are normal behaviours that everyone reports) - may be most effective. Also, interventions to enhance memory at donor screening may help to address poor recall of sexual behaviours. Indeed, the use of smartphones was perceived as an effective and appropriate method to enhance recall. Awareness was not influenced by frames but varied by demographics. For example, people from Black or Asian ethnic communities were more likely to report greater awareness of non-compliance factors.

Implications for Transfusion Services

There are a number of clear implications for transfusion services. First, as more transfusion services adopt a donor eligibility policy based on individual sexual behaviour and sexual health, they should be greater consideration around the framing of communications in terms of safety and/or the recipient (Canadian Blood Services, 2022). Also, blood services should consider that people from ethnic minority communities are more likely to indicate being deterred by such a policy. Therefore, targeted work to understand attitudes and identify strategies to encourage donations in ethnic minority groups is critical (Ferguson et al., 2022) to ensure donor pool diversity and improve donor-recipient matching for more effective treatments (Boulware et al., 2002; Josephson et al., 2007; Shaz et al., 2008). Future research should focus on developing strategies to reduce embarrassment (normalizing sexual behaviour questions), campaigns highlighting the importance of donor compliance, and memory aids (smartphones) at screening should be trialled.

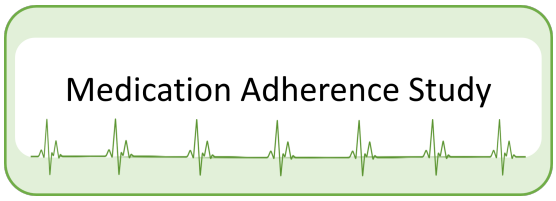
Appendix A

Appendix: Chapter 1

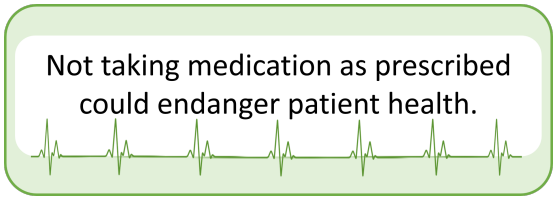
A.1 Design

A.1.1 Messages

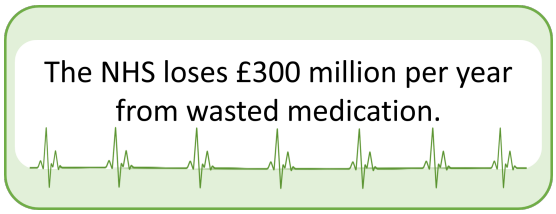
Control message:



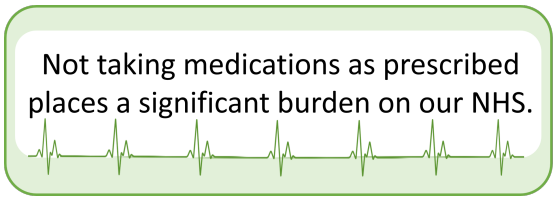
Health message:



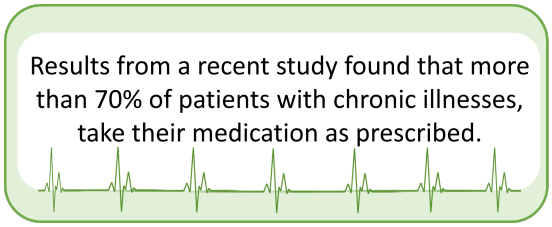
£-NHS message:



Burden-NHS message:



Descriptive norm message:



A.1.2 Experimental instructions

Instructions were encountered by all subjects except when stated otherwise. Here, we report exact question wording and response fields. The questionnaire was programmed in Qualtrics. The experiments in Study 1, 2, and 3 follow the same structure. Differences between the experiements are highlighted. All programmes are available upon request. For an overview of the question structure:

Q1 Informed consent

Q2 - 6 Demographics

Q7 Currently taking medication for long-term illness screener

Q8 Illness group

Q9 - 13 Self-report using the MARS-5 (Chan et al., 2020)

Q14 - 24 Medication regime characteristics

Q25 - 27 Message evaluation questions (for questions included in Study 3 see Appendix B.1.1)

Q28 - 29 Personality scales

Q30 - 31 Additional control questions

[Q1]

Welcome to the research study!

We are interested in understanding medication adherence. You will be presented with information relevant to medication adherence and asked to answer some questions about it. Please be assured that your responses will be kept completely confidential.

The study should take you around 12-15 minutes to complete, and you will receive £4 for your participation. Your participation in this research is voluntary. You have the right to withdraw at any point during the study, for any reason, and without any prejudice. If you would like to contact the Principal Investigator in the study to discuss this research, please e-mail Sarah at sarah.bowen@nottingham.ac.uk.
By clicking the button below, you acknowledge that your participation in the study is voluntary and that you are aware that you may choose to terminate your participation in the study at any time and for any reason. All your responses will be kept anonymous.
Please note that we follow the established rules of the Centre for Decision Research and Experimental Economics (CeDEx) at the University of Nottingham and therefore will never employ deception in any of our studies.
This survey will be best displayed on a laptop or desktop computer. Some features may be less compatible for use on a mobile device.

□ I have read and understand the explanations and I voluntarily consent to participate in this study.

Please enter your Prolific ID here: [**Text input**]

Before we get started, we would like to know some information about you.
[Q2] Age [**Numeric input**]
[Q3] Gender [**Text input**]
[Q4] What is the highest qualification or level of schooling you have completed?

- GCSEs or equivalent
- Apprenticeship
- A Levels or equivalent
- Undergraduate Degree (for example BA, BSc)
- Masters Degree (for example MA, MSc)
- Professional Qualification (for example teaching, nursing, accountancy)
- PhD
- Other vocations/work-related qualifications
- No Qualifications
- Other

[Q5] Think of this ladder as a representation of where people stand in the United Kingdom. At the top of the ladder are the people who are the best off – those who have the most money, the most education, and the most respected jobs. At the bottom are the people who are the worst off – who have the least money, least education, and the least respected jobs or no job. The higher up you are on this ladder the closer you

are to the people at the very top; the lower you are, the closer you are to the people at the very bottom. Where would you place yourself on this ladder?

- Top rung - 10
- 9
- 8
- 7
- 6
- 5
- 4
- 3
- 2
- Bottom rung - 1

[Q6] Information about income is important to understand how people are doing financially these days. Your answers are confidential. What is your best guess of the total yearly income of all the members of your household living with you in 2018, before taxes? This figure should include income from all sources, including salaries, wages, benefits, pensions, dividends, interest, and all other income.

- Up to £9,999
- £10,000 and up to £24,999
- £25,000 and up to £49,999
- £50,000 and up to £99,999
- £100,000 and above

[**In Study 3, this question was asked during the pre-screening stage, which was a survey subjects had to complete before being invited to the main experiment in Study 3. See Appendix A.1.5 for the experimental instructions and questions asked during the pre-screening stage.**]

[Q7] Do you take at least one prescribed course of medication to control a long-term chronic health condition such as diabetes, heart disease, stroke, high blood pressure etc.?

- Yes
- No
- Prefer not to say

[**Respondents responding "Yes" proceeded onto the next stage. Otherwise, in Study 1, respondents were directed to an alternative hypothetical version of the experiment (see Appendix A.1.3), before returning to the main survey flow at [Q25]. Alternatively, in Study 2, respondents were removed from the study and shown the following message:**

I'm sorry, but you are ineligible for this study. We thank you for your interest in our survey! Please return your submission on Prolific by selecting the 'Stop without completing' button and you will be rewarded with a goodwill payment of £0.40.]

[**In Study 1 and 2 all subjects answered [Q8]. In Study 3 a more expansive list of chronic illnesses was presented to subjects as shown in [Q8*]**]

[Q8] From the list below, please select ONE chronic health condition/course of treatment that applies to you, even if you have more than one chronic condition. If you select "Other", then please write down ONE chronic health condition that you are currently taking medication for in the text box.

- Hypertension (High Blood Pressure)
- Heart Disease (Antiplatelets)
- Anticoagulant therapy (Blood clots, strokes, heart attacks)
- Asthma or COPD
- Type 2 Diabetes
- Other [**Text input**]
- Prefer not to say

[Q8*] In this survey, for some questions we will ask you to think about a course of medication you take regularly for the treatment of a chronic condition. It is important that you only think about a prescribed course of medication to control ONE chronic condition. In the dropdown list below, we present a non-exhaustive list of chronic conditions. Please indicate a chronic condition you will think of, when answering questions about your medication taking behaviour. Your response is confidential and anonymous, and your answer will not affect the rest of the survey. If your chronic condition is not in the drop down list, please select "other", you will have a chance to write in your condition if it is not included in the list.

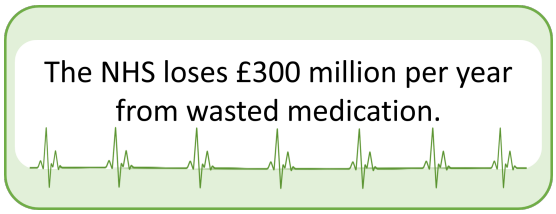
- Alzheimer disease and dementia
- Arthritis
- Asthma
- Cancer
- Cardiovascular disease
- Chronic Kidney Disease
- COPD

- Cystic fibrosis
- Diabetes
- Endometriosis
- Epilepsy
- Fibromyalgia
- HIV/AIDS
- Hypertension
- Inflammatory bowel disease (ulcerative colitis, Crohn's disease, etc.)
- Mood disorders (bipolar, cyclothymic, and depression)
- Motor Neurone Disease
- Multiple sclerosis
- Osteoporosis
- Parkinson disease
- Stroke
- Other [**Text input**]
- Prefer not to say

[**Text highlighted in red was seen only by subjects who completed the TF-self-report.**]

For one reason or another, many people can't or don't always take medication as prescribed. We want to know what barriers to medication adherence people do, or might expect to, experience.

This research is important because:



[**The message shown to subjects depended on message treatment condition. This is an example of what subjects assigned to the £-Burden condition saw during the experiment. For all messages used in the experiment see A.1.1.**]

On the next screens, we will ask you 5 questions about your medication adherence [**in the last two weeks**]. When thinking about your answers, please keep the medication used to control the health condition you disclosed on the previous screen in mind, and answer the questions to the best of your ability.

[**Subjects saw the message again above each MARS-5 item [Q9 - 13]**.]

[Q9] I forget to take my medication [**OR** In the last two weeks, I forgot to take my medication]

- Never
- Sometimes
- About half the time
- Most of the time
- Always

[Q10] I change the dosage of my medication [**OR** In the last two weeks, I changed the dosage of my medication]

- Never
- Sometimes
- About half the time
- Most of the time
- Always

[Q11] I stop taking my medication for a while [**OR** In the last two weeks, I stopped taking my medication for a while]

- Never
- Sometimes
- About half the time
- Most of the time
- Always

[Q12] I decide to skip one of my medication dosages [**OR** In the last two weeks, I skipped one of my medication dosages]

- Never
- Sometimes
- About half the time
- Most of the time

- Always

[Q13] I use my medication less than is prescribed [**OR** In the last two weeks, I used my medication less than prescribed]

- Never
- Sometimes
- About half the time
- Most of the time
- Always

[**At this point in the experiment in Study 2, for each question [Q9 - 13] a subject reported to at least "sometimes" not adhere to their medications, a follow up question about the intentionality of their non-adherence was asked (see A.1.4)**.]

[Q14] Consider the medication(s) we have asked you to think about when answering questions. How long have you been taking the course of medication(s) prescribed for your condition?

- 0 - 3 months
- 4 - 6 months
- 7 - 11 months
- 1 - 3 years
- 3+ years
- Prefer not to say

[Q15] Are you prescribed more than one type of medication to control a chronic illness?

- Yes (if Yes, then how many are you prescribed: [**Numeric input**])
- No
- Prefer not to say

[Q16] How many times a day do you take one or more of your medications? [**Text input**])

[Q17] In your opinion, how complex is your medication regime?

- Extremely simple
- Moderately simple
- Neither simple nor complex
- Moderately complex
- Extremely complex

[Q18] How important is strictly adhering to your medication to maintaining your health?

- Not at all important - I can miss a couple of doses and it doesn't make much difference to my health
- Moderately important - It is ideal if I never miss a dose, but it is not the end of the world if it happens
- Extremely important - It is vital that I adhere to my medication regime exactly as prescribed to maintain a good level of health

[Q19] How do you access health care services in the UK

- Public (NHS) only
- Private only
- Public (NHS) and Private

[**Only subjects in Study 1 and 2 answered question [Q20 - 21]. Subjects in Study 3 instead answered [Q22-24]**.]

[Q20] On average, how often do you find yourself travelling away from home?

- Never
- Once in a year
- Once in 6 months
- Once in 3 months
- Every month
- Every week

[Q21] To what extent do you agree or disagree with the following statement: Travelling away from home makes adhering to my medications significantly more difficult.

- Strongly disagree
- Slightly disagree
- Neither agree nor disagree
- Slightly agree
- Strongly agree

[Q22] How has the coronavirus pandemic impacted your ability to take your medication exactly as prescribed?

- Reduced ability
- Not at all
- Increased ability
- Prefer not to say

[Q23] How has the coronavirus pandemic impacted your motivation to take your medication exactly as prescribed?

- Reduced motivation
- Not at all
- Increased motivation
- Prefer not to say

[Q24] Is your medication immunosuppressive?

- Yes
- No
- Prefer not to say

[**In Study 1, subjects completed [Q25] and [Q26]. In Study 2, subjects completed these questions plus [Q27]. In Study 3 respondents completed different message evaluation questions (see Appendix B.1.1) immediately after the self-report [Q9 - 13].**]

[Q25] Consider this message [**message shown here**], and then please answer the questions below:

[Q25.i] This message is trying to manipulate me

- Strongly disagree
- Slightly disagree
- Neither agree nor disagree
- Slightly agree
- Strongly agree

[Q25.ii] The information in this message is overblown

- Strongly disagree
- Slightly disagree
- Neither agree nor disagree
- Slightly agree
- Strongly agree

[Q25.iii] This message annoys me

- Strongly disagree
- Slightly disagree
- Neither agree nor disagree

- Slightly agree
- Strongly agree

[Q26] Local community pharmacies are considering producing stickers with the following message printed on them: [**message shown here**]. These stickers could be used by patients as a device to remember to take medications as prescribed. Reminders could be posted in multiple places around the house such as, the Kitchen fridge, Bathroom mirror, Desktop computer, and Bedside cabinet.
[Q26.i] Would you be willing to buy a set of these stickers?

- Yes (If you answered Yes, then what is the maximum you would be willing to pay for a set of 10 stickers (in GBP £)?: [**Numeric input**])
- No

[**In Study 2:** Q27] Suppose you were given a set of these stickers. How likely would you be to use them to help take your medication properly (i.e. use them to help you remember to take a scheduled dose)?

- Extremely likely
- Somewhat likely
- Neither likely nor unlikely
- Somewhat unlikely
- Extremely likely

[**In Study 3, subjects completed the shortened version of this scale. The items included in Study 3 are indicated with an asterix below.**]
[Q28] Please rate the extent to which each statement is true of you on a scale of 1 to 7 (where 1 = Not True and 7 = Very True). Please indicate your answer by choosing a number between 1 and 7.

- My first impressions of people usually turn out to be right
- It would be hard for me to break any of my bad habits
- I don't care to know what other people really think of me
- ∗ I have not always been honest with myself
- ∗ I always know why I like things
- When my emotions are aroused, it biases my thinking
- Once I've made up my mind, other people can seldom change my opinion
- I am not a safe driver when I exceed the speed limit
- I am fully in control of my own fate

* It's hard for me to shut off a disturbing thought
* I never regret my decisions
* I sometimes lose out on things because I can't make up my mind soon enough
- The reason I vote is because my vote can make a difference
- My parents were not always fair when they punished me
* I am a completely rational person
- I rarely appreciate criticism
* I am very confident of my judgments
* I have sometimes doubted my ability as a lover
- It's all right with me if some people happen to dislike me
- I don't always know the reasons why I do the things I do
* I sometimes tell lies if I have to
* I never cover up my mistakes
* There have been occasions when I have taken advantage of someone
- I never swear
* I sometimes try to get even rather than forgive and forget
- I always obey laws, even if I'm unlikely to get caught
* I have said something bad about a friend behind his/her back
* When I hear people talking privately, I avoid listening
- I have received too much change from a salesperson without telling him or her
- I always declare everything at customs
- When I was young I sometimes stole things
- I have never dropped litter on the street
- I sometimes drive faster than the speed limit
- I never visit sexy websites
- I have done things that I don't tell other people about
* I never take things that don't belong to me
- I have taken sick-leave from work or school even though I wasn't really sick
- I have never damaged a library book or store merchandise without reporting it

- I have some pretty awful habits
- ∗ I don't gossip about other people's business

[**In Study 1, subjects completed a general locus of control scale, in Study 2 a health-domain specific locus of control scale was implemented**]

[**If Study 1:** Q29] For each question select the statement (a or b) that you agree with the most.

i.a Many of the unhappy things in people's lives are partly due to bad luck.

i.b People's misfortunes result from the mistakes they make

ii.a One of the major reasons why we have wars is because people don't take enough interest in politics

ii.b There will always be wars, no matter how hard people try to prevent them.

iii.a In the long run, people get the respect they deserve in this world

iii.b Unfortunately, an individual's worth often passes unrecognised no matter how hard he tries

iv.a The idea that teachers are unfair to students is nonsense

iv.b Most students don't realise the extent to which their grades are influenced by accidental happenings

v.a Without the right breaks, one cannot be an effective leader

v.b Capable people who fail to become leaders have not taken advantage of their opportunities

vi.a No matter how hard you try, some people just don't like you

vi.b People who can't get others to like them don't understand how to get along with others

vii.a I have often found that what is going to happen will happen

vii.b Trusting to fate has never turned out as well for me as making a decision to take a definite course of action

viii.a In the case of the well-prepared student, there is rarely, if ever, such a thing as an unfair test.

viii.b Many times exam questions tend to be so unrelated to course work that studying is really useless

ix.a Becoming a success is a matter of hard work; luck has little or nothing to do with it.

ix.b Getting a good job depends mainly on being in the right place at the right time

x.a The average citizen can have an influence on government decisions

x.b This world is run by the few people in power, and there is not much the little guy can do about it.

xi.a When I make plans, I am almost certain that I can make them work

xi.b It is not always wise to plan too far ahead because many things turn out to be a matter of luck anyway

xii.a In my case, getting what I want has little or nothing to do with luck

xii.b Many times we might just as well decide what to do by flipping a coin

xiii.a What happens to me is my own doing

xiii.b Sometimes I feel that I don't have enough control over the direction my life is taking.

[**If Study 2:** Q29] To what extent do you agree with the following statements on a scale from 1 "Strongly agree"' to 7 "Strongly disagree"?

- If I get sick, it is my own behaviour, which determines how soon I get well again
- No matter what I do, if I am going to get sick, I will get sick
- Having regular contact with my physician is the best way for me to avoid illness
- Most things that affect my health happen to me by accident
- Whenever I don't feel well, I should consult a medically trained professional
- I am in control of my health
- My family has a lot to do with my becoming sick or staying healthy
- When I get sick, I am to blame
- Luck plays a big part in determining how soon I will recover from an illness
- Health professionals control my health
- My good health is largely a matter of good fortune
- The main thing, which affects my health, is what I myself do
- If I take care of myself, I can avoid illness
- When I recover from an illness, it's usually because other people (for example, doctors, nurses, family, friends) have been taking good care of me
- No matter what I do, I am likely to get sick
- If it's meant to be, I will stay healthy
- If I take the right actions, I can stay healthy
- Regarding my health, I can only do what my doctor tells me to do

---

[**The questions [Q30] and [Q31] were asked in both Study 1 and 2, with the wording changed slightly in Study 2 to account for the fact that the study was run in 2019, a year after Study 1. Wording specific to Study 1 is shown in round parentheses, and to Study 2, square parentheses**]

[Q30] These are the last questions of the survey! Thank you for your participation. The NHS has recently celebrated a landmark anniversary (this year) [in 2018]. Do you know how old the NHS turned (this) [last] year?

- Yes
- No

[**If Q30 = Yes**] [Q30.i] (How old do you believe the NHS is currently?) [Which landmark anniversary did the NHS celebrate in 2018?]

- 60
- 70
- 80
- 90
- 100

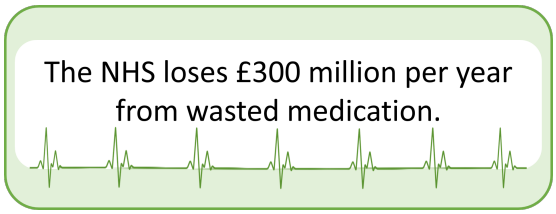
[Q31] With health services facing unprecedented financial and operational pressures, the future of the NHS has risen to the top of the political agenda. In March 2017, the Prime Minister announced that a new long-term plan for the NHS will be published, backed by a multi-year funding settlement. To what extent do you agree with increasing the proportion of UK government spending allocated to the NHS?

- Strongly disagree
- Slightly disagree
- Neither agree nor disagree
- Slightly agree
- Strongly agree

A.1.3 Study 1: Hypothetical version

[**In Study 1, subjects who reported not to be taking medication prescribed for a long-term illness (for exact question wording see [Q7], completed a hypothetical version of experiment. In this section, we summarise the question wording and response modes for the hypothetical version of the self-report.**] For one reason or another, many people can't or don't always take medication as prescribed. We want to know what barriers to medication adherence people do, or might expect to, experience.

This research is important because...



Consider the following scenario...

Today, you have been diagnosed with high blood pressure. You may need to take blood pressure medication for the rest of your life. But your doctor might be able to reduce or stop your treatment if your blood pressure stays under control for several years. It's really important to take your medications as directed. If you miss doses, it won't work as effectively. The medication won't necessarily make you feel any different, but this doesn't mean it's not working. Medications used to treat high blood pressure can have side effects, but most people don't experience any. If you do, asking your doctor about changing medication will often help. You have been prescribed two drugs (A and B) to take in the following way:

Drug	**Directions**	**Recommendations**
A	Take one tablet once a day.	Take in the morning when you wake up.
B	Take two tablets separately once a day.	Leave at least 6 hours between tablets and ingest with food.

[**Subjects saw the message again above each item**.]

I might forget to take my medication

- Never
- Sometimes
- About half the time
- Most of the time
- Always

I might change the dosage of my medication

- Never
- Sometimes
- About half the time
- Most of the time
- Always

I might stop taking my medication for a while

- Never
- Sometimes
- About half the time
- Most of the time
- Always

I might decide to skip one of my medication dosages

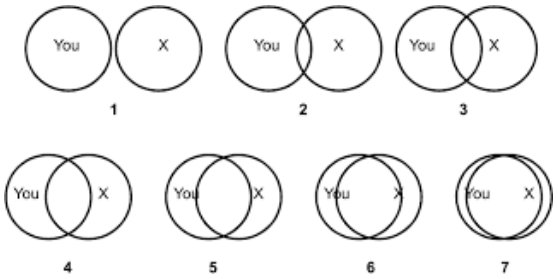
- Never
- Sometimes
- About half the time
- Most of the time
- Always

I might use my medication less than is prescribed

- Never
- Sometimes
- About half the time
- Most of the time
- Always

Have you known anyone that has, or previously had high blood pressure?

- Yes
- No



[**If Yes**] In the figure above, we ask you to consider which of these pairs of circles best describes your relationship with the person you have known that has (had) high blood pressure. In the figure "X" serves as a placeholder for this person, that is, you should think of "X" as being this individual. By selecting the appropriate number please indicate to what extent you and this person are connected. If you know (have known) more than one person with high blood pressure, then you should think about the person that you have the closest relationship to.

- 1
- 2

- 3
- 4
- 5
- 6
- 7

How likely do you think it is that you will develop high-blood pressure during the next five years?

- Extremely unlikely
- Somewhat unlikely
- Neither likely nor unlikely
- Somewhat likely
- Extremely likely

A.1.4 Study 2: Intentionality of non-adherence

Across the previous screens, you reported having not taken your medication as prescribed at some point over the last 2 weeks. In your experience, was this on purpose (Intentional) or was this something that just happened (Unintentional)? On a scale from 1 "Always intentional" to 5 "Always unintentional":

1. When you forgot to take your medication this was...
2. When you changed the dosage of your medication this was...
3. When you stopped taking your medication this was...
4. When you skipped a dose of your medication this was...
5. When you used less than was prescribed of your medication this was..

A.1.5 Study 3: Pre-screening

Participants who completed the main experiment in Study 3, first completed a pre-screening survey. The pre-screening study was also programmed in Qualtrics. In this section, we provide the question wording and response fields in the pre-screening study.

[PS1] A chronic condition is defined broadly as a condition that lasts 3 months or more and requires ongoing medical attention or limits activities of daily living or both. Do you take at least one prescribed course of medication to control a long-term chronic health condition such as diabetes, heart disease, stroke, high blood pressure etc.?

- Yes
- No

[**If "no", then subjects were exited from the survey.**]

[**The order in which scales [PS2] and [PS3] were presented to subjects was randomised.**]

[PS2] For each of the statements below, please indicate whether or not the statement is characteristic of you, or of what you believe about yourself from 1 = "Absolutely disagree" to 7 = "Absolutely agree"

- If I reflect on my past, I see that I tend to be afraid of feeling emotions
- I feel that I need to experience strong emotions regularly
- Emotions help people to get along in life
- I find strong emotions overwhelming and therefore try to avoid them
- I think that it is important to explore my feelings
- I would prefer not to experience either the lows or highs of emotion
- I do not know how to handle my emotions, so I avoid them
- It is important for me to be in touch with my feelings
- It is important for me to know how others are feeling
- Emotions are dangerous and they tend to get me into situations that I would rather avoid

[PS3] For each of the statements below, please indicate whether or not the statement is characteristic of you, or of what you believe about yourself from 1 = "Absolutely disagree" to 7 = "Absolutely agree"

- I would prefer complex to simple problems
- I like to have the responsibility of handling a situation that requires a lot of thinking
- Thinking is not my idea of fun
- I would rather do something that requires little thought than something that is sure to challenge my thinking abilities
- I really enjoy a task that involves coming up with new solutions to problems
- I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought

---

[PS4] Thinking about your chronic condition, and the medication for which you are prescribed. On a scale of 1 "Not at all" to 7 "Completely", how important is it for you to improve how you are taking your medication?

- 1

- 2
- 3
- 4
- 5
- 6
- 7

[PS5] On a scale of 1 "Not at all" to 7 "Completely", how happy are you with your health at the moment as it relates to your chronic condition?

- 1
- 2
- 3
- 4
- 5
- 6
- 7

A.2 Additional analysis

A.2.1 Tests for joint orthogonality

Tests for joint orthogonality confirm that subject characteristics were overall balanced between message condition and, in Study 3, between self-report measure condition. We use multinomial logit regression to investigate the balance of subject characteristics. Full sets of demographic (age, gender, education, subjective socioeconomic score, household income), medication regime (medication prescribed for 3+ years, taking more than one prescribed medication, taking more than one dose a day on average, subjective complexity, importance of adherence, access to health care via NHS only, condition group), illness group, social desirability concern measures (SDE score, IM score), and in the case of Study 3, the additional COVID-19 control measures (effect of the pandemic on motivation to be adherent, effect of the pandemic on ability to be adherent, taking immunosuppressive medication).

Study	Outcome	Prob $> \chi 2$
1	Message condition	0.1173
2	Message condition	0.5434
3	Message condition	0.6717
3	Time-framed self-report condition	0.0798

Table A.1: Summary of joint orthogonality tests

A.2.2 Distribution of NTF- and TF-self-reports

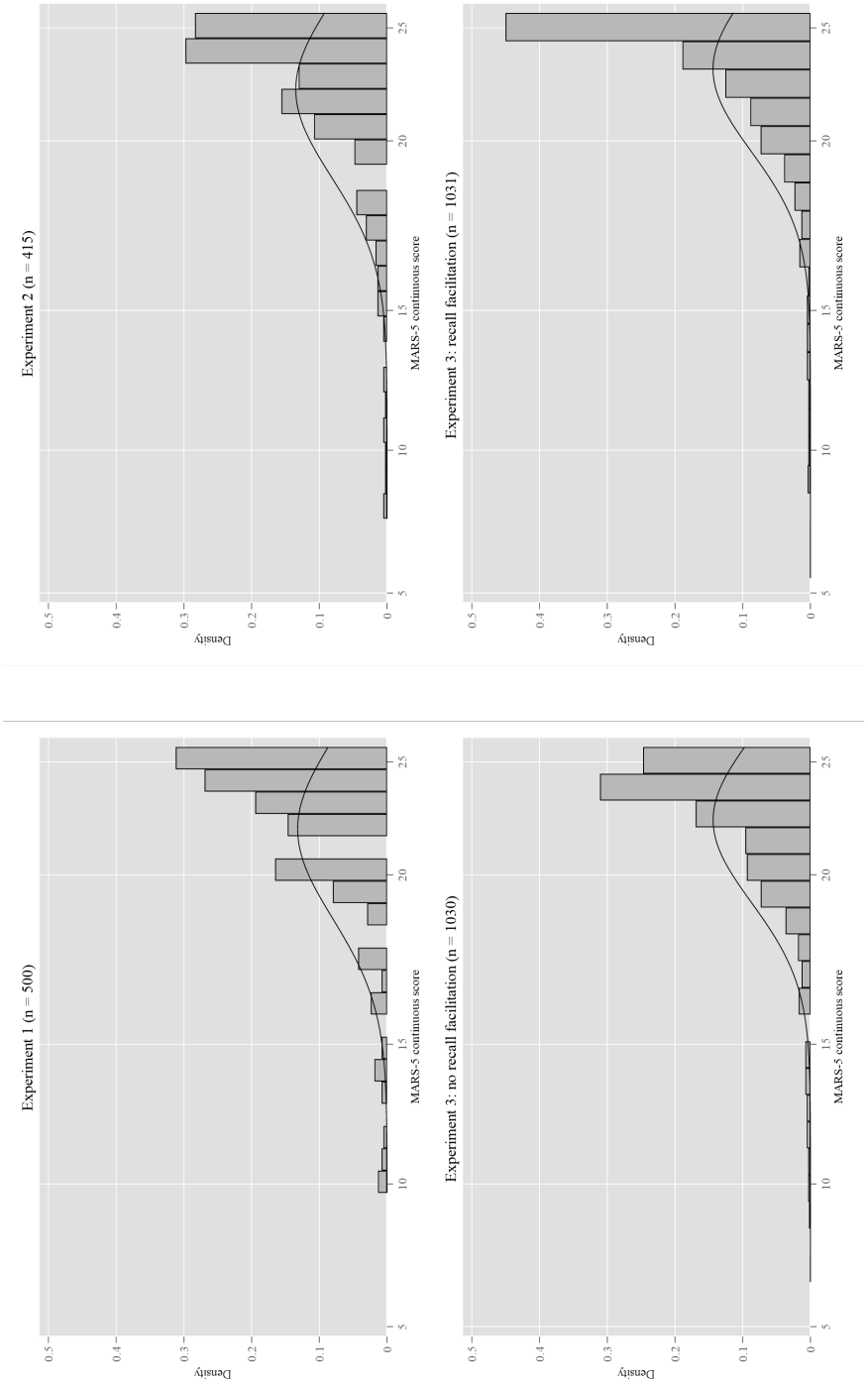


Figure A.1: Distribution of self-reports of adherence

A.2.3 Summary of sample charactistics between Study 1, 2 and 3

Table A.2 shows that there were some differences in sample characteristics across studies. In terms of demographics, respondents in Study 3, were slightly older on average (M = 45.69, SD =14.38) than those in Study 1 (M = 42.67, SD = 13.62) and 2 (M = 42.74, SD = 13.21) ($F(2, 2973) = 14.26, p = 0.000$). We also find small socio-economic differences in respondents across studies. Subjects in Study 1 reported the lowest average subjective socioeconomic status score (F(2, 2973) = 10.46, p = 0.000), and subjects in Study 3 were reported higher average household income (χ2(8) = 56.78, p = 0.000).

In terms of medication regime characteristics, participants in Study 2 were more likely to belong to the recently prescribed group (<1 year) (M = 0.13, χ2(2) = 15.08, p = 0.001). Subjective ratings of complexity in Study 3 (M = 2.18, SD = 1.18) and 2 (M = 2.12, SD = 1.19) were higher than those in Study 1 (M = 1.93, SD = 1.06) (F(2, 2973) = 9.16, p = 0.001). Also, there were fewer participants reporting polypharmacy in experiment 1 (M = 0.61, χ2(2), p = 0.000).

Finally participants recruited in Study 3 reported higher average IM scores relative to those in Study 1 and Study 2 (F(2, 2972) = 9.66, p = 0.000).

Table A.2: Summary of sample characteristics across independent samples in Study 1, 2, and 3

	Study 1		Study 2		Study 3		
	M	SD	M	SD	M	SD	p
Demographics							
Age	42.67	13.62	42.74	13.21	45.69	14.38	0.000
Gender							0.168
Female	0.72	0.45	0.66	0.48	0.70	0.46	
Male	0.27	0.44	0.34	0.47	0.29	0.45	
Non-binary	0.01	0.09	0.01	0.07	0.01	0.10	

Continued on next page

Table A.2 – continued from previous page

	Study 1		Study 2		Study 3		
	M	SD	M	SD	M	SD	p
Prefer not to say	0.00	0.00	0.00	0.00	0.00	0.05	
Highest level of education							0.754
No qual	0.01	0.12	0.01	0.12	0.01	0.11	
GCSEs or eq.	0.18	0.38	0.16	0.36	0.14	0.35	
A-levels or eq.	0.23	0.42	0.27	0.44	0.24	0.43	
University degree or eq.	0.43	0.50	0.42	0.49	0.45	0.50	
Post-grad degree	0.15	0.36	0.14	0.35	0.16	0.36	
Other	0.00	0.06	0.01	0.07	0.00	0.06	
Prefer not to say	0.00	0.00	0.00	0.00	0.00	0.04	
Subjective socio-economic	4.92	1.53	5.24	1.74	5.28	1.58	0.000
Household income							0.000
up to £9,999	0.10	0.30	0.09	0.28	0.06	0.23	
£10000 to £24999	0.34	0.48	0.33	0.47	0.25	0.44	
£25000 to £49999	0.36	0.48	0.37	0.48	0.40	0.49	
£50000 to £99999	0.18	0.38	0.18	0.38	0.24	0.43	
£100000 +	0.02	0.13	0.04	0.19	0.05	0.21	
Chronic illness							0.000
Hypertension	0.15	0.35	0.10	0.30	0.08	0.27	
Cardiovascular disease	0.02	0.15	0.03	0.17	0.04	0.21	
Stroke	0.02	0.15	0.04	0.20	0.01	0.11	
Asthma/COPD	0.20	0.40	0.14	0.35	0.10	0.30	
Diabetes	0.19	0.39	0.23	0.42	0.21	0.41	
Arthritis	0.02	0.13	0.03	0.17	0.07	0.26	
Cancer	0.00	0.06	0.00	0.05	0.03	0.17	
Endometriosis	0.00	0.00	0.00	0.00	0.02	0.14	
Fibromyalgia	0.00	0.05	0.00	0.05	0.05	0.21	
IBD	0.01	0.09	0.01	0.09	0.07	0.26	
Mood disorders	0.04	0.20	0.03	0.18	0.07	0.25	

Continued on next page

Table A.2 – continued from previous page

	Study 1		Study 2		Study 3		
	M	SD	M	SD	M	SD	p
Multiple Sclerosis	0.01	0.09	0.00	0.05	0.02	0.15	
Hyperthyroidism	0.05	0.22	0.07	0.26	0.02	0.13	
Autoimmune condition	0.03	0.17	0.02	0.15	0.01	0.09	
Other	0.25	0.43	0.28	0.45	0.19	0.40	
Prefer not to say	0.01	0.12	0.01	0.10	0.01	0.07	
Prescribed medication							
Time taking medication							0.002
0-6 months	0.04	0.20	0.07	0.25	0.04	0.19	
7-11 months	0.04	0.20	0.06	0.24	0.03	0.18	
1-3 years	0.22	0.42	0.23	0.42	0.19	0.39	
3+ years	0.70	0.46	0.65	0.48	0.73	0.44	
Prefer not to say	0.00	0.00	0.00	0.00	0.00	0.05	
>1 medication prescribed	0.61	0.49	0.66	0.47	0.72	0.45	0.000
>1 dose per day	0.64	0.48	0.66	0.47	0.66	0.47	0.691
Subjective complexity	1.93	1.06	2.12	1.19	2.18	1.18	0.000
Importance of adherence	2.36	0.66	2.40	0.64	2.40	0.62	0.380
NHS only healthcare	0.92	0.28	0.90	0.30	0.89	0.32	0.185
Travel > once every 3 months	0.38	0.49	0.41	0.49	0	0	0.437
Travel complexity	2.79	1.36	2.93	1.38	0	0	0.133
COVID-19							
Immunosuppressive medication					−0.08	0.36	
Ability to be adherent					0.01	0.49	
Motivation to be adherent					0.19	0.39	
Social desirability bias							
SDE	3.83	0.94	3.78	0.90	3.84	0.96	0.556

Continued on next page

Table A.2 – continued from previous page

	Study 1		Study 2		Study 3		
	M	SD	M	SD	M	SD	p
IM	4.09	1.07	4.11	1.02	4.49	1.00	0.000

Note: F-tests and $\chi 2$ tests were used to compare subject characteristics across the independent samples collected in each study. Sample size was: n = 500 for Study 1, n = 415 for Study 2, and n = 2061 for Study 3.

A.2.4 Discussion of potential effect of the COVID-19 pandemic on results

Could the results of Study 3 have been affected by the context of a global pandemic? Certainly, the effect of the pandemic on medication adherence in chronic patient populations has been complex and heterogeneous across illness and therapy type. At the time of writing, most studies investigating the impact of the pandemic on patient behaviour conclude adherence likely worsened during the pandemic (Esposti et al., 2020; Gannon et al., 2020; Racette et al., 2022). With the exception of patients living with Asthma/COPD, for whom some studies suggest adherence may have improved during the early months of 2020 (Dhruve et al., 2022; Kaye et al., 2020). However, many studies probing the effect of the pandemic on adherence behaviour use standardised self-report scales as the only measure of adherence. During the pandemic, patients experienced decreased access to healthcare facilities, follow-up, and medications due to intermittent lockdowns and overburdened healthcare systems (Ágh et al., 2021; Clement et al., 2021). Also increased levels of stress and anxiety (Mariotti, 2015), likely contributed to poorer adherence behaviour (Johnson, 2002). Fear of immunosuppressive effects of therapy was also shown to have contributed to poorer adherence (Khabbazi et al., 2020).

Could the effect of the pandemic on adherence explain why we find nudge bias in TF-self-reports in Study 3, but not Study 2. If we take this observation at face value, we might conclude that respondents simply reported a greater degree of nudge bias in self-reports in Study 3, relative to Study 2. On the other hand, Study 2 may have been underpowered, and a larger sample size was necessary to detect nudge bias in self-reports. It should be noted that the main result of Study 1 - that the £-NHS message produced nudge biased NTF-self-reports - was replicated in Study 3, despite comparable difference between global contexts and power. This suggests that the nudge bias in NTF-self-reports is not necessarily driven by the same psychological mechanisms as the nudge bias in TF-self-reports (more on this in Chapter 2).

To control for the possible effects of COVID-19 on actual adherence and perceptions of adherence promoting nudge messages, we included additional measures in Study 3.

When these additional measures are included as covariates in the regression analysis (see Table A.3), both NTF- and TF-self-reports become more nudge-robust. Subjects who report that the pandemic had reduced their motivation to be adherent to their medication, reported lower NTF ($\beta = -1.615, p < 0.001$) and TF adherence ($\beta = -1.848, p < 0.001$). Also, for NTF-self-reports, there was a negative relationship between reporting that the pandemic had reduced a subject's ability to be adherent and the level of adherence reported ($\beta = -0.817, p = 0.002$). However, we also find that subjects who reported that the pandemic had increased their motivation and ability to be adherent reported lower adherence than those who said the pandemic had no effect on their adherence.

As such, while including additional pandemic control questions into regression analysis reduces the degree of variation in self-reports explained by message condition, further investigation into the interaction between effect of the pandemic on motivation and ability and message condition reveals mixed results.

Reporting that the pandemic reduced motivation to be adherent resulted in lower TF-self-reports in the burden-NHS message treatment, relative to subjects who reported no motivational change due to the pandemic and those who saw the control message ($\beta = -1.86, p = 0.049$). We also find a marginally significant interaction between reduced motivation due to the pandemic and being in the health message condition on TF-self-reports ($\beta = 1.74, p = 0.089$). These results highlight the heterogeneous effects of the nudges on TF-self-reports for subjects reporting reduced motivation due to the pandemic. Impact of the pandemic on ability to be adherent, did not moderate the effect of message condition of self-reports.

For NTF-self-reports, we find reduced motivation moderated self-reports in the £-burden ($\beta = -2.80, p = 0.045$) and burden-NHS conditions ($\beta = 1.57, p = 0.098$). Also, for subjects in the health message condition, we find reduced moitvation due to the pandemic moderated NTF-self-reports ($\beta = -1.45, p = 0.075$).

The results suggest that self-reports of subjects who report determinants of their adherence (motivation and ability) are negatively affected by COVID-19, are more sensitive to the presence of nudge messages than those who report no impact of the pandmic on motiv-

ation and ability. However, the effect of nudge messages on self-reports differs depending on whether they are time-framed. In the case of the health message, reduced motivation due to the pandemic lead to higher TF-self-reports and lower NTF-self-reports relative to the control. The opposite is true for the NHS-burden message. An important caveat here is that many of these moderation effects are only marginally significant. This is potentially due to low power, as the sample was not designed to specifically investigate the impact of the pandemic on adherence as a moderator of nudge bias. Also, we cannot rule out the effect of omitted variable bias. Therefore, any inferences about the determinants of nudge bias based on these interaction effects, should be made with care.

	(1) NTF-self-reports	(2) TF-self-reports
Health	0.326	0.448*
	(0.266)	(0.253)
£-NHS	0.347	0.464
	(0.250)	(0.275)
Burden-NHS	0.250	0.286
	(0.262)	(0.261)
Norm	0.084	−0.441*
	(0.263)	(0.262)
COVD-19 effect on adherence		
Reduced ability	−0.817***	−0.502*
	(0.262)	(0.275)
Increased ability	−1.012**	−0.253
	(0.483)	(0.526)
Reduced motivation	−1.615***	−1.848***
	(0.279)	(0.259)
Increased motivation	−0.437*	−0.568**
	(0.232)	(0.241)
% Immunosuppressive	0.232	0.350
	(0.197)	(0.197)
Constant	18.57***	18.77***
	(0.694)	(0.625)
Controls	YES	YES
R^2	0.231	0.207
F	12.50	10.93
N	978	986

Note: Significance indicated: * p < 0.1, ** p < 0.05, *** p < 0.01. Models 1 and 2 are OLS. All models included controls for demographics (age, education), medication regime (recency, subjective complexity, importance), and social desirability bias scores (SDE, IM). Baseline group: message condition = control message; Ability = No change; Motivation = No change.

Table A.3: Study 3 regression analysis with controls for pandemic effects

Appendix B

Appendix: Chapter 2

B.1 Design

B.1.1 Study 3: Cognitive and affective evaluation of messages

[**In Study 3, subjects were asked to evaluate the message in terms of affect and cognition. The order within which subjects completed the affective and cognitive evalautation was randomised to control for order effects. Subjects in the control group were only asked to evaluate the control message in terms of affect.**]

[**Affect**] While reading the statement [**message shown**], which emotions do you experience to any degree? Select all that apply.

- Anger
- Wanting
- Dread
- Sad
- Easygoing
- Grossed out
- Happy
- Terror
- Rage
- Grief
- Nausea
- Anxiety
- Chilled out
- Desire
- Nervous
- Lonely
- Scared
- Mad
- Satisfaction
- Sickened
- Empty
- Craving

- Panic
- Longing
- Calm
- Fear
- Relaxation
- Revulsion
- Worry
- Enjoyment
- Pissed off
- Liking
- None of the above

[**For each emotion selected:**] When reading the statement, how strongly did you feel the emotions on a scale of "Slightly" to "Very much".

- Slightly
- Moderately
- Quite a bit
- Very much

[**Cognition**] Thinking about the message [**message shown**], to what extent do you agree with the following statements from 1 = "not at all" and 7 ="completely".

- The message is interesting
- I like the message
- The message is worth sharing
- The message is relevant to me
- The message makes me think about my actions
- The message grabs my attention
- The message is memorable
- The message is easy to comprehend
- The information in this message is common knowledge
- The information in this message is new to me
- The message makes me want to improve my adherence

- The message will help me to improve my adherence
- The message reminds me of something I ought to do
- The message reminds me of something I ought not to do
- The message is credible
- The message is believable
- The message is overblown
- The message is annoying
- The message is manipulative
- The message is trying to get me do something that will be good for me
- The message reminds me of something I want to do

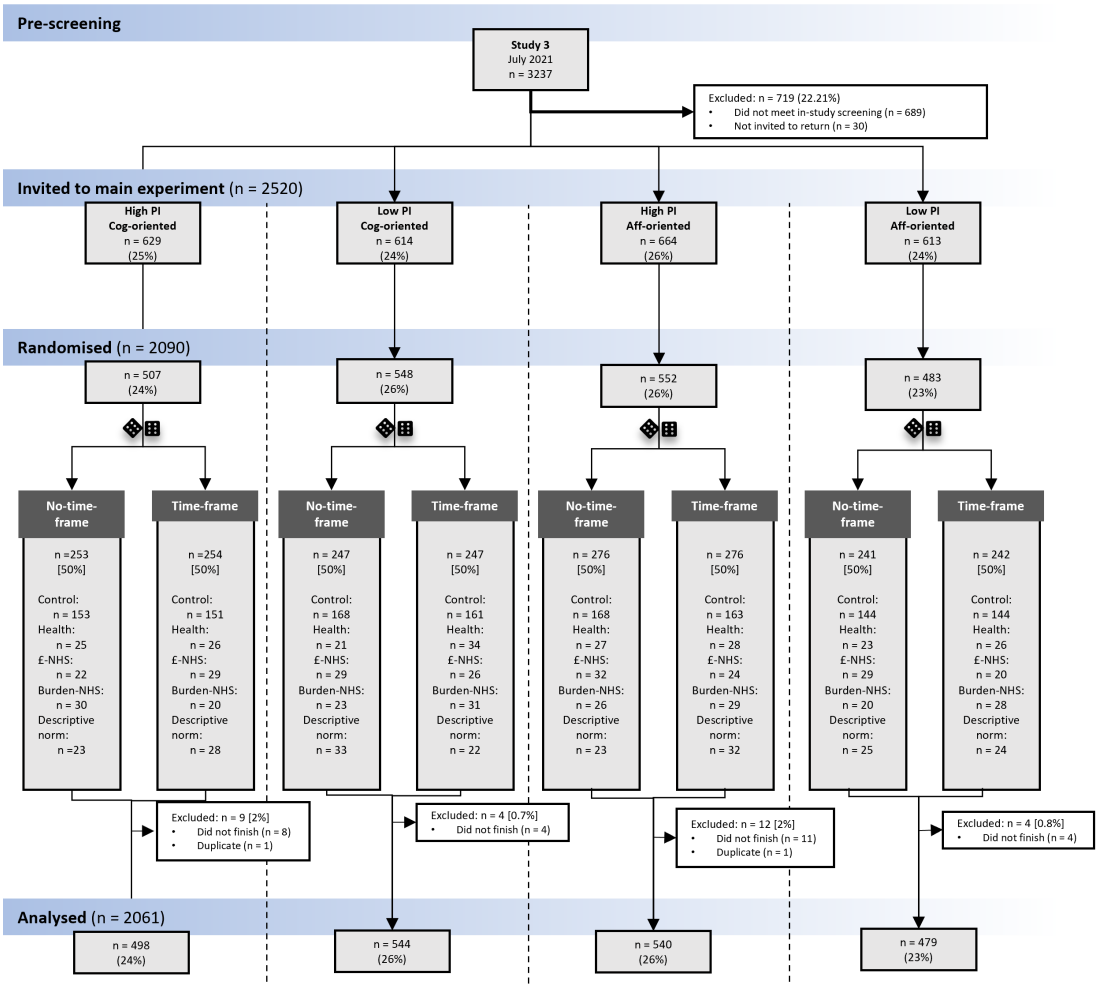
B.2 Additional analysis

B.2.1 Study 3 detailed trial profile

The number of respondents recruited, randomised, and analysed is summarised in Figure B.1.

At pre-screening, 3,237 subjects were recruited from Prolific. Within this sample, 689 (21.29%) subjects were coded as ineligible and not invited to the main experiment. Only eligible participants were invited to the main experiment (n = 2,520), of which, 629 are coded as cognitive-oriented with high personal involvement, 614 as cognitive-oriented with low personal involvement, 664 as affective-oriented with high personal involvement, and 613 as affective-oriented with low personal involvement.

In total, 2,061 participants completed all stages of the main experiment and are included in the main analysis: 1,031 (50.02%) subjects completed the time-framed self-report, and 1,030 the self-report with no time-framing. In terms of message condition, 1,236 (59.97%) of subjects saw the neutral control messages, 208 (10.09%) saw the health nudge, 207 (10.04%) saw the £-NHS nudge, 204 (9.90%) the burden-NHS nudge, and 206 (10%) the descriptive norm nudge.



Note: percentages in round parentheses use total number of respondents included at the relevant stage as denominator, percentages in square parentheses use total number randomised within each stratified group as denominator.

Figure B.1: Study 3 trial profile with randomisation by strata information

B.2.2 Sample characteristics by strata

Table B.1: Sample characteristics in Study 3 by strata

	NFC-oriented				NFA-oriented				
	High PI		Low PI		High PI		Low PI		
	n = 498		n = 544		n = 540		n = 479		
	M	*SD*	M	*SD*	M	*SD*	M	*SD*	p
Demographics									
Age	44.50	13.87	48.41	14.94	43.91	13.62	43.91	13.62	**0.000**
Gender									
Female	0.66	0.48	0.59	0.49	0.80	0.40	0.80	0.40	**0.000**
Male	0.32	0.47	0.41	0.49	0.19	0.39	0.19	0.39	
Non-binary	0.02	0.13	0.01	0.09	0.01	0.11	0.01	0.11	
Prefer not to say	0.01	0.09	0.00	0.00	0.00	0.00	0.00	0.00	
Education									
No qual	0.01	0.08	0.01	0.07	0.02	0.13	0.02	0.13	**0.000**
GCSEs or eq.	0.11	0.31	0.11	0.32	0.18	0.38	0.18	0.38	
A-levels or eq.	0.24	0.43	0.19	0.39	0.30	0.46	0.30	0.46	
University degree or eq.	0.47	0.50	0.49	0.50	0.39	0.49	0.39	0.49	
Post-grad degree	0.18	0.38	0.19	0.40	0.11	0.31	0.11	0.31	
Other	0.00	0.05	0.01	0.07	0.01	0.07	0.01	0.07	
Prefer not to say	0.00	0.05	0.00	0.00	0.00	0.06	0.00	0.06	
Subjective socio-economic	5.14	1.62	5.61	1.58	4.97	1.53	4.97	1.53	**0.000**
Household income									**0.002**
up to £9,999	0.07	0.25	0.04	0.20	0.05	0.22	0.05	0.22	
£10000 to £24999	0.26	0.44	0.22	0.41	0.31	0.46	0.31	0.46	
£25000 to £49999	0.40	0.49	0.40	0.49	0.41	0.49	0.41	0.49	
£50000 to £99999	0.23	0.42	0.29	0.45	0.20	0.40	0.20	0.40	
£100000 +	0.04	0.21	0.06	0.24	0.03	0.16	0.03	0.16	

Continued on next page

Table B.1 – continued from previous page

	NFC-oriented				NFA-oriented				
	High PI		Low PI		High PI		Low PI		
	n = 498		n = 544		n = 540		n = 479		
	M	*SD*	M	*SD*	M	*SD*	M	*SD*	p
Chronic illness									**0.000**
Hypertension	0.06	0.23	0.09	0.29	0.05	0.21	0.05	0.21	
Cardiovascular disease	0.05	0.22	0.06	0.24	0.03	0.18	0.03	0.18	
Stroke	0.01	0.12	0.02	0.13	0.01	0.11	0.01	0.11	
Asthma/COPD	0.10	0.31	0.10	0.30	0.09	0.28	0.09	0.28	
Diabetes	0.24	0.43	0.17	0.37	0.27	0.44	0.27	0.44	
Arthritis	0.09	0.28	0.06	0.23	0.09	0.28	0.09	0.28	
Cancer	0.02	0.15	0.04	0.19	0.03	0.16	0.03	0.16	
Endometriosis	0.02	0.15	0.02	0.15	0.02	0.15	0.02	0.15	
Fibromyalgia	0.07	0.25	0.02	0.13	0.07	0.26	0.07	0.26	
IBM	0.06	0.24	0.06	0.24	0.07	0.25	0.07	0.25	
Mood disorders	0.06	0.24	0.06	0.24	0.06	0.24	0.06	0.24	
Multiple Sclerosis	0.01	0.08	0.03	0.18	0.03	0.18	0.03	0.18	
Hyperthyroidism	0.01	0.12	0.03	0.16	0.02	0.12	0.02	0.12	
Autoimmune condition	0.01	0.09	0.01	0.09	0.01	0.10	0.01	0.10	
Other	0.18	0.38	0.22	0.42	0.16	0.37	0.16	0.37	
Prefer not to say	0.01	0.08	0.01	0.09	0.01	0.07	0.01	0.07	
Prescribed medication information									
Time taking medication									**0.000**
0-6 months	0.06	0.25	0.02	0.15	0.05	0.21	0.05	0.21	
7-11 months	0.04	0.20	0.03	0.16	0.05	0.21	0.05	0.21	
1-3 years	0.22	0.41	0.14	0.35	0.21	0.41	0.21	0.41	
3+ years	0.67	0.47	0.81	0.40	0.69	0.46	0.69	0.46	
Prefer not to say	0.00	0.06	0.00	0.06	0.00	0.04	0.00	0.04	

Continued on next page

Table B.1 – continued from previous page

	NFC-oriented				NFA-oriented				
	High PI		Low PI		High PI		Low PI		
	n = 498		n = 544		n = 540		n = 479		
	M	*SD*	M	*SD*	M	*SD*	M	*SD*	p
% >1 medication prescribed	0.78	0.42	0.65	0.48	0.76	0.43	0.76	0.43	**0.000**
% > 1 dose per day	0.74	0.44	0.58	0.49	0.74	0.44	0.74	0.44	**0.000**
Subjective complexity	2.49	1.24	1.86	1.03	2.47	1.23	2.47	1.23	**0.000**
Importance of adherence	2.47	0.58	2.28	0.63	2.51	0.58	2.51	0.58	**0.000**
% NHS only healthcare	0.88	0.33	0.86	0.35	0.90	0.31	0.90	0.31	**0.009**
COVID-19									
% Immunosuppressive	-0.16	0.42	-0.06	0.32	-0.07	0.38	-0.07	0.38	0.085
Ability to be adherent	-0.08	0.57	0.06	0.42	-0.01	0.53	-0.01	0.53	**0.000**
Motivation to be adherent	0.17	0.37	0.17	0.38	0.21	0.41	0.21	0.41	**0.000**
Social desirability bias									
SDE	3.67	0.97	4.01	0.91	3.70	0.98	3.70	0.98	**0.000**
IM	4.43	1.01	4.56	0.97	4.49	1.01	4.49	1.01	0.210

B.2.3 Correlation between moderators and mediators

1 of 3	1.	1a.	1b.	2.	2a.	2b.	3.
1. Orientation	1.0000						
1a. NFC	-0.7009*	1.0000					
1b. NFA	0.5546*	0.2047*	1.0000				
2. PI	0.0316	-0.0915*	-0.0633*	1.0000			
2a. Behaviour	0.0200	-0.0123	0.0132	0.8042*	1.0000		
2b. Outcome	-0.0293	0.1372*	0.1198*	-0.7332*	-0.1855*	1.0000	
3. IM	-0.0058	0.1291*	0.1427*	-0.0303	0.0225	0.0758*	1.0000
4. Appraisal	0.0570	0.0163	0.0966*	0.0998*	0.1482*	0.0060	0.1011*
5. Engagement	0.0666	0.0287	0.1241*	0.1705*	0.2089*	-0.0413	0.0539
6. Quality	0.0348	0.0405	0.0947*	-0.0383	-0.0138	0.0477	0.1228*
7. Intentions	0.0726	-0.0188	0.0768	0.1837*	0.2530*	-0.0121	0.0018
8. Reactance	-0.0885	0.0243	-0.0919*	0.0054	-0.0296	-0.0435	-0.1385*
9. Desirability	0.0542	0.0149	0.0911*	0.0809	0.1352*	0.0223	0.0333
10. % Anger	-0.0105	-0.0133	-0.0300	0.0767*	0.0571*	-0.0614*	-0.0670*
11. % Anxiety	-0.0028	-0.021	-0.0284	0.1734*	0.1297*	-0.1382*	-0.0993*
12. % Disgust	0.0387	-0.0533	-0.0091	0.1132*	0.0686*	-0.1086*	-0.0572*
13. % Fear	0.0204	-0.0630*	-0.0455	0.1267*	0.0905*	-0.1060*	-0.0513
14. % Sadness	-0.047	-0.0108	-0.0770*	0.1410*	0.0918*	-0.1280*	-0.0618*
15. % Desire	-0.0063	-0.0039	-0.0132	0.0637*	0.0690*	-0.0264	-0.0525
16. % Relaxed	0.0114	0.0527	0.0771*	-0.1120*	-0.0463	0.1321*	0.0652*
17. % Happiness	0.0422	0.0276	0.0900*	-0.0383	0.0352	0.1037*	0.0741*

Note: * $p < 0.001$.

2 of 3		4.	5.	6.	7.	8.	9.
4.	Appraisal	1.0000					
5.	Engagement	0.6929*	1.0000				
6.	Quality	0.5808*	0.5167*	1.0000			
7.	Intentions	0.5368*	0.7207*	0.3913*	1.0000		
8.	Reactance	-0.4751*	-0.3355*	-0.5001*	-0.2704*	1.0000	
9.	Adherence SD	0.4857*	0.6422*	0.4171*	0.6391*	-0.2141*	1.0000
10.	% Anger	0.0496	0.031	0.0487	0.0187	0.0514	-0.0518
11.	% Anxiety	0.0273	0.1298*	-0.0099	0.1365*	0.0585	0.1075*
12.	% Disgust	0.0892	0.0888	0.0619	0.0658	0.0143	0.0157
13.	% Fear	0.0069	0.0534	-0.042	0.0758	0.049	0.0771
14.	% Sadness	0.0439	0.0485	0.0331	0.0521	-0.0289	-0.027
15.	% Desire	0.0067	0.045	-0.0039	0.0197	0.0128	0.0562
16.	% Relaxed	0.0564	0.0021	-0.0115	-0.0416	-0.0654	-0.0049
17.	% Happiness	0.1264*	0.0769	0.0823	0.0479	-0.0934*	0.0834

Note: * $p < 0.001$.

3 of 3		10.	11.	12.	13.	14.	15.	16.
10.	% Anger	1.0000						
11.	% Anxiety	0.3242*	1.0000					
12.	% Disgust	0.3805*	0.2610*	1.0000				
13.	% Fear	0.3371*	0.4694*	0.3501*	1.0000			
14.	% Sadness	0.4312*	0.3691*	0.2698*	0.3246*	1.0000		
15.	% Desire	0.3232*	0.2707*	0.2265*	0.2950*	0.3516*	1.0000	
16.	% Relaxed	-0.1035*	-0.2418*	-0.0707*	-0.0916*	-0.0983*	0.1192*	1.0000
17.	% Happiness	0.1340*	0.0609*	0.0935*	0.1027*	0.0818*	0.3031*	0.3841*

Note: * $p < 0.001$.

Table B.2: Correlation between moderator and mediator variables

B.2.4 Simple slopes moderation analysis

IM

Moderator	DV	IV (T_i)	dydx	SE	p	95% CI	
IM	TF-self-reports	Control (n = 612)	0.388	0.104	0.000	0.184	0.593
		Health (n = 113)	0.100	0.230	0.663	-0.352	0.553
		£-NHS (n = 96)	-0.063	0.251	0.802	-0.556	0.430
		Burden-NHS (n = 105)	0.213	0.243	0.380	-0.264	0.690
		Descriptive norm (n = 105)	0.331	0.261	0.205	-0.182	0.844
	NTF-self-reports	Control (n =624)	0.113	0.106	0.285	-0.094	0.320
		Health (n = 95)	0.288	0.266	0.279	-0.234	0.810
		£-NHS (n = 111)	-0.034	0.262	0.897	-0.548	0.481
		Burden-NHS (n = 99)	-0.056	0.233	0.810	-0.514	0.401
		Descriptive norm (n = 101)	0.266	0.242	0.273	-0.210	0.741

Note: Average marginal effects estimated from OLS regressions with controls for demographic (age, education) and medication characteristics (recency, complexity, importance, more than one dose per day) and social desirability subscales. Average marginal effects calculated for each message condition.

Table B.3: Simple slopes analysis on IM as moderator

PI & NFC

Moderator	DV	IV (T_i)	dydx	SE	p	95% CI	
PI	TF-self-reports	Control (n = 612)	-0.604	0.102	0.000	-0.805	-0.404
		Health (n = 113)	-0.185	0.253	0.466	-0.682	0.312
		£-NHS (n = 96)	-0.091	0.266	0.732	-0.613	0.431
		Burden-NHS (n = 105)	-0.202	0.272	0.459	-0.736	0.332
		Descriptive norm (n = 105)	-0.364	0.247	0.140	-0.805	0.120
	NTF-self-reports	Control (n =624)	-0.369	0.107	0.001	-0.580	-0.159
		Health (n = 95)	-0.461	0.262	0.079	-0.976	0.053
		£-NHS (n = 111)	-0.491	0.247	0.047	-0.975	-0.007
		Burden-NHS (n = 99)	-0.059	0.239	0.803	-0.527	0.409
		Descriptive norm (n = 101)	-0.261	0.246	0.289	-0.743	0.222
NFC	TF-self-reports	Control (n = 612)	-0.116	0.105	0.271	-0.322	0.090
		Health (n = 113)	-0.081	0.281	0.772	-0.632	0.469
		£-NHS (n = 96)	0.101	0.227	0.655	-0.343	0.546
		Burden-NHS (n = 105)	-0.391	0.285	0.170	-0.950	0.168
		Descriptive norm (n = 105)	-0.452	0.224	0.044	-0.892	-0.012
	NTF-self-reports	Control (n =624)	-0.047	0.105	0.659	-0.253	0.160
		Health (n = 95)	-0.348	0.270	0.199	-0.878	0.183
		£-NHS (n = 111)	-0.173	0.246	0.482	-0.656	0.310
		Burden-NHS (n = 99)	0.207	0.250	0.408	-0.284	0.699
		Descriptive norm (n = 101)	-0.201	0.294	0.493	-0.777	0.375

Note: Average marginal effects estimated from OLS regressions with controls for demographic (age, education) and medication characteristics (recency, complexity, importance, more than one dose per day) and social desirability subscales. Average marginal effects calculated for each message condition.

Table B.4: Simple slopes analysis on PI and NFC as moderators

NFA

Moderator	DV	IV (T_i)	dydx	SE	p	95% CI	
NFA	TF-self-reports	Control (n = 612)	0.080	0.106	0.449	-0.128	0.289
		Health (n = 113)	-0.044	0.229	0.849	-0.494	0.406
		£-NHS (n = 96)	-0.282	0.277	0.308	-0.826	0.261
		Burden-NHS (n = 105)	0.196	0.241	0.416	-0.277	0.669
		Descriptive norm (n = 105)	0.176	0.238	0.459	-0.291	0.643
	NTF-self-reports	Control (n =624)	0.251	0.102	0.014	0.051	0.450
		Health (n = 95)	-0.391	0.297	0.189	-0.974	0.192
		£-NHS (n = 111)	0.155	0.251	0.536	-0.337	0.647
		Burden-NHS (n = 99)	-0.202	0.253	0.424	-0.698	0.294
		Descriptive norm (n = 101)	0.256	0.283	0.367	-0.300	0.812

Note: Average marginal effects estimated from OLS regressions with controls for demographic (age, education) and medication characteristics (recency, complexity, importance, more than one dose per day) and social desirability subscales. Average marginal effects calculated for each message condition.

Table B.5: Simple slopes analysis on NFA as moderator

B.2.5 Message affective response by NTF- or TF-self-report

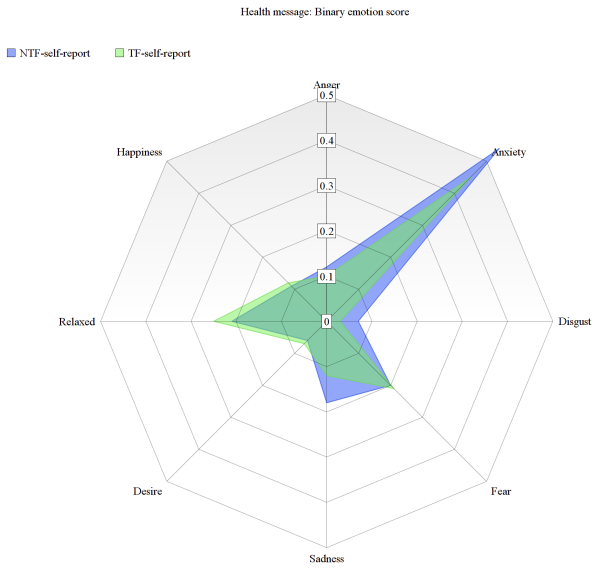


Figure B.2: Health message affective response: by subjects who completed NTF- or TF-self-report

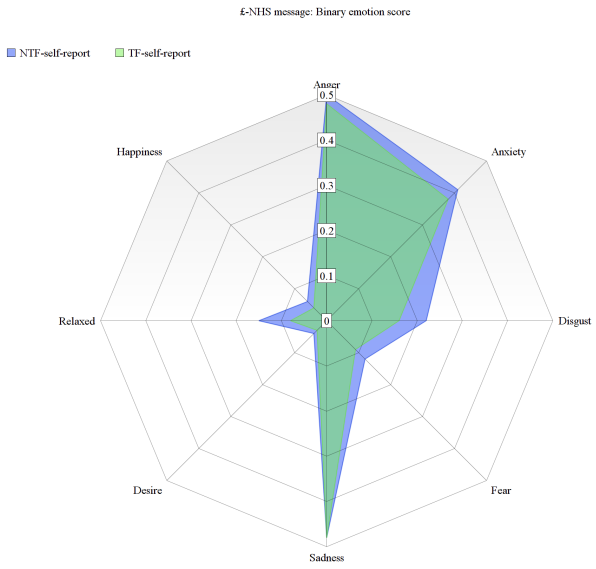


Figure B.3: £-NHS message affective response: by subjects who completed NTF- or TF-self-report

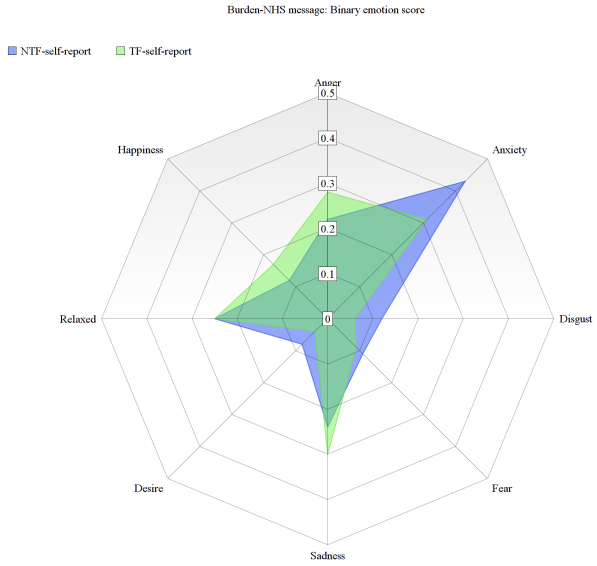


Figure B.4: Burden-NHS message affective response: by subjects who completed NTF- or TF-self-report

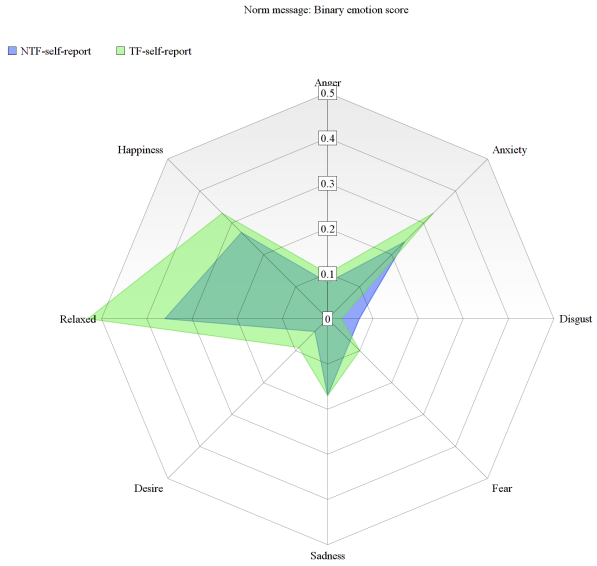


Figure B.5: Descriptive norm message affective response: by subjects who completed NTF- or TF-self-report

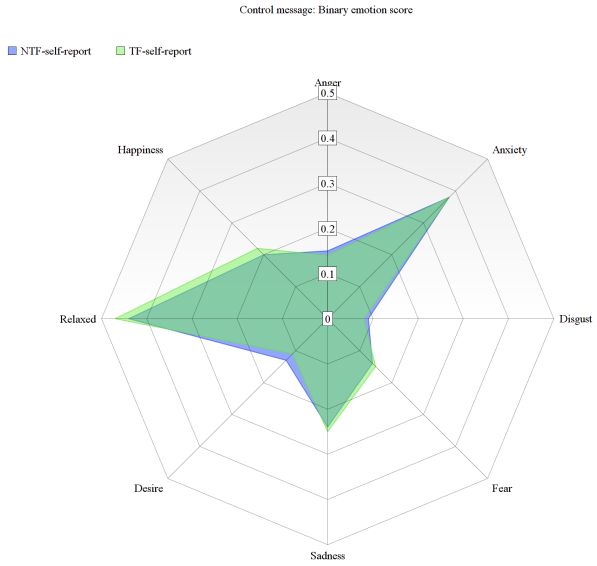


Figure B.6: Control message affective response: by subjects who completed NTF- or TF-self-report

Appendix C

Appendix: Chapter 3

C.1 Design

C.1.1 Experimental Instructions

Instructions were encountered by all subjects except when stated otherwise. Here, we report exact question wording and response fields. The questionnaire was programmed in Qualtrics. For an overview of the question structure:

Q1 Informed consent

Q9 - 12 Demographics

Q13 - 17 Blood donation history

Q19 Communication manipulation

Q20 - 39 Outcomes

Q40 Behavioural outcome

Q41 Debrief

[**Participants were given detailed information about the study in the Prolific study description, including what they would be asked, how their data would be used, how long the experiment would take approximately, that they could leave the study at any time, the contact details of the researchers, and what they would be paid for participating. This information can be provided on demand.**]

[Q1] Please enter your Prolific ID here: [**Text input**]

[Q2] I confirm that I have read and understand the information about the study.

- Yes
- No

[Q3] I have received enough information about the study.

- Yes
- No

[Q4] I understand I can withdraw from the study at any time, prior to submitting the study with no negative consequences.

- Yes
- No

[Q5] I understand my responses will remain anonymous.

- Yes
- No

[Q6] I am consenting for my information to be used as data towards this research provided my anonymity is completely protected.

- Yes
- No

[Q7] I understand that should I choose to withdraw, the information collected so far cannot be erased and may still be used in the project analysis.

- Yes
- No

[Q8] I can confirm I have read the information about this study and I voluntarily agree to take part in the study.

- Yes
- No

[**If respondents answered "no" to any of these informed consent questions, they were directed immediately to the debrief information (see [Q41]) and left the study.**]

Demographic Questionnaire

[Q9] How old are you? [**Text input**]

[Q10] How would you personally describe your gender?

- Female
- Male
- Gender non-conforming
- Other
- Prefer not to say

[Q11] What is your ethnicity?

- English/Welsh/Scottish/Northern Irish/British
- Irish

- Gypsy or Irish Traveller
- Other White background
- White and Black Caribbean
- White and Black African
- White and Asian
- Other Mixed/multiple ethnic background
- Indian
- Pakistani
- Bangladeshi
- Chinese
- Other Asian background
- African
- Caribbean
- Other Black/African/Caribbean background
- Arab
- Other ethnic group
- Prefer not to say

[Q12] What is your sexual orientation?

- Asexual
- Bisexual
- Gay
- Heterosexual/straight
- Lesbian
- Queer
- Pansexual
- Bi-curious
- Prefer not to say

[Q13] Have you ever donated blood?

- Yes

- No
- Prefer not to say

[Q14] Have you ever been a recipient of donated blood or it's components?

- Yes
- No
- Prefer not to say

[Q15] Would you consider donating blood in the future?

- Yes
- No
- Prefer not to say

[**Only respondents who answer "yes" to [Q13], were then asked to answer [Q16] and [Q17] to get more information about their blood donation history.**]

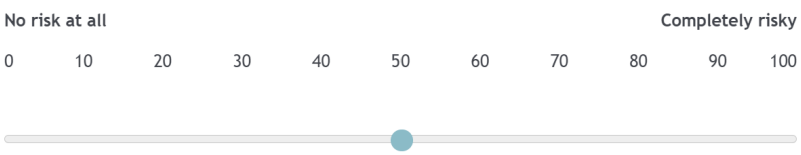
[Q16] Have you donated blood in the UK?

- Yes
- No
- Prefer not to say

[Q17] When was the last time you donated?

- Less than a month ago
- 2 to 12 months ago
- 12 months to 2 years ago
- Longer than 2 years ago
- I cannot remember

[Q18] To what extent do you think patients in the UK are at risk of being infected with viruses such as HIV or hepatitis by blood transfusion from infected blood donors? Please use the slider to answer on a scale of 0 = No Risk At All, to 100 = Completely Risky.



[Q19] Take some time to read the following policy statement carefully. In the next part of the survey, you will be asked some questions about your opinion on the statement. [**The message shown to participants was exogenously manipulated. Below is the message shown to subjects in the "donor" focus and "safety" frame condition. To see all the communications used in the experiment see Appendix C.1.2.**]

The Donor Safety Check

The rules on blood donation in the UK will change in the summer of 2021. Previously, men-who-have-sex-with-men (MSM) were not able to donate blood unless they had abstained from sexual behaviour for 3 months. The new policy will mean that all potential donors, regardless of age, sex, sexuality, or ethnicity, will complete the same 'Donor Safety Check' before they donate.

The 'Donor Safety Check' covers a wide range of questions of health and travel to ensure that the donor is healthy to donate blood. The 'Donor Safety Check' will now include several new questions on sexual behaviours. These questions are asked to further ensure that donors are safe to give blood.

Thus, donation decisions are based on donor's behaviour, not their sexuality.

[Q20] I am willing to donate blood.

- Yes
- No

[Q21] I plan to donate blood in the near future.

- Yes
- No

[Q22] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely. To what extent does this statement make you feel that blood is screened to ensure it is safe?

- 1
- 2
- 3

- 4
- 5
- 6
- 7

[Q23] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely. To what extent does this statement make you feel that donor selection is equitable and fair?

- 1
- 2
- 3
- 4
- 5
- 6
- 7

[Q24] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely. To what extent does the statement make you feel that the blood patients receive is safe?

- 1
- 2
- 3
- 4
- 5
- 6
- 7

[Q25] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely. To what extent would the statement put you off donating blood?

- 1
- 2
- 3
- 4
- 5
- 6

- 7

[Q26] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely. To what extent does the statement make you think about the patients who receive blood?

- 1
- 2
- 3
- 4
- 5
- 6
- 7

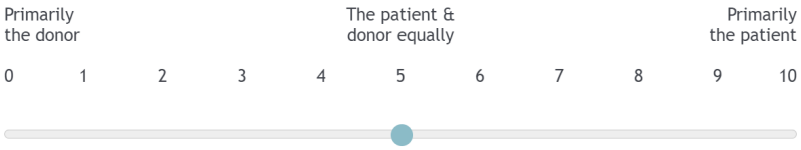
[Q27] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely. To what extent do you think that this statement would put other people off donating blood?

- 1
- 2
- 3
- 4
- 5
- 6
- 7

[Q28] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely. To what extent do you feel that this statement would encourage a wider diversity of people to donate blood?

- 1
- 2
- 3
- 4
- 5
- 6
- 7

[Q29] In your opinion, who is the focus of the statement? The donor, the patient receiving the blood, or a combination of both? Use the slider to indicate your response where: 0 = Primarily the donor and 10 = Primarily the patient.



[Q30] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely, how would you describe the statement?

Q30.1 Understandable

Q30.2 Informative

Q30.3 Memorable

Q30.4 Clear

Q30.5 Helpful

Q30.6 Upsetting

Q30.7 Upbeat

Q30.8 Encouraging

Q30.9 Inspiring

[Q31] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely. To what extent does the statement promote a sense of equality?

- 1
- 2
- 3
- 4
- 5
- 6
- 7

[Q32] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely. To what extent does the statement promote a sense of fairness?

- 1
- 2

- 3
- 4
- 5
- 6
- 7

[Q33] To what extent do you think asking people about their sexual behaviour is acceptable to assess blood donor eligibility?

- Extremely unacceptable
- Somewhat unacceptable
- Neither acceptable nor unacceptable
- Somewhat acceptable
- Extremely acceptable

[Q34] All potential donors will be asked to report on aspects of their sexual behaviour over the last 3 months. How accurate do you think people will be when reporting their sexual behaviour over the last 3 months?

- Reports will be a complete guess
- Inaccurate
- Somewhat accurate
- Mostly accurate
- Very accurate

[Q35] On a scale of 1 to 7, where 1 = Not at all, and 7 = Completely, how would you describe the statement?

Q35.1 They have forgotten aspects of their previous sexual behaviour?

Q35.2 They feel ashamed to report on aspects of their sexual behaviour?

Q35.3 They feel embarrassed to report on aspects of their sexual behaviour?

Q35.4 They feel that the questions are too personal and do not wish to report on them?

Q35.5 They may think that the questions are not relevant as blood is tested anyway and decide not to report them?

To what extent do you think the following strategies will be effective and acceptable to help people remember their sexual behaviour more accurately over the last 3 months?

[Q36] Asking potential donors to use their smartphones to help them recall their activities over the last 3 months?

Q36.1 How acceptable do you think this would be?

Q36.2 How effective do you think this would be?

[Q37] Reminding potential donors that everyone is asked about the same sexual behaviours and that these behaviours are all common and current behaviours reported by people?

Q37.1 How acceptable do you think this would be?

Q37.2 How effective do you think this would be?

[Q38] When it becomes the UK policy to use a donor health check of this type (focused on every donor's sexual behaviour), what do you feel the level of infection risk to a patient receiving blood will be?

- No risk at all
- A small risk
- A large risk
- A very large risk
- A extremely large risk

[Q39] To what extent do you think this risk is acceptable? Please indicate your answer using the slider on a scale of 0 = Unacceptable, to 10 = Acceptable.

[Q40] Thank you for your participation in our survey. If you wish to find out more about blood donation in the UK, we have provided some links you might want to check out.

UK Blood Donation Links

Do you want to become a registered blood donor? Click here to sign up online and become a blood donor. Are you already a registered blood donor, and want to donate blood? Create your own online account or login, to manage and book your next donation.

Are you feeling charitable? Check out the YouGov official list of the most popular charities in the UK!

[Q41] Title: Evaluation of Self-Other Statements for Blood Donor Health Check

Students: Sarah Bowen: Sarah.bowen@nottingham.ac.uk

Supervisor: Eamonn Ferguson: Eamonn.ferguson@nottingham.ac.uk

Ethics reference: S1306

Rationale: This study is designed to examine people's attitudes and beliefs towards the upcoming change to the UK's blood donor selection policy.

Background Information: By way of some background information, the blood donor screening process will officially change to enable a more individualised way of assessing safe blood donations from Summer 2021.

- https://www.gov.uk/government/news/landmark-change-to-blood-donation-criteria

This policy change is evidenced by recommendations from the FAIR (For the Assessment of Individualised Risk) steering group.

- https://www.blood.co.uk/news-and-campaigns/news-and-statements/fair-steering-group/

Data Analysis: Responses will be analysed using standard statistic procedures to better understand how people may respond to the UK's blood donation policy change. If you would like to know more about the study or have any questions, please contact the project supervisor Eamonn Ferguson (Eamonn.ferguson@nottingham.ac.uk). If you have any concerns or complaints about the study, please contact Stephen Jackson (Ethics Committee Chair) Stephen.jackson@nottingham.ac.uk.

Please click on the arrow below to complete and submit the survey and thank you for your participation.

C.1.2 Structure of the Policy Communication Messages

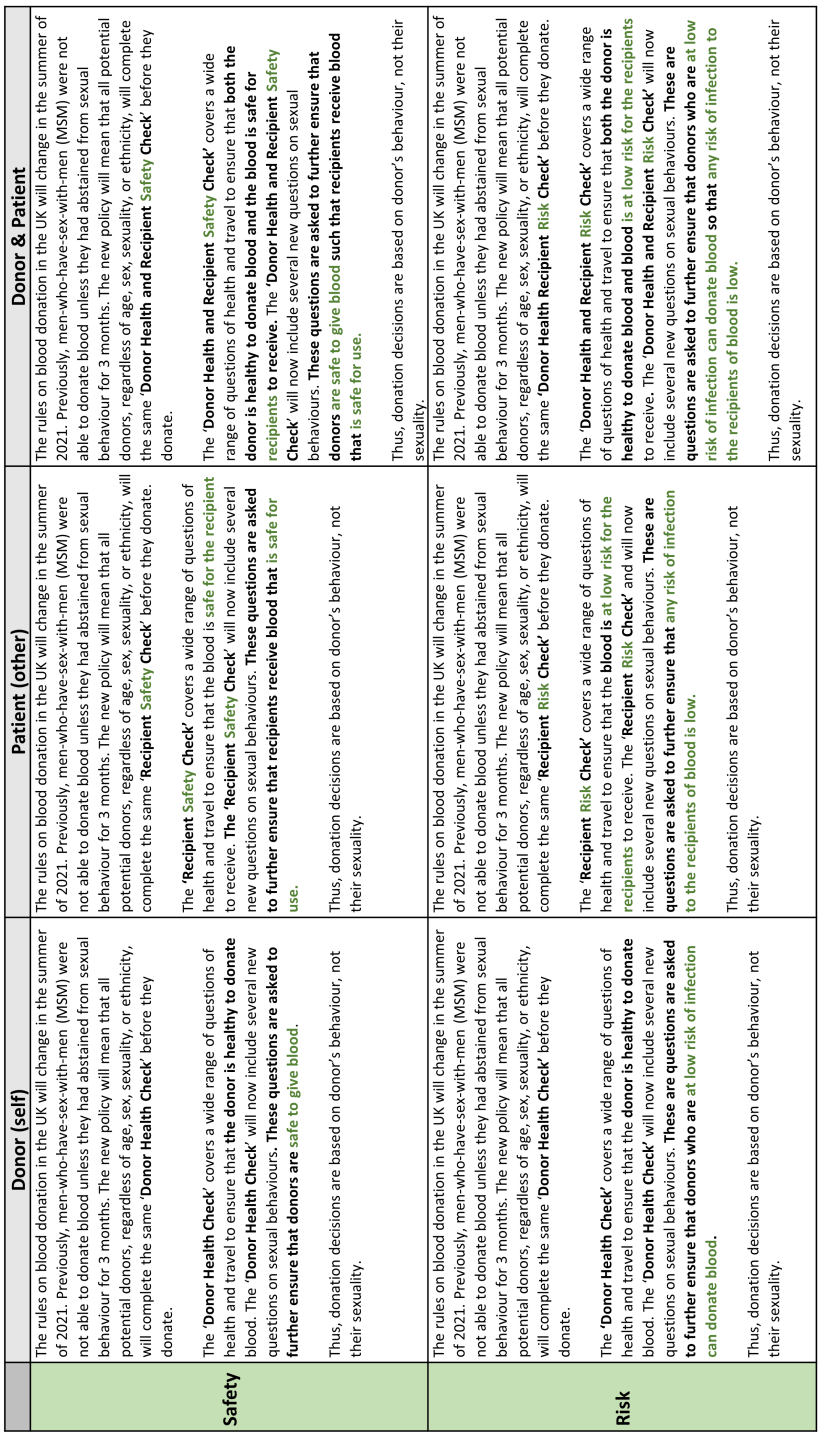


Figure C.1: Risk (risk vs safety) by Focus (donor, recipient, or both) frames

C.2 Additional Analysis

C.2.1 Sample structure

Below are the ONS categorisation used to code the study data. The white sample does not include white minorities (e.g, Gypsy, Roma or Irish Traveller groups or specified as "White other").

	n	Valid Percentage
Asexual	50	1.9
Bi-Curious	31	1.2
Bisexual	366	14.0
Gay	118	4.5
Heterosexual/Straight	1882	71.9
Lesbian	93	3.6
Pansexual	48	1.8
Queer	31	1.2
Missing	0	
Total n	2677	

Table C.1: Coding of sexuality

	n	Valid Percentage
People from Asian ethnic backgrounds		
Indian	135	5.06
Pakistani	69	2.58
Bangladeshi	58	2.17
Chinese	80	3.00
Any other Asian background, please describe	85	3.18
	427	
People Black and Caribbean backgrounds		
African	98	3.67
Caribbean	55	2.06
Any other Black/African/Caribbean background, please describe	22	0.82
	175	
People from mixed ethnic backgrounds		
White and Black Caribbean	37	1.39
White and Black African	15	0.56
White and Asian	79	2.96
Any other Mixed/Multiple ethnic backgrounds, please describe	83	3.11
	214	
Other ethnic groups-		
Arab	21	0.79
Any other ethnic group, please describe	30	1.12
	51	
White People		
English/Welsh/Scottish/Northern Irish/British	1.598	59.85
Irish	18	0.67
Gypsy or Irish Traveller	0	
Any other White background, please describe	177	6.63
	1793	
Prefer not to say	10	0.37
Missing	7	0.26
Total n	**2677**	

Table C.2: Coding of ONS ethnicity

C.2.2 Standardized scores

Normalizing scores The indices of Approach and Avoidance were normalized to between 0 and 1 using the following formulae. Where: $maxscore$ = maximum possible score on the scale, $minscore$ = minimum possible score on the scale, $score_i$ = the individual actual score on the scale.

$$NormalisedScore_i = \frac{score_i - minscore}{maxscore - minscore}$$

Creating normalized approach-avoidance indices These operationalize the relative strength of approach and avoidance an approach-avoidance index (AAI) to range from -1 to 1, where a score of -1 represents the stronger force towards avoidance, 0 represents the case when the forces cancel each other out, and 1 presents the stronger force towards approach.

$$SelfAAI_i = \frac{Approach_i - 0}{2 - 0} - \frac{SelfDefer_i - 2}{7 - 1}$$

$$NormativeAAI_i = \frac{Approach_i - 0}{2 - 0} - \frac{NormativeDefer_i - 2}{14 - 2}$$

C.2.3 Secondary Outcomes

We assessed several secondary outcomes ("safety", "fairness", and "potential mechanism to mitigate non-compliance").

Perceived safety of blood: The sum of two items indexes beliefs about the safety of the blood supply: (i) To what extent does this statement make you feel that blood is screened to ensure it is safe, and (ii) To what extent does the statement make you feel that the blood patients receive is safe (1 = "Not at all", to 7 = "Completely") ($r = 0.8195, p < 0.001$).

Perceived fairness and equality: This was indexed by the sum of 4-item asking to what extent the policy communication promotes (i) a sense of equality, (ii) a sense of fairness, (iii) will encourage a wider diversity of people to donate, and (iv) that donor selection feels fair (1 = "Not at all", to 7 = "Completely") (Cronbach's $\alpha = 0.717$).

Potential mechanism to mitigate non-compliance: participants to indicate the extent to which they felt asking people to use their mobile phone to aid recall over the last 3 months would be an effective strategy to increase compliance (1 = "Not at all", to 7 = "Completely").

C.2.4 Balance tests

The balance tests are presented in Tables C.3 (risk frames) and C.5 (altruism frames) for the full samples and Tables C.4 (risk frames) and C.6 (altruism frames) for samples minus the recipients of blood. There were no significant effects showing that the randomizationn worked and was maintained after the recipients of blood were removed.

	Risk		**Safety**		
	n	Mean (sd)	n	Mean (sd)	p
Age	1452	38.29 (15.64)	1452	38.61 (15.69)	0.584
Male	1428	0.43 (0.49)	1430	0.40 (0.49)	0.089
Ethnicity					
Asian	1449	0.16 (0.37)	1409	0.15 (0.36)	0.523
Black	1449	0.07 (0.26)	1409	0.07 (0.25)	0.638
Mixed	1449	0.07 (0.25)	1409	0.06 (0.25)	0.901
White	1449	0.69 (0.46)	1409	0.70 (0.46)	0.490
Current donor	1452	0.11 (0.32)	1453	0.10 (0.30)	0.164
Blood recipient	1444	0.07 (0.26)	1443	0.08 (0.26)	0.563
% LGBTQ+	1455	0.29 (0.45)	1452	0.30 (0.46)	0.416
% MSM	1452	0.08 (0.27)	1453	0.08 (0.28)	0.417
Prior infection belief	1439	20.10 (22.52)	1445	19.79 (22.46)	0.710

Note: p is the p-value of the frame manipulation coefficient in a simple linear regression of frame on the variable.

Table C.3: Balance tables across risk framing manipulation (full sample)

	Risk		**Safety**		
	n	Mean (sd)	n	Mean (sd)	p
Age	1,343	37.74 (15.43)	1,333	37.87 (15.43)	0.821
Male	1,322	0.44 (0.5)	1,313	0.4 (0.49)	0.052
Ethnicity					
Asian	1,340	0.17 (0.37)	1,330	0.15 (0.36)	0.306
Black	1,340	0.07 (0.25)	1,330	0.06 (0.24)	0.734
Mixed	1,340	0.08 (0.27)	1,330	0.08 (0.27)	0.954
White	1,340	0.67 (0.47)	1,330	0.68 (0.47)	0.572
Non-donor	1,340	0.65 (0.48)	1,330	0.66 (0.47)	0.637
Current donor	1,340	0.22 (0.41)	1,330	0.23 (0.42)	0.451
Lapsed donor	1,340	0.13 (0.33)	1,330	0.11 (0.31)	0.095
% LGBTQ+	1,337	0.29 (0.45)	1,333	0.30 (0.46)	0.285
% MSM	1,343	0.08 (0.27)	1,334	0.09 (0.28)	0.494
Prior infection belief	1,332	20.05 (12.63)	1,327	19.70 (22.17)	0.684

Note: p is the p-value of the frame manipulation coefficient in a simple linear regression of frame on the variable.

Table C.4: Balance tables across risk framing manipulation (no recipients)

	Donor		**Recipient**		**Both**		
	n	Mean (sd)	n	Mean (sd)	n	Mean (sd)	p
Age	969	38.57 (15.48)	964	38.57 (15.80)	971	38.21 (15.72)	0.846
Male	953	0.42 (0.49)	949	0.39 (0.49)	956	0.43 (0.50)	0.183
Ethnicity							
Asian	946	0.16 (0.36)	941	0.15 (0.36)	941	0.17 (0.37)	0.663
Black	946	0.07 (0.26)	941	0.08 (0.27)	941	0.06 (0.23)	0.240
Mixed	946	0.08 (0.27)	941	0.08 (0.27)	941	0.09 (0.28)	0.787
White	946	0.69 (0.46)	941	0.70 (0.46)	941	0.69 (0.46)	0.980
Current donor	970	0.11 (0.31)	964	0.10 (0.30)	971	0.11 (0.31)	0.621
Blood recipient	967	0.07 (0.25)	958	0.07 (0.25)	962	0.08 (0.27)	0.472
% LGBTQ+	966	0.28 (0.45)	960	0.30 (0.46)	971	0.30 (0.46)	0.606
% MSM	970	0.08 (0.27)	964	0.08 (0.26)	971	0.08 (0.28)	0.774
Prior infection belief	961	20.48 (23.61)	958	19.82 (21.84)	965	19.54 (21.97)	0.641

Note: p is the p-value of an F-test of equality across statement altruism manipulation coefficients.

Table C.5: Balance tables across altruism framing manipulation (full sample)

	Donor		Recipient		Both		
	n	Mean (sd)	n	Mean (sd)	n	Mean (sd)	p
Age	899	38.06 (15.30)	893	27.96 (15.58)	884	37.39 (15.42)	0.617
Male	883	0.43 (0.49)	879	0.4 (0.49)	873	0.44 (0.5)	0.185
Ethnicity							
Asian	897	0.16 (0.36)	890	0.16 (0.36)	883	0.17 (0.37)	0.748
Black	897	0.07 (0.26)	890	0.07 (0.26)	883	0.05 (0.23)	0.260
Mixed	897	0.08 (0.27)	890	0.08 (0.27)	883	0.09 (0.28)	0.725
White	897	0.67 (0.47)	890	0.68 (0.47)	883	0.66 (0.47)	0.872
Non-donor	897	0.66 (0.47)	890	0.67 (0.47)	883	0.63 (0.48)	0.187
Current donor	897	0.21 (0.41)	890	0.22 (0.42)	883	0.24 (0.43)	0.167
Lapsed donor	897	0.13 (0.34)	890	0.1 (0.3)	883	0.12 (0.33)	0.221
% LGBTQ+	896	0.28 (0.45)	890	0.3 (0.46)	884	0.3 (0.46)	0.533
% MSM	900	0.08 (0.28)	893	0.08 (0.27)	884	0.08 (0.28)	0.920
Prior infection belief	891	20.44 (23.58)	889	19.71 (21.65)	879	19.78 (21.92)	0.639

Note: p is the p-value of an F-test of equality across statement altruism manipulation coefficients.

Table C.6: Balance tables across altruism framing manipulation (no recipients)

C.2.5 Altruistic frames manipulation checks

In terms of focus, across all treatments, participants reported the statement focused on the donor slightly more than the recipient ($M = 3.42, CI_{95\%} = [3.33, 3.51]$) (Table C.7).[1]

In terms of salience, participants reported to think about the patient more in the combined donor-recipient condition ($M = 4.76, CI_{95\%} = [4.65, 4.86]$) than to the donor only ($M = 4.58, CI_{95\%} = [4.47, 6.69], p = 0.021$) and the recipient only condition ($M = 4.56, CI_{95\%} = [4.44, 4.67], p = 0.012$) (Table C.7).[2] Overall, on average participants reported to think more about the patient than the donor ($M = 4.63, CI_{95\%} = [4.57, 4.69], t_{one-sample} =$

[1] Focus assessed with the following question: Who is the focus of the statement? On a 11-point scale from 0 "the donor only" to 10 "the patient only"

[2] Salience assessed with the following question: To what extent does the statement make you think about the patients who receive blood? On a 7-point scale from 1 "Not at all", to 7 "Completely".

$37.73; p < 0.0001$)

The extent of thinking about the parient was not significantly different compared to the mid-point (3.5) for the 'Donor' frame ($t_{one-sample} = -0.952; p = 0.341$) or the 'recipient' frame ($t_{one-sample} = -1.250; p = 0.212$) but was for the 'donor and recipient' frame ($t_{one-sample} = 2.346; p = 0.019$). Thus, we find the reader thought about the recipient more in the 'donor and recipient' altruism frame.

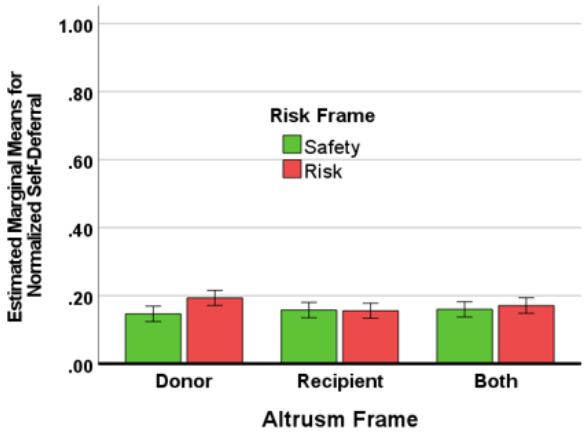
	Mean	CI [95%]		p
Focus (-5 Donor to +5 Recipient)				
Overall	**3.42**	**3.33, 3.51**		
1. Donor	3.37	3.21, 3.42	1 v 2	0.843
2. Recipient	3.40	3.23, 3.56	1 v 3	0.317
3. Both	3.49	3.33, 3.65	2 v 3	0.423
Salience of the recipient				
Overall	**4.63**	**4.57, 4.69**		
1. Donor	4.58	4.47, 4.69	1 v 2	0.809
2. Recipient	4.56	4.44, 4.67	**1 v 3**	**0.021**
3. Both	4.76	4.65, 4.86	**2 v 3**	**0.012**

Table C.7: Focus manipulation check

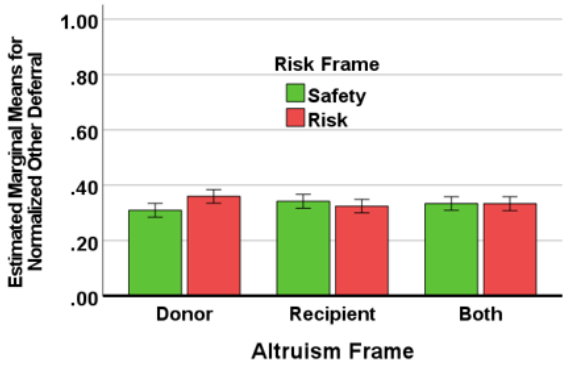
C.2.6 Margins analysis of approach-avoidance interactions

The margin analyses for the interactions observed in Table 3.3 in the main text are given in Table C.8 and Figure C.2 below. The results show that the combination of a risk and donor-focused frame leads to higher levels of consideration of feelings of self-deter, feelings that others would be deterred, and normative deter.

(a) Self-deter



(b) Other-deter



(c) Normative-deter

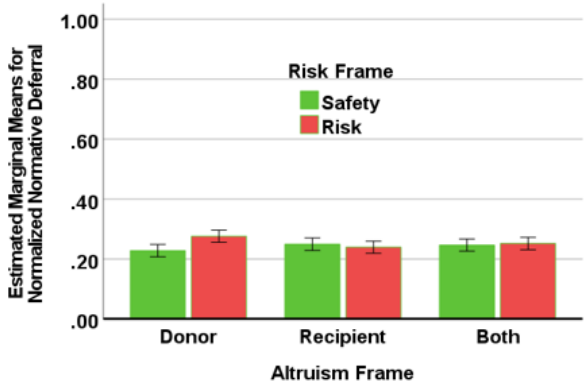


Figure C.2: The interactions of risk by altruism frame by deter scores

	Delta-method											
	Self-deferral				Other-Deferral				Normative Deferral			
	dy/dx	P=	95% interval		dy/dx	P=	95% interval		dy/dx	P=	95% interval	
	se				se				se			
			Lower	Upper			Lower	Upper			Lower	Upper
Donor Frame *(baseline)*												
Recipient Frame												
Risk Frame												
Safety	0.0107 (0.0162)	.508	-0.0210	0.0424	0.0327 (0.0179)	.069	-0.0025	0.0679	0.0216 (0 .0147)	.143	-0.0073	0.0504
Risk	**-0.0367 (0.0158)**	**.020**	**-0.0677**	**-0.0057**	**-0.0351 (0.0176)**	**.046**	**-0.0696**	**-0.0006**	**-0.0360 (0.0144)**	**.012**	**-0.0643**	**-0.0078**
Donor and Recipient Frame												
Risk Frame												
Safety	0.01278 (0.01601)	.425	-0.0186	0.0442	0.0245 (0.0178)	.168	-0.0104	0.0594	0.0185 (0.0146)	.204	-0.0100	0.0471
Risk	-0.0208 (0.0160)	.195	-0.0522	0.0107	-0.0248 (0.0178)	.164	-0.0598	0.0101	-0.0228 (0.0146)	.119	-0.0514	0.0058

Table C.8: Margins for risk frame by altrusim frame interactions

C.2.7 Regression models for awareness of mechanisms linked to the anticipation of under-reporting sexual behaviour

Table C.9 details the OLS regression analyses for the predictors of the three anticipated mechanisms associated with the under-reporting of sexual behaviour.

	(1) Forgetting		(2) Negative Emotions		(3) Irrelevance	
	β	CI [95%]	β	CI [95%]	β	CI [95%]
Risk						
Safety	0.02	[-0.189,0.229]	−0.024	[-0.188,0.141]	0.106	[-0.104,0.315]
Altruism						
Recipient	0.027	[-0.180,0.235]	0.105	[-0.058,0.268]	0.054	[-0.153,0.261]
Both	0.069	[-0.141,0.279]	0.056	[-0.108,0.221]	−0.032	[-0.242,0.179]
Risk × *Altruism*						
Safety × Recipient	−0.050	[-0.346,0.246]	0.035	[-0.197,0.267]	−0.107	[-0.403,0.189]
Safety × Both	−0.091	[-0.388,0.206]	0.018	[-0.215,0.251]	0.027	[-0.270,0.324]
Controls						
Age	**-0.010*** **	[-0.015,-0.005]	**-0.005** **	[-0.009,-0.002]	0.002	[-0.003,0.007]
Male	**-0.300*** **	[0.425,-0.176]	**-0.246*** **	[-0.344,-0.149]	**-0.443*** **	[-0.568,-0.319]
LGBQ+	0.069	[-0.081,0.218]	−0.104	[-0.222,0.013]	−0.081	[-0.231,0.068]
Ethnicity						
Asian	**0.348*** **	[0.167,0.529]	0.075	[-0.067,0.217]	0.123	[-0.058,0.304]
Black	**0.322***	[0.067,0.576]	**0.328** **	[0.128,0.529]	**0.303***	[0.048,0.558]
Mixed	0.148	[-0.083,0.380]	0.174	[-0.007,0.356]	0.074	[-0.158,0.305]
Donor status						
Lapsed donor	−0.131	[-0.289,0.026]	0.031	[-0.093,0.154]	−0.097	[-0.255,0.060]
Current donor	−0.129	[-0.321,0.062]	**-0.190***	[-0.341,-0.040]	−0.148	[-0.340,0.043]
Constant	4.434***	[4.165,4.703]	5.611***	[5.400,5.822]	5.064***	[4.796,5.333]
R^2	0.040		0.029		0.026	
N	2552		2548		2551	

Note: * p < 0.05, ** p < 0.01, *** p < 0.001. Baselines groups: Risk = risk frame; Altrusim = donor frame; Ethnicity = people from white ethnicity backgrounds; Donor status = non-donors. Coefficients are unstandardized.

Table C.9: OLS regressions for swareness of mechanisms leading to under-reporting of sexual behaviours

C.2.8 Safety and fairness regression models

There were no significant framing effects on perceived safety to the blood supply or equality and fairness of the policy (Table C.10). However, there were demographic effects. Older respondents, and respondents from Asian, Black , and Mixed ethnic backgrounds relative to white respondents, reported lower perceived safety and fairness. Lapsed donors and current donors reported higher perceived safety scores than non-donors. Also, LGBQ+ respondents reported greater perceptions of safety than straight identifying respondents. Lapsed donors reported higher perceived fairness than non-donors.

	(1) Perceived safety		(2) Fairness and Equality	
	β	CI [95%]	β	CI [95%]
Risk				
Safety	0.038	[-0.311,0.387]	0.092	[-0.347,0.531]
Altruism				
Recipient	−0.141	[0.486,0.205]	−0.150	[-0.585,0.284]
Both	−0.144	[-0.495,0.206]	−0.106	[-0.547,0.334]
Risk × *Altruism*				
Safety x Recipient	−0.041	[-0.535,0.452]	0.044	[-0.577,0.665]
Safety x Both	−0.091	[-0.586,0.405]	0.003	[-0.620,0.626]
Controls				
Age	**-0.020*** **	[-0.027,-0.012]	**-0.013***	[-0.023,-0.003]
Male	**-0.212***	[-0.420,-0.004]	**-0.666*** **	[-0.927,-0.405]
LGBQ+	**0.656*** **	[0.407,0.906]	0.084	[-0.230,0.398]
Ethnicity				
Asian	**-0.821*** **	[-1.122,-0.519]	**-0.963*** **	[-1.343,-0.584]
Black	**-0.740*** **	[-1.165,-0.315]	−0.181	[-0.715,0.354]
Mixed	**-0.398***	[-0.784,-0.012]	**-0.733****	[-1.218,-0.247]
Donor status				
Lapsed donor	**0.576*** **	[0.313,0.839]	**0.502****	[0.172,0.833]
Current donor	**0.941*** **	[0.622,1.261]	0.262	[-0.140,0.664]
Constant	11.921***	[11.473,12.368]	25.148***	[24.584,25.711]
R^2	0.065		0.029	
N	2551		2553	

Note: * p < 0.05, ** p < 0.01, *** p < 0.001. Baselines groups: Risk = risk frame; Altrusim = donor frame; Ethnicity = people from white ethnicity backgrounds; Donor status = non-donors. Coefficients are unstandardized.

Table C.10: OLS regressions on perceived safety and fairness/equality.

References

ABO (2022). The Risk-Based Decision-Making Framework.

Adler, N., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology*, *19*(6), 586–592.

Adler, N., & Stewart, J. (2007). The MacArthur Scale of Subjective Social Status. *San Francisco: MacArthur Research Network on SES & Health*.

Ágh, T., van Boven, J. F., Wettermark, B., Menditto, E., Pinnock, H., Tsiligianni, I., Petrova, G., Potočnjak, I., Kamberi, F., & Kardas, P. (2021). A Cross-Sectional Survey on Medication Management Practices for Noncommunicable Diseases in Europe During the Second Wave of the COVID-19 Pandemic. *Frontiers in Pharmacology*, *12*, 1433.

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In *Action Control*, (pp. 11–39). Berlin, Heidelberg: Springer Berlin Heidelberg.

Akerlof, K., & Kennedy, C. (2013). Nudging toward a Healthy Natural Environment. How behavioral change research can inform conservation. *George Mason University*, (p. 90).

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9-10), 1082–1095.

Antin, J., & Shaw, A. (2012). Social desirability bias and self-reports of motivation. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, (p. 2925). New York, New York, USA: ACM Press.

Appel, M., Gnambs, T., & Maio, G. R. (2012). A short measure of the need for affect. *Journal of Personality Assessment*, *94*(4), 418–426.

Aquino, A., Alparone, F. R., Pagliaro, S., Haddock, G., Maio, G. R., Perrucci, M. G., & Ebisch, S. J. (2020). Sense or sensibility? The neuro-functional basis of the structural matching effect in persuasion. *Cognitive, Affective and Behavioral Neuroscience*, *20*(3), 536–550.

Arnet, I., Walter, P. N., & Hersberger, K. E. (2013). Polymedication electronic monitoring system (POEMS) - a new technology for measuring adherence. *Frontiers in Pharmacology*, *4 MAR*, 26.

Aubé, E., Lewin, A., O'Brien, S. F., Grégoire, Y., Pillonel, J., Steele, W. R., Custer, B., Davison, K. L., Germain, M., Seed, C. R., & Camirand Lemyre, F. (2021). HIV residual risk in Canada for apheresis source plasma donation without deferral for men who have sex with men. *Vox Sanguinis*.

Bach, D. R., Guitart-Masip, M., Packard, P. A., Miró, J., Falip, M., Fuentemilla, L., & Dolan, R. J. (2014). Human hippocampus arbitrates approach-avoidance conflict. *Current Biology*, *24*(5), 541–547.

Bae, S. G., Kam, S., Park, K. S., Kim, K. S. K. Y., Hong, N. S., Kim, K. S. K. Y., Lee, Y. M., Lee, W. K., & Choe, M. S. P. (2016). Factors related to intentional and unintentional medication nonadherence in elderly patients with hypertension in rural community. *Patient Preference and Adherence*, *10*, 1979–1989.

Banerjee, S., & John, P. (2021). Nudge plus: incorporating reflection into behavioral public policy. *Behavioural Public Policy*, (pp. 1–16).

Bao, J., & Ho, B. (2015). Heterogeneous Effects of Informational Nudges on Pro-social Behavior. *B.E. Journal of Economic Analysis and Policy*, *15*(4), 1619–1655.

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.

Beard, E., West, R., Lorencatto, F., Gardner, B., Michie, S., Owens, L., & Shahab, L. (2019). What do cost-effective health behaviour-change interventions contain? A comparison of six domains. *PLoS ONE*, *14*(4).

Belli, R. F., Traugott, M. W., Young, M., & McGonagle, K. A. (1999). Reducing vote overreporting in surveys: Social desirability, memory failure, and source monitoring. *Public Opinion Quarterly*, *63*(1), 90–108.

Berkowitz, L., & Harmon-Jones, E. (2004). Toward an understanding of the determinants of anger.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., & Milkman, K. L. (2015). The Effect of Providing Peer Information on Retirement Savings Decisions. *Journal of Finance*, *70*(3), 1161–1201.

Betsch, C., Böhm, R., & Korn, L. (2013). Inviting free-riders or appealing to prosocial behavior? Game-theoretical reflections on communicating herd immunity in vaccine advocacy. *Health Psychology*, *32*(9), 978–985.

Betsch, C., Böhm, R., Korn, L., & Holtmann, C. (2017). On the benefits of explaining herd immunity in vaccine advocacy. *Nature Human Behaviour*, *1*(3), 56.

Bicchieri, C. (2006). *The Grammar of Society*. Cambridge University Press.

Boulware, L. E., Ratner, L. E., Cooper, L. A., Sosa, J. A., LaVeist, T. A., & Powe, N. R. (2002). Understanding disparities in donor behavior: Race and gender differences in willingness to donate blood and cadaveric organs. *Medical Care*, *40*(2), 85–95.

Braverman, J. (2008). Testimonials versus informational persuasive messages: The moderating effect of delivery mode and personal involvement. *Communication Research*, *35*(5), 666–694.

Brehm, S., & Brehm, J. (2013). *Psychological reactance: A theory of freedom and control*. Academic Press.

Brenner, P. S., & DeLamater, J. (2016). Lies, Damned Lies, and Survey Self-Reports? Identity as a Cause of Measurement Bias. *Social Psychology Quarterly*, *79*(4), 333–354.

Brenner, P. S., & DeLamater, J. D. (2014). Social Desirability Bias in Self-reports of Physical Activity: Is an Exercise Identity the Culprit? *Social Indicators Research*, *117*(2), 489–504.

Brewer, N. T., Chapman, G. B., Rothman, A. J., Leask, J., & Kempe, A. (2017). Increasing Vaccination: Putting Psychological Science Into Action. *Psychological Science in the Public Interest*, *18*(3), 149–207.

Brian Haynes, R., Wayne Taylor, D., Sackett, D. L., Gibson, E. S., Bernholz, C. D., & Mukherjee, J. (1980). Can simple clinical measurements detect patient noncompliance? *Hypertension*, *2*(6), 757–764.

Burnier, M. (2019). Is there a threshold for medication adherence? Lessons learnt from electronic monitoring of drug adherence.

Busch, M. P., Glynn, S. A., Stramer, S. L., Strong, D. M., Caglioti, S., Wright, D. J., Pappalardo, B., & Kleinman, S. H. (2005). A new strategy for estimating risks of transfusion-transmitted viral infections based on rates of detection of recently infected donors. *Transfusion*, *45*(2), 254–264.

Byrne, S., & Hart, P. S. (2009). The Boomerang Effect A Synthesis of Findings and a Preliminary Theoretical Framework. *Annals of the International Communication Association*, *33*(1), 3–37.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology*, *42*(1), 116.

Caffrey, N., Goldman, M., Osmond, L., Yi, Q. L., Fan, W., & O'Brien, S. F. (2022). HIV incidence and compliance with deferral criteria over three progressively shorter time deferrals for men who have sex with men in Canada. *Transfusion*, *62*(1), 125–134.

Cambridge University Press (n.d.). Message. *Retrieved September 24, 2022 from https://dictionary.cambridge.org/dictionary/english/message*.

Campos, S., Doxey, J., & Hammond, D. (2011). Nutrition labels on pre-packaged foods: A systematic review.

Canadian Blood Services (2022). Canadian Blood Services to remove eligibility criteria specific to men who have sex with men. *Canadian Blood Services*.

Carey, R. N., Connell, L. E., Johnston, M., Rothman, A. J., De Bruin, M., Kelly, M. P., & Michie, S. (2018). Behavior Change Techniques and Their Mechanisms of Action: A Synthesis of Links Described in Published Intervention Literature. *Annals of Behavioral Medicine*, *53*(8), 693–707.

Caruso, J., Germain, M., Godin, G., Myhal, G., Pronovost, F., Morin, M., & Otis, J. (2019). â€˜One step closer': Acceptability of a programme of plasma donation for fractionation from men who have sex with men. *Vox Sanguinis*, *114*(7), 675–686.

Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and Emotion*, *30*(2), 105–110.

Carver, C. S., & Harmon-Jones, E. (2009). Anger Is an Approach-Related Affect: Evidence and Implications. *Psychological Bulletin*, *135*(2), 183–204.

Chaiken, S. (1987). The heuristic model of persuasion BT - Social influence: The Ontario symposium. *Social influence: The Ontario symposium*, *5*, 3–39.

Chan, A. H. Y., Horne, R., Hankins, M., & Chisari, C. (2020). The Medication Adherence Report Scale: A measurement tool for eliciting patients' reports of nonadherence. *British Journal of Clinical Pharmacology*, *86*(7).

Chapman, G. B., Li, M., Vietri, J., Ibuka, Y., Thomas, D., Yoon, H., & Galvani, A. P. (2012). Using Game Theory to Examine Incentives in Influenza Vaccination Behavior. *Psychological Science*, *23*(9), 1008–1015.

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.

Clarke, P. M., Fiebig, D. G., & Gerdtham, U. G. (2008). Optimal recall length in survey design. *Journal of Health Economics*, *27*(5), 1275–1284.

Clement, J., Jacobi, M., & Greenwood, B. N. (2021). Patient access to chronic medications during the Covid-19 pandemic: Evidence from a comprehensive dataset of US insurance claims. *PLoS ONE*, *16*(4 April), e0249453.

Corr, P. J. (2013). Approach and avoidance behaviour: Multiple systems and their interactions. *Emotion Review*, *5*(3), 285–290.

Cross, A. J., Elliott, R. A., Petrie, K., Kuruvilla, L., & George, J. (2020). Interventions for improving medication-taking ability and adherence in older adults prescribed multiple medications.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment.

Cutts, J. C., Quinn, B., Seed, C. R., Kotsiou, G., Pearson, R., Scott, N., Wilson, D. P., Harrod, M. E., Maher, L., Caris, S., Thompson, A. J., Farrell, M., Pink, J., & Hellard, M. E. (2021). A Systematic Review of Interventions Used to Increase Blood Donor Compliance with Deferral Criteria. *Transfusion Medicine and Hemotherapy*, *48*(2), 118–129.

Dai, H., Mao, D., Riis, J., Volpp, K. G., Relish, M. J., Lawnicki, V. F., & Milkman, K. L. (2017). Effectiveness of Medication Adherence Reminders Tied to "Fresh Start" Dates. *JAMA Cardiology*, *2*(4), 453.

Dalziel, K., Li, J., Scott, A., & Clarke, P. (2018). Accuracy of patient recall for self-reported doctor visits: Is shorter recall better? *Health Economics (United Kingdom)*, *27*(11), 1684–1698.

Davison, K. L., Reynolds, C. A., Andrews, N., Brailsford, S. R., Kohli, H., Carter, M., Field, S., Miflin, G., Murdock, J., Maguire, K., & Wickenden, C. (2021). Blood donation by men who have sex with men: using evidence to change policy. *Vox Sanguinis*, *116*(3), 260–272.

De Geest, S., & Sabaté, E. (2003). Adherence to long-term therapies: Evidence for action.

De Kort, W., Mayr, W., Jungbauer, C., Vuk, T., Kullaste, R., Seifried, E., Grazzini, G., De Wit, J., & Folléa, G. (2016). Blood donor selection in European Union directives: Room for improvement. *Blood Transfusion*, *14*(2), 101–108.

Dempsey, R. C., McAlaney, J., & Bewick, B. M. (2018). A critical appraisal of the social norms approach as an interventional strategy for health-related behavior and attitude change.

Dhruve, H., D'Ancona, G., Holmes, S., Dhariwal, J., Nanzer, A. M., & Jackson, D. J. (2022). Prescribing Patterns and Treatment Adherence in Patients with Asthma During the COVID-19 Pandemic. *Journal of Allergy and Clinical Immunology: In Practice*, *10*(1), 100–107.e2.

Dunlop, S. M., Wakefield, M., & Kashima, Y. (2010). Pathways to persuasion: Cognitive and experiential responses to health-promoting mass media messages. *Communication Research*, *37*(1), 133–164.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, *34*(3), 169–189.

Elliott, R. A., Boyd, M. J., Salema, N.-E. E., Davies, J., Barber, N., Mehta, R. L., Tanajewski, L., Waring, J., Latif, A., Gkountouras, G., Avery, A. J., Chuter, A., & Craig, C. (2016). Supporting adherence for people starting a new medication for a long-term

condition through community pharmacies: a pragmatic randomised controlled trial of the New Medicine Service. *BMJ quality & safety*, *25*(10), 747–758.

Esposti, L. D., Buda, S., Nappi, C., Paoli, D., & Perrone, V. (2020). Implications of covid-19 infection on medication adherence with chronic therapies in italy: A proposed observational investigation by the fail-to-refill project.

FAIR Steering Group (2020). Can donor selection policy move from a population-based donor selection policy to one based on a more individualised risk assessment? Conclusions from the For the Assessment of Individualised Risk (FAIR) group. Tech. Rep. December.

Ferguson, E., & Chandler, S. (2005). A stage model of blood donor behaviour: Assessing volunteer behaviour. *Journal of Health Psychology*, *10*(3), 359–372.

Ferguson, E., Dawe-Lane, E., Khan, Z., Reynolds, C., Davison, K., Edge, D., & Brailsford, S. R. (2022). Trust and distrust: Identifying recruitment targets for ethnic minority blood donors. *Transfusion Medicine*.

Ferraro, P. J., & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics*, *95*(1), 64–73.

Fisayo, T. (2021). Science in action? A critical view of UK blood donation deferral policy and men who have sex with men. *International Journal of Health Planning and Management*, *36*(4), 1207–1222.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404.

Fisher, R., & Tellis, G. (1998). Removing Social Desirability Bias With Indirect Questioning: Is the Cure Worse Than the Disease? *Advances in Consumer Research*, *25*(10), 563–567.

Fisher, R. J. (1993). Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, *20*(2), 303.

Furnham, A., & Henderson, M. (1983). Response bias in self-report measures of general health. *Personality and Individual Differences*, *4*(5), 519–525.

Furniss, D., Barber, N., Lyons, I., Eliasson, L., & Blandford, A. (2014). Unintentional non-Adherence: Can a spoon full of resilience help the medicine go down? *BMJ Quality and Safety*, *23*(2), 95–98.

Gadkari, A. S., & McHorney, C. A. (2012). Unintentional non-adherence to chronic prescription medications: How unintentional is it really? *BMC Health Services Research*, *12*(1), 98.

Gannon, J. M., Conlogue, J., Sherwood, R., Nichols, J., Ballough, J. R., Fredrick, N. M., & Chengappa, K. N. (2020). Long acting injectable antipsychotic medications: Ensuring care continuity during the COVID-19 pandemic restrictions. *Schizophrenia Research*, *222*, 532–533.

Garcia-Marcos, P. W., Brand, P. L., Kaptein, A. A., & Klok, T. (2016). Is the MARS questionnaire a reliable measure of medication adherence in childhood asthma? *Journal of Asthma*, *53*(10), 1085–1089.

Germain, M. (2016). The risk of allowing blood donation from men having sex with men after a temporary deferral: Predictions versus reality. *Transfusion*, *56*(6), 1603–1607.

Gerrard, M., Gibbons, F. X., Houlihan, A. E., Stock, M. L., & Pomery, E. A. (2008). A dual-process approach to health risk decision making: The prototype willingness model. *Developmental Review*, *28*(1), 29–61.

Glasgow, R. E., Knoepke, C. E., Magid, D., Grunwald, G. K., Glorioso, T. J., Waughtal, J., Marrs, J. C., Bull, S., & Ho, P. M. (2021). The NUDGE trial pragmatic trial to enhance cardiovascular medication adherence: study protocol for a randomized controlled trial. *Trials*, *22*(1), 1–16.

Gnambs, T., & Kaspar, K. (2016). Socially Desirable Responding in Web-Based Questionnaires: A Meta-Analytic Review of the Candor Hypothesis. *Assessment*, *24*(6), 746–762.

Godinho, C. A., Alvarez, M. J., & Lima, M. L. (2016). Emphasizing the losses or the gains: Comparing situational and individual moderators of framed messages to promote fruit and vegetable intake. *Appetite*, *96*, 416–425.

Goldman, M., Shih, A. W., O'Brien, S. F., & Devine, D. (2018). Donor deferral policies for men who have sex with men: past, present and future. *Vox Sanguinis*, *113*(2), 95–103.

Gordon, L., Graves, N., Hawkes, A., & Eakin, E. (2007). A review of the cost-effectiveness of face-to-face behavioural interventions for smoking, physical activity, diet and alcohol. *Chronic Illness*, *3*(2), 101–129.

Gray, J. A., & McNaughton, N. (2000). *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*. Oxford: Oxford University Press, 2nd edn ed.

Grimm, P. (2010). Social Desirability Bias. In *Wiley International Encyclopedia of Marketing*. John Wiley & Sons, Ltd.

Gunzler, D., Chen, T., Wu, P., & Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai Archives of Psychiatry*, *25*(6), 390–394.

Haddock, G., Maio, G. R., Arnold, K., & Huskinson, T. (2008). Should persuasion be affective or cognitive? The moderating effects of need for affect and need for cognition. *Personality and Social Psychology Bulletin*, *34*(6), 769–778.

Haidt, J. (2003). *The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.)*, vol. Handbook o. Oxford University Press.

Haidt, J. (2007). The new synthesis in moral psychology.

Haire, B., Whitford, K., & Kaldor, J. M. (2018). Blood donor deferral for men who have sex with men: still room to move. *Transfusion*, *58*(3), 816–822.

Hall, M. G., Sheeran, P., Noar, S. M., Ribisl, K. M., Boynton, M. H., & Brewer, N. T. (2017). A brief measure of reactance to health warnings. *Journal of Behavioral Medicine*, *40*(3), 520–529.

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PLoS ONE*, *11*(8), e0159915.

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *SAGE Open*, *5*(4), 215824401562111.

Haynes, R. B., McDonald, H., Garg, A., & Montague, P. (2002). Interventions for helping patients to follow prescriptions for medications. In *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd.

Heinrich, C., & Kuiper, R. A. (2012). Using Handheld Devices to Promote Medication Adherence in Chronic Illness. *The Journal for Nurse Practitioners*, *8*(4), 288–293.

Hill, N. L., Mogle, J., Whitaker, E. B., Gilmore-Bykovskyi, A., Bhargava, S., Bhang, I. Y., Sweeder, L., Tiwari, P. A., & Van Haitsma, K. (2019). Sources of Response Bias in Cognitive Self-Report Items: "which Memory Are You Talking About?". *Gerontologist*, *59*(5), 912–924.

Horne, R., & Weinman, J. (2002). Self-regulation and self-management in asthma: Exploring the role of illness perceptions and treatment beliefs in explaining non-adherence to preventer medication. *Psychology and Health*, *17*(1), 17–32.

Hull, C. L. (1952). *Behavior System: An introduction to Behaviour Theory concerning the individual organinces*. New Haven: Yale University Press.

Hume, S., John, P., Sanders, M., & Stockdale, E. (2020). Nudge in the Time of Coronavirus: The Persistence of Behavioural Messages during Crisis. *SSRN Electronic Journal*.

Isler, O., Isler, B., Kopsacheilis, O., & Ferguson, E. (2020). Limits of the social-benefit motive among high-risk patients: A field experiment on influenza vaccination behaviour. *BMC Public Health*, *20*(1).

Jachimowicz, J. M., Gladstone, J. J., Berry, D., Kirkdale, C. L., Thornley, T., & Galinsky,

A. D. (2019). Making medications stick: improving medication adherence by highlighting the personal health costs of non-compliance. *Behavioural Public Policy*, *5*(396-416).

Jimmy, B., & Jose, J. (2011). Patient medication adherence: Measures in daily practice.

Jo, M. S., Nelson, J. E., & Kiecker, P. (1997). A model for controlling social desirability bias by direct and indirect questioning. *Marketing Letters*, *8*(4), 429–437.

Johnson, M. J. (2002). The Medication Adherence Model: a guide for assessing medication taking. *Research and theory for nursing practice*, *16*(3), 179–92.

Josephson, C. D., Su, L. L., Hillyer, K. L., & Hillyer, C. D. (2007). Transfusion in the Patient With Sickle Cell Disease: A Critical Review of the Literature and Transfusion Guidelines. *Transfusion Medicine Reviews*, *21*(2), 118–133.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263 – 291.

Kamal, A. K., Shaikh, Q., Pasha, O., Azam, I., Islam, M., Memon, A. A., Rehman, H., Akram, M. A., Affan, M., Nazir, S., Aziz, S., Jan, M., Andani, A., Muqeet, A., Ahmed, B., & Khoja, S. (2015). A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored Short Messaging Service (SMS)-SMS4Stroke study. *BMC Neurology*, *15*(1), 1–11.

Karamitros, G., Kitsos, N., & Karamitrou, I. (2017). The ban on blood donation on men who have sex with men: Time to rethink and reassess an outdated policy. *Pan African Medical Journal*, *27*, 1937–8688.

Kaye, L., Theye, B., Smeenk, I., Gondalia, R., Barrett, M. A., & Stempel, D. A. (2020). Changes in medication adherence among patients with asthma and COPD during the COVID-19 pandemic. *Journal of Allergy and Clinical Immunology: In Practice*, *8*(7), 2384–2385.

Kesby, M., & Sothern, M. (2014). Blood, sex and trust: The limits of the population-based risk management paradigm. *Health and Place*, *26*, 21–30.

Khabbazi, A., Kavandi, H., Paribanaem, R., Khabbazi, R., & Malek Mahdavi, A. (2020). Adherence to medication in patients with rheumatic diseases during COVID-19 pandemic. *Annals of the Rheumatic Diseases*, *0*(0).

Kim, S. C., Pei, D., Kotcher, J. E., & Myers, T. A. (2021). Predicting Responses to Climate Change Health Impact Messages From Political Ideology and Health Status: Cognitive Appraisals and Emotional Reactions as Mediators. *Environment and Behavior*, *53*(10), 1095–1117.

Kimmel, S. E., Troxel, A. B., Loewenstein, G., Brensinger, C. M., Jaskowiak, J., Doshi, J. A., Laskin, M., & Volpp, K. (2012). Randomized trial of lottery-based incentives to improve warfarin adherence. *American Heart Journal*, *164*(2), 268–274.

Kjellsson, G., Clarke, P., & Gerdtham, U. G. (2014). Forgetting to remember or remembering to forget: A study of the recall period length in health care survey questions. *Journal of Health Economics*, *35*(1), 34–46.

Kleinsinger, F. (2018). The Unmet Challenge of Medication Nonadherence. *The Permanente journal*, *22*, 18–033.

Kramer, K., Zaaijer, H. L., & Verweij, M. F. (2017). The Precautionary Principle and the Tolerability of Blood Transfusion Risks. *American Journal of Bioethics*, *17*(3), 32–43.

Kruglanski, A. W., Chen, X., Pierro, A., Mannetti, L., Erb, H. P., & Spiegel, S. (2006). Persuasion according to the unimodel: Implications for cancer communication. *Journal of Communication*, *56*(SUPPL.), S105–S122.

Kwan, Y. H., Cheng, T. Y., Yoon, S., Ho, L. Y., Huang, C. W., Chew, E. H., Thumboo, J., Østbye, T., & Low, L. L. (2020). A systematic review of nudge theories and strategies used to influence adult health behaviour and outcome in diabetes management. *Diabetes and Metabolism*, *46*(6), 450–460.

Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research*, *61*(5), 534–547.

Latkin, C. A., Edwards, C., Davey-Rothwell, M. A., & Tobin, K. E. (2017). The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. *Addictive Behaviors*, *73*, 133–136.

Latkin, C. A., Mai, N. V. T., Ha, T. V., Sripaipan, T., Zelaya, C., Minh, N. L., Morales, G., & Go, V. F. (2016). Social desirability response bias and other factors that may influence self-reports of substance use and HIV risk behaviors: A qualitative study of drug users in Vietnam. *AIDS Education and Prevention*, *28*(5), 417–425.

Lau, J. Y. C., Lee, C. K., Chan, C. P., Leung, J. N. S., Poon, C. M., & Lee, S. S. (2021). Compliance and attitudes of blood donors following transitioning from permanent to 12-month deferral of men who have sex with men in Hong Kong. *Vox Sanguinis*, *116*(5), 504–512.

Layder, D., & Stryker, S. (1982). Symbolic Interaction: A Social Structural Version. *The British Journal of Sociology*, *33*(3), 445.

Lee, C. S., Tan, J. H. M., Sankari, U., Koh, Y. L. E., & Tan, N. C. (2017). Assessing oral medication adherence among patients with type 2 diabetes mellitus treated with polytherapy in a developed Asian community: A cross-sectional study. *BMJ Open*, *7*(9), e016317.

Lewin, K. (2004). Intention, will and need. In *Organization and pathology of thought: Selected sources.*, (pp. 95–153). Columbia University Press.

Li, M., Taylor, E. G., Atkins, K. E., Chapman, G. B., & Galvani, A. P. (2016). Stimulating influenza vaccination via prosocial motives. *PLoS ONE*, *11*(7).

Lindenberg, S., & Papies, E. K. (2019). Two kinds of nudging and the power of cues: Shifting salience of alternatives and shifting salience of goals. *International Review of Environmental and Resource Economics*, *13*(3-4), 229–263.

Lins de Holanda Coelho, G., H. P. Hanel, P., & J. Wolf, L. (2020). The Very Efficient Assessment of Need for Cognition: Developing a Six-Item Version*. *Assessment*, *27*(8), 1870–1885.

Loewenstein, G. F., Hsee, C. K., Weber, E. U., & Welch, N. (2001). Risk as Feelings. *Psychological Bulletin*, *127*(2), 267–286.

Loh, E., Kurth-Nelson, Z., Berron, D., Dayan, P., Duzel, E., Dolan, R., & Guitart-Masip, M. (2017). Parsing the role of the hippocampus in approach-avoidance conflict. *Cerebral Cortex*, *27*(1), 201–215.

Lu, M., Safren, S. A., Skolnik, P. R., Rogers, W. H., Coady, W., Hardy, H., & Wilson, I. B. (2008). Optimal Recall Period and Response Task for Self-Reported HIV Medication Adherence. *Springer*, *12*(1), 86–94.

Lüke, T., & Grosche, M. (2018). What do I think about inclusive education? It depends on who is asking. Experimental evidence for a social desirability bias in attitudes towards inclusion. *International Journal of Inclusive Education*, *22*(1), 38–53.

Luong, P., Glorioso, T. J., Grunwald, G. K., Peterson, P., Allen, L. A., Khanna, A., Waughtal, J., Sandy, L., Ho, P. M., & Bull, S. (2021). Text Message Medication Adherence Reminders Automated and Delivered at Scale Across Two Institutions: Testing the Nudge System: Pilot Study. *Circulation: Cardiovascular Quality and Outcomes*, *14*(5), E007015.

Mahler, C., Hermann, K., Horne, R., Ludt, S., Haefeli, W. E., Szecsenyi, J., & Jank, S. (2010). Assessing reported adherence to pharmacological treatment recommendations. Translation and evaluation of the Medication Adherence Report Scale (MARS) in Germany. *Journal of Evaluation in Clinical Practice*, *16*(3), 574–579.

Maio, G. R., & Esses, V. M. (2001). The need for affect: Individual differences in the motivation to approach or avoid emotions. *Journal of Personality*, *69*(4), 583–614.

Manullrdby, A. C., Back, A., Horne, R., Landen, M., & Sundell, K. A. (2012). The medication adherence report scale (MARS-5) in a swedish sample with bipolar disorder-a pilot study. *Pharmacoepidemiology and Drug Safety*, *21*(2), 317.

Mariotti, A. (2015). The effects of chronic stress on health: New insights into the molecular mechanisms of brain-body communication. *Future Science OA*, *1*(3).

McAuliffe, T. L., DiFranceisco, W., & Reed, B. R. (2007). Effects of question format and collection mode on the accuracy of retrospective surveys of health risk behavior: A comparison with daily sexual activity diaries. *Health Psychology*, *26*(1), 60–67.

McElroy, T., & Seta, J. J. (2003). Framing effects: An analytic-holistic perspective. *Journal of Experimental Social Psychology*, *39*(6), 610–617.

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.

Mertens, S., Herberz, M., Hahnel, U. J., & Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(1).

Mikkelsen, C., Mori, G., van Walraven, S. M., Castrén, J., Zahra, S., MacLennan, S., Seidel, K., Fontana, S., Veropalumbo, E., Cannata, L., Pupella, S., Kvist, M., Happel, M., Korkalainen, P., Chandrasekar, A., Paulus, U., Bokhorst, A., Wulff, B., Fernandez-Sojo, J., Eguizabal, C., Urbano, F., Vesga, M. A., van Kraaij, M., Merz, E. M., van den Hurk, K., Hansen, M. B., Slot, E., & Ullum, H. (2021). How donor selection criteria can be evaluated with limited scientific evidence: lessons learned from the TRANSPOSE project. *Vox Sanguinis*, *116*(3), 342–350.

Molloy, G. J., Messerli-Bürgy, N., Hutton, G., Wikman, A., Perkins-Porras, L., & Steptoe, A. (2014). Intentional and unintentional non-adherence to medications following an acute coronary syndrome: A longitudinal study. *Journal of Psychosomatic Research*, *76*(5), 430–432.

Morisky, D. E., & DiMatteo, M. R. (2011). Improving the measurement of self-reported medication nonadherence: Final response. *Journal of Clinical Epidemiology*, *64*(3), 262–263.

Mowrer, O. H. (2009). *Learning theory and behavior.*. John Wiley & Sons Inc.

Mulder, J., & de Bruijne, M. (2019). Willingness of Online Respondents to Participate in Alternative Modes of Data Collection. *Survey Practice*, *12*(1), 1–11.

Müller, S., Kohlmann, T., & Wilke, T. (2015). Validation of the Adherence Barriers Questionnaire - An instrument for identifying potential risk factors associated with medication-related non-adherence Quality, performance, safety and outcomes. *BMC Health Services Research*, *15*(1), 1–12.

Nabi, R. L. (1999). A cognitive-functional model for the effects of discrete negative emotions on information processing, attitude change, and recall. *Communication Theory*, *9*(3), 292–320.

National Institute for Health and Clinical Excellence (NICE) (2007). Behaviour change :individual approaches. Tech. Rep. October, Public health guideline [PH49].

National Institute for Health and Clinical Excellence (NICE) (2009). Medicines adherence: involving patients in decisions about prescribed medicines and supporting adherence. Tech. Rep. January.

Nieuwkerk, P. T., De Boer-Van Der Kolk, I. M., Prins, J. M., Locadia, M., & Sprangers, M. A. (2010). Self-reported adherence is more predictive of virological treatment response among patients with a lower tendency towards socially desirable responding. *Antiviral Therapy*, *15*(6), 913–916.

Nieuwlaat, R., Wilczynski, N., Navarro, T., Hobson, N., Jeffery, R., Keepanasseril, A., Agoritsas, T., Mistry, N., Iorio, A., Jack, S., Sivaramalingam, B., Iserman, E., Mustafa, R. A., Jedraszewski, D., Cotoi, C., Haynes, R. B., Ackloo, E., Sahota, N., McDonald,

H. P., & Yao, X. (2014). Interventions for enhancing medication adherence. *Cochrane Database of Systematic Reviews*, *2014*(11).

O'Brien, S. F., Goldman, M., Robillard, P., Osmond, L., Myhal, G., & Roy, É. (2021). Donor screening question alternatives to men who have sex with men time deferral: Potential impact on donor deferral and discomfort. *Transfusion*, *61*(1), 94–101.

O'Brien, S. F., Osmond, L., Fan, W., Yi, Q. L., & Goldman, M. (2019). Compliance with time-based deferrals for men who have sex with men. *Transfusion*, *59*(3), 916–920.

Osterberg, L., & Blaschke, T. (2005). Adherence to medication. *New England Journal of Medicine*, *353*(5), 487–497.

Panksepp, J. (2004). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*(3), 598–609.

Paulhus, D. L. (1988). Balanced Inventory of Desirable Responding (BIDR). *Acceptance and Commitment Therapy. Measures Package*, *41*, 79586–79587.

Peer, E., Egelman, S., Harbach, M., Malkin, N., Mathur, A., & Frik, A. (2020). Nudge me right: Personalizing online security nudges to people's decision-making styles. *Computers in Human Behavior*, *109*.

Persson, M., & Solevid, M. (2014). Measuring political participation - Testing social desirability bias in a web-survey experiment. *International Journal of Public Opinion Research*, *26*(1), 98–112.

Peterson, C. K., & Harmon-Jones, E. (2012). Anger and testosterone: Evidence that situationally-induced anger relates to situationally-induced testosterone. *Emotion*, *12*(5), 899–902.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In *Communication and Persuasion*, (pp. 1–24). Springer, New York, NY.

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2015). A Clearer Picture: The Contribution of Visuals and Text to Framing Effects. *Journal of Communication*, *65*(6), 997–1017.

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2019). Framing fast and slow: a dual processing account of multimodal framing effects. *Media Psychology*, *22*(4), 572–600.

Racette, L., Abu, S. L., Poleon, S., Thomas, T., Sabbagh, N., & Girkin, C. A. (2022). The Impact of the Coronavirus Disease 2019 Pandemic on Adherence to Ocular Hypotensive Medication in Patients with Primary Open-Angle Glaucoma. *Ophthalmology*, *129*(3), 258–266.

Raiff, B. R., Jarvis, B. P., & Dallery, J. (2016). Text-message reminders plus incentives increase adherence to antidiabetic medication in adults with type 2 diabetes. *Journal of Applied Behavior Analysis*, *49*(4), 947–953.

Ramanath, K. V., Balaji, D. B., Nagakishore, C. H., Mahesh Kumar, S., & Bhanuprakash, M. (2012). A study on impact of clinical pharmacist interventions on medication adherence and quality of life in rural hypertensive patients. *Journal of Young Pharmacists*, *4*(2), 95–100.

Reese, P. P., Kessler, J. B., Doshi, J. A., Friedman, J., Mussell, A. S., Carney, C., Zhu, J., Wang, W., Troxel, A., Young, P., Lawnicki, V., Rajpathak, S., & Volpp, K. (2016). Two Randomized Controlled Pilot Trials of Social Forces to Improve Statin Adherence among Patients with Diabetes. *Journal of General Internal Medicine*, *31*(4), 402–410.

Renzetti, C., & Lee, R. (1993). Researching Sensitive Topics. In *CRVAW Faculty Book Gallery*.

Reynolds-Tylus, T. (2019). Psychological Reactance and Persuasive Health Communication: A Review of the Literature. *Frontiers in Communication*, *4*, 56.

Roseleur, J., Harvey, G., Stocks, N., & Karnon, J. (2019). Behavioral Economic Insights to Improve Medication Adherence in Adults with Chronic Conditions: A Scoping Review. *Patient*, *12*(6), 571–592.

Sandy, R., & Connor, U. (2015). Variation in medication adherence across patient behavioral segments: A multi-country study in hypertension. *Patient Preference and Adherence*, *9*, 1539–1548.

Sanquin (2021). Men in enduring, monogamous homosexual relationships welcome as blood donors.

Sasaki, S., Kurokawa, H., & Ohtake, F. (2021). Effective but fragile? Responses to repeated nudge-based messages for preventing the spread of COVID-19 infection. *Japanese Economic Review*, *72*(3), 371–408.

Schmidt, A. T., & Engelen, B. (2020). The ethics of nudging: An overview. *Philosophy Compass*, *15*(4), e12658.

Schöning, C., Matt, C., & Hess, T. (2019). Personalised nudging for more data disclosure? On the adaption of data usage policies format to cognitive styles. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, vol. 2019-Janua, (pp. 4395–4404).

Scribano, M. L., Caprioli, F., Michielan, A., Contaldo, A., Privitera, A. C., Bozzi, R. M., Calabrese, E., Castiglione, F., Ciccaglione, A. F., Delle Fave, G., Bodini, G., Costantino, G., Horne, R., Saettone, S., Usai, P., Vernia, P., Di Fino, S., Gualberti, G., di Fonzo, M., Merolla, R., & Orlando, A. (2019). Translation and initial validation of the Medication Adherence Report Scale (MARS) in Italian patients with Crohn's Disease. *Digestive and Liver Disease*, *51*(5), 640–647.

Shaz, B. H., Zimring, J. C., Demmons, D. G., & Hillyer, C. D. (2008). Blood Donation and Blood Transfusion: Special Considerations for African Americans. *Transfusion Medicine Reviews*, *22*(3), 202–214.

Shepherd, J., Kavanagh, J., Picot, J., Cooper, K., Harden, A., Barnett-Page, E., Jones, J., Clegg, A., Hartwell, D., Frampton, G. K., & Price, A. (2010). The effectiveness and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13-19: A systematic review and economic evaluation. *Health Technology Assessment*, *14*(7), 1–230.

Sherman, D. K., Mann, T., & Updegraff, J. A. (2006). Approach/avoidance motivation, message framing, and health behavior: Understanding the congruency effect. *Motivation and Emotion*, *30*(2), 165–169.

Shim, E., Chapman, G. B., Townsend, J. P., & Galvani, A. P. (2012). The influence of altruism on influenza vaccination decisions. *Journal of the Royal Society Interface*, *9*(74), 2234–2243.

Sieck, W., & Yates, J. F. (1997). Exposition effects on decision making: Choice and confidence in choice. *Organizational Behavior and Human Decision Processes*, *70*(3), 207–219.

Simon, A. F., Fagley, N. S., & Halleran, J. G. (2004). Decision framing: Moderating effects of individual differences and cognitive processing. *Journal of Behavioral Decision Making*, *17*(2), 77–93.

Sjölander, M., Eriksson, M., & Glader, E. (2013). The association between patients' beliefs about medicines and adherence to drug treatment after stroke: a cross-sectional questionnaire survey. *BMJ*, *3*, 3551.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*(3), 1333–1352.

Stentzel, U., van den Berg, N., Schulze, L. N., Schwaneberg, T., Radicke, F., Langosch, J. M., Freyberger, H. J., Hoffmann, W., & Grabe, H. J. (2018). Predictors of medication adherence among patients with severe psychiatric disorders: Findings from the baseline assessment of a randomized controlled trial (Tecla). *BMC Psychiatry*, *18*(1), 1–8.

Stirratt, M. J., Dunbar-Jacob, J., Crane, H. M., Simoni, J. M., Czajkowski, S., Hilliard, M. E., Aikens, J. E., Hunter, C. M., Velligan, D. I., Huntley, K., Ogedegbe, G., Rand, C. S., Schron, E., & Nilsen, W. J. (2015). Self-report measures of medication adherence behavior: recommendations on optimal use. *Translational Behavioral Medicine*, *5*(4), 470–482.

Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing Continuous and Dichotomous Scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment*, *78*(2), 370–389.

Stodel, M. (2015). But what will people think? Getting beyond social desirability bias by increasing cognitive load. *International Journal of Market Research*, *57*(2), 313–321.

Stone, J. K., Shafer, L. A., Graff, L. A., Lix, L., Witges, K., Targownik, L. E., Haviva, C., Sexton, K., & Bernstein, C. N. (2021). Utility of the MARS-5 in Assessing Medication Adherence in IBD. *Inflammatory Bowel Diseases*, *27*(3), 317–324.

Stull, D. E., Leidy, N. K., Parasuraman, B., & Chassany, O. (2009). Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. *Current Medical Research and Opinion*, *25*(4), 929–942.

Sunstein, C. R. (2016). People prefer system 2 nudges (kind of). *Duke Law Journal*, *66*(1), 121–168.

Takemura, K. (1992). Effect of decision time on framing of decision: A case of risky choice behavior. *Psychologia: An International Journal of Psychology in the Orient*, *35*(3), 180–185.

Takemura, K. (1994). Influence of elaboration on the framing of decision. *Journal of Psychology: Interdisciplinary and Applied*, *128*(1), 33–39.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, *58*(345), 345–372.

Tao, D., Xie, L., Wang, T., & Wang, T. (2015). A meta-analysis of the use of electronic reminders for patient adherence to medication in chronic disease care. *Journal of Telemedicine and Telecare*, *21*(1), 3–13.

Tesfaye, W., & Peterson, G. (2022). Self-reported medication adherence measurement tools: Some options to avoid a legal minefield. *Journal of Clinical Pharmacy and Therapeutics*, *47*(3), 363–368.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, New Haven & London.

Thompson, K., Kulkarni, J., & Sergejew, A. A. (2000). Reliability and validity of a new Medication Adherence Rating Scale (MARS) for the psychoses. *Schizophrenia research*, *42*(3), 241–7.

Thorneloe, R. J., Griffiths, C. E., Emsley, R., Ashcroft, D. M., Cordingley, L., Barker, J., Benham, M., Burden, D., Evans, I., Hussain, S., Kirby, B., Lawson, L., Mason, K., McElhone, K., Murphy, R., Ormerod, A., Owen, C., Reynolds, N., Smith, C., Warren, R., Barnes, M., Payne, K., Ryder, S., & Stocken, D. (2018). Intentional and Unintentional Medication Non-Adherence in Psoriasis: The Role of Patients' Medication Beliefs and Habit Strength. *Journal of Investigative Dermatology*, *138*(4), 785–794.

Timmermann, C. (2017). The Precautionary Principle and the Social Institution of Blood Donation. *American Journal of Bioethics*, *17*(3), 52–54.

Tommelein, E., Mehuys, E., Van Tongelen, I., Brusselle, G., & Boussery, K. (2014). Accuracy of the Medication Adherence Report Scale (MARS-5) as a Quantitative Measure of

Adherence to Inhalation Medication in Patients With COPD. *Annals of Pharmacotherapy*, *48*(5), 589–595.

Torrance, E. P., & Brehm, J. W. (1968). A Theory of Psychological Reactance. *The American Journal of Psychology*, *81*(1), 133.

Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, *133*(5), 859–883.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and Biases. *Science*, *185*, 267–286.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Q J Econ*, *106*, 1039–1061.

Ungemach, C., Camilleri, A. R., Johnson, E. J., Larrick, R. P., & Weber, E. U. (2018). Translated attributes as choice architecture: Aligning objectives and choices through decision signposts. *Management Science*, *64*(5), 2445–2459.

van de Mortel, T. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, *25*(4), 40–48.

Vidrine, J. I., Simmons, V. N., & Brandon, T. H. (2007). Construction of smoking-relevant risk perceptions among college students: The influence of need for cognition and message content. *Journal of Applied Social Psychology*, *37*(1), 91–114.

Volpp, K. G., Loewenstein, G., Troxel, A. B., Doshi, J., Price, M., Laskin, M., & Kimmel, S. E. (2008). A test of financial incentives to improve warfarin adherence. *BMC Health Services Research*, *8*(1), 272.

Walsh, J. C., Mandalia, S., & Gazzard, B. G. (2002). Responses to a 1 month self-report on adherence to antiretroviral therapy are consistent with electronic data and virological treatment outcome. *AIDS*, *16*(2), 269–277.

Wansink, B., & Pope, L. (2015). When do gain-framed health messages work better than fear appeals? *Nutrition Reviews*, *73*(1), 4–11.

Waughtal, J., Luong, P., Sandy, L., Chavez, C., Ho, P. M., & Bull, S. (2021). Nudge me: Tailoring text messages for prescription adherence through N-of-1 interviews. *Translational Behavioral Medicine*, *11*(10), 1832–1838.

Wentz, A. E., Merchant, R. C., Clark, M. A., Liu, T., Rosenberger, J. G., Bauermeister, J. A., & Mayer, K. H. (2019). Blood Donation, Sexual Practices, and Self-Perceived Risk for HIV in the United States Among Young Adult Men Who Have Sex With Men. *Public Health Reports*, *134*(1), 36–46.

Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. *Communication Monographs*, *59*(4), 329–349.

Wright, N. D., Morris, L. S., Guitart-Masip, M., & Dolan, R. J. (2013). Manipulating the contribution of approach-avoidance to the perturbation of economic choice by valence. *Frontiers in Neuroscience*, *0*(7 DEC), 228.

Wright, N. D., Symmonds, M., Hodgson, K., Fitzgerald, T. H., Crawford, B., & Dolan, R. J. (2012). Approach-avoidance processes contribute to dissociable impacts of risk and loss on choice. *Journal of Neuroscience*, *32*(20), 7009–7020.

Wroe, A. L. (2002). Intentional and Unintentional Nonadherence: A Study of Decision Making. *Journal of Behavioral Medicine*, *25*(4), 355–372.

Zhao, X., Villagran, M. M., Kreps, G. L., & McHorney, C. (2012). Gain Versus Loss Framing in Adherence-Promoting Communication Targeting Patients With Chronic Diseases: The Moderating Effect of Individual Time Perspective. *Health Communication*, *27*(1), 75–85.