----------Final Submitted version----------

You're Free When You Choose What I Like: Flexible Attributions of Free Choice Under Manipulation

Christos Bechlivanidis[1] & Magda Osman[2]

1. Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP England. UK.
2. Judge Business School, University of Cambridge, Trumpington St, Cambridge CB2 1AG. England. UK.

Abstract

The existence of mechanisms that operate below conscious awareness (e.g. unconscious bias, microtargeting, dark patterns) and the degree to which they may threaten our autonomy fascinate and challenge us in equal measure. Although concepts like free will and consciousness are being extensively debated in (experimental) philosophy, we know remarkably little about how everyday people perceive threats to their autonomy. Still, ethical debates and ensuing policy often assume that any manipulative attempt, conscious or otherwise, is perceived as worrisome and autonomy-threatening. The present work evaluates how these positions align with public perceptions of freedom and autonomy under conscious or unconscious manipulation. We take a novel approach by grounding experimental scenarios in situations members of the public have volunteered. Across three studies (N=1462), findings show that in realistic and relatable scenarios, people hold nuanced, pragmatic views. They are generally reluctant to conclude that choices are not free, even in cases of direct, successful, and inaccessible manipulation. When evaluating freedom of choice, people consider choice importance, victim identity (self vs. others), and the counterfactual case. Intriguingly, whether participants themselves like the outcome or not affects their evaluation of the victim's freedom, even if the victim's preferences are subverted. Overall, people judge there to be a lack of freedom only after a search across plausible causal factors fails to account for anomalous choice outcomes. We propose this as a viable rational cognitive process, contrasting with claims of public irrationality, underappreciation, or misunderstanding of core philosophical concepts.

There are arguably few concepts whose definition or even mere ontological existence have been as heavily debated as those of consciousness and free will. As is common with philosophical discourse, the excitement of the experts is not reflected in everyday people, and, as a result, we know very little about how the "folk" understand these concepts and their relationships (Knobe & Prinz, 2008; Nahmias et al., 2020). This state of affairs would be unexceptional and go unnoticed, if it wasn't for the encroachment of these concepts into the popular domain, through the common assumption that there are techniques that operate below the threshold of consciousness and are capable or eroding people's freedom of will. The most popular examples concern one's own latent beliefs affecting decision-making (unconscious bias) or the attempts by benevolent or malicious third parties to influence decision-making processes, through techniques known as nudging, subliminal messaging, psychological priming, hypnosis, microtargeting etc[1]. Despite the active debates about the ethical status of such techniques (Sunstein, 2014a) and whether governments and other entities should intervene (Council Directive 2018/1808/EC, 2018; Halpern & Costa, 2019), these discussions often rest on psychological assumptions that have yet to be empirically examined. In fact, very little is known about how people view unconscious mechanisms, whether they perceive them to be potent enough to diminish free choice[2] and, if so, whether that's at all concerning to them.

Some evidence comes from recent work in experimental philosophy (or X-Phi) which has focused attention on people's intuitions on philosophical concepts such as consciousness (Knobe & Prinz, 2008; Sytsma & Machery, 2009), free will (Baumeister et al., 2009; Genschow et al., 2017; Nahmias et al., 2005; Nichols & Knobe, 2007), and, critically for our interests, their relationship (Nahmias et al., 2020; Shepherd, 2012, 2015; Sytsma & Snater, 2023). One line of work, in particular,

[1] Despite a recent resurgence, it is by no means the first time that the unconscious and the possibility of manipulating it have been capturing our imagination. From Franz Mesmer's animal magnetism in the 18th century (Tallis, 2002) to Judas Priest's backmasked messages (Schmaltz, 2016), we seem to be particularly worried and/or intrigued by the prospect that our choices and behaviour may be influenced or determined without our conscious awareness or consent.

[2] Although the concepts of "free will" and "free choice" are arguably distinct, for simplicity, we will be using the terms interchangeably.

has examined whether people take consciousness to be a necessary prerequisite for free action[3], and whether, by extension, any influence that isn't consciously accessible would result in unfree actions. Researchers, for example, have asked participants whether there is free will in the absence of consciousness either directly (Shepherd, 2012) or through imaginative scenarios featuring split brain patients with conscious and unconscious brain parts, and humanoids (Shepherd, 2012) that were either pre-programmed of capable of learning from interactions (Nahmias et al, 2020). Overall, the results indicate that consciousness is positively correlated it free will (Shepherd, 2012, 2015) although the relationship with moral responsibility is less clear (Nahmias et al, 2020).

In a parallel line of work, however, consciousness is shown to be less relevant, on the grounds that any act of manipulation fundamentally compromises free will (Kane, 1985; Pereboom, 2006). According to "deep self" theories (Frankfurt, 1971; Sripada, 2016; Watson, 1975; Wolf, 1990), an action is free when it agrees with the agent's values and preferences, and thus the agent's "true self". Therefore, any manipulation that successfully subverts the agent's preferences leads to actions that are not free. Supporting evidence comes from experiments using similar methodology, with vignettes featuring evil scientists (Sripada, 2012) and omniscient tourists (Murray & Lombrozo, 2017; although see Phillips & Shaw, 2015 for a mundane scenario).

However, despite the volume of work on intuitions about consciousness and manipulation, there is little consensus and frequent inconsistencies in the pattern of findings (Genschow et al., 2023; Osman & Bechlivanidis, 2024). While, for some researchers as we saw, people treat consciousness as necessary for free will (Knobe & Prinz, 2008), others argue that people don't even have a concept of consciousness (Sytsma & Machery, 2009; Sytsma & Snater, 2023). Similarly, some work provides evidence that people are incompatibilists, whereby determinism precludes free will (Murray & Nahmias, 2014; Nahmias et al., 2005), whereas others present evidence to the contrary (Nichols & Knobe, 2007).

[3] That could be the case for various reasons, e.g. because free will requires one to have conscious access to alternative options or to reasoning processes (Fischer & Ravizza, 1998), that for freely willed actions one must be consciously aware and endorse the underlying reasons (Wolf, 1990), or that a free agent must be capable of certain conscious phenomenal states such as guilt, gratitude (P. Strawson, 1962), or concern for others (Frankfurt, 1999).

These discrepancies often stem from terminological misunderstandings, sometimes due to researchers inadvertently targeting different concepts (Block, 1995), or, most commonly, due to ordinary people conceptualising terms very differently compared to philosophers (e.g. Murray & Nahmias, 2014).

A more fundamental concern involves how well these experimental approaches reflect real-world thinking. Experimental philosophy generally prioritizes validating philosophical assumptions over understanding human cognition (Knobe & Prinz, 2008). To that end, studies in the field usually provide participants with definitions of complex philosophical terms and/or ask them to navigate highly counterfactual vignettes that require them to suspend their understanding of reality (Osman & Bechlivanidis, 2024). As a result, it is not clear whether responses on scenarios involving sentient robots, split-brain patients or omniscient beings, correspond to real, stable intuitions, that translate to the everyday situations people are confronted with and care about. Is the perceived absence of free will in the case of the killer raised by the evil scientist, analogous and transferable to the case where marketing influences the consumer into buying the glass cleaner they wouldn't otherwise go for? Does finding out that the learning robot is freer than the pre-programmed one, afford us any predictions about the perceived free will of a person with learning difficulties?

To be clear, we are not arguing against the value of X-Phi studies; rather, we believe it is important to caution against misinterpreting abstract theoretical views as everyday intuitions. Addressing this distinction, an alternative line of work takes a more bottom-up approach by asking people to generate their own illustrative examples. Stillman and colleagues (2011), for example, asked people to describe either free or unfree actions and analysed responses across multiple dimensions. They find that people tend to mention positive outcomes and goal attainment more when describing free actions, while there was also an effect of consciousness, albeit weaker.

In recent work, Osman (2020) followed a similar approach, by asking people to suggest examples in which free will is being manipulated without the victim being aware. This resulted in a very rich body of examples, in contexts ranging from social media to marketing, psychological research and therapy. Rather than having independent researchers analyse these examples, different groups of

participants were subsequently asked to evaluate the degree to which they believe the technique mentioned in each example targets the unconscious, diminishes free will, or raises concern (Osman, 2020).

In another set of experiments, the level of personalisation was varied by asking participants either to judge the general case without specifying any particular victim, or to try to imagine themselves in the situation mentioned in the scenario (Osman & Bechlivanidis, 2021, 2022, 2023). At first glance, the results agreed with previous work (Nahmias et al., 2020; Shepherd, 2012, 2015), with people attributing a lower degree of free will and control in the presence of unconscious manipulation, i.e. in the absence of conscious access to the decision-making process (Fischer & Ravizza, 1998). However, whereas one would expect a very strong negative correlation, i.e. that free will is seen to diminish as the influence of unconscious manipulation gets stronger, the correlation was non-existent to weak ($r<0.3$) (Osman & Bechlivanidis, 2022). Although people recognise the relationship that theoretical philosophy and past empirical evidence propose, they still allow for free will in the absence of consciousness, and this is not accounted for by the degree of the perceived success of the manipulation. When attempting to identify the factors that modulate this relationship, there was little evidence for individual differences (e.g. age, gender, ethnicity, religiosity, education, political affiliation), showing remarkable agreement between and within countries (Osman, 2020; Osman & Bechlivanidis, 2023). What seems to be making a difference is the particular examples of unconscious manipulation, with scenarios such as hypnosis for therapeutic purposes, subliminal messaging and microtargeting on social media, showing the strongest negative correlations (Osman & Bechlivanidis, 2022). A second factor that played a consistent role, and which was eventually manipulated experimentally (e.g. Osman & Bechlivanidis, 2021, 2022, 2024), was the degree of personalisation. The more concrete the instructions became in placing oneself in the scenarios, the stronger the negative correlation between manipulation strength and free will became. This reached the strongest levels when participants imagined themselves as victims in a choice scenario that unfolded in the present. Even then, the strength of the correlation remained far below the levels aforementioned theories would predict.

Taken together, factors such as the realism of the case being considered, its abstract or concrete formulation, and whether one is asked to adopt a third- or first-person perspective, are all critical factors when studying people's intuitions on concepts like free will and consciousness. Although the experimenter is losing some control over the experimental conditions when using ecologically valid scenarios that people have themselves generated, the findings will correspond more closely to folk intuitions and will therefore be relevant to everyday situations and debates (Dhami et al, 2004). The findings also bring about novel and puzzling questions that the present work considers in detail. The present study tightened the experimental controls employed, while maintaining, where possible, the concreteness and relatedness of the situations presented to participants and avoiding highly stylised and artificial departures from reality.

Current work

The overarching aim of the work presented here is to examine the factors that affect people's attributions of free choice in realistic and relatable settings. The focus is to understand the persistence of relatively high ratings of free will in the face of unconscious manipulation observed in previous work (Osman & Bechlivanidis, 2022, 2024), and, relatedly, the weaker than expected negative correlation between successful unconscious manipulations and free will. With that in mind, out of the examples that participants in previous studies had volunteered (Osman, 2020, Osman & Bechlivanidis, 2024), we chose a scenario that refers to microtargeting in social media. Although the potency of this technique has recently been under scrutiny (Bodó et al., 2017; Tappin et al., 2023), it was perceived as one of the most capable manipulations in reducing free will for participants in previous work (Osman & Bechlivanidis, 2022). The technique is relevant to philosophical theories since it attempts to influence the preferences that lead to some decision (Sripada, 2016), and it has recently been capturing the headlines as a factor influencing electoral outcomes (e.g. Stracqualursi, 2020; Weaver, 2018).

To ensure relatedness, participants are asked to express their actual preferences before being asked to imagine that these have been manipulated. These preferences are related to everyday decisions about a supermarket item, a fashion item and a political candidate. The objects of choice are

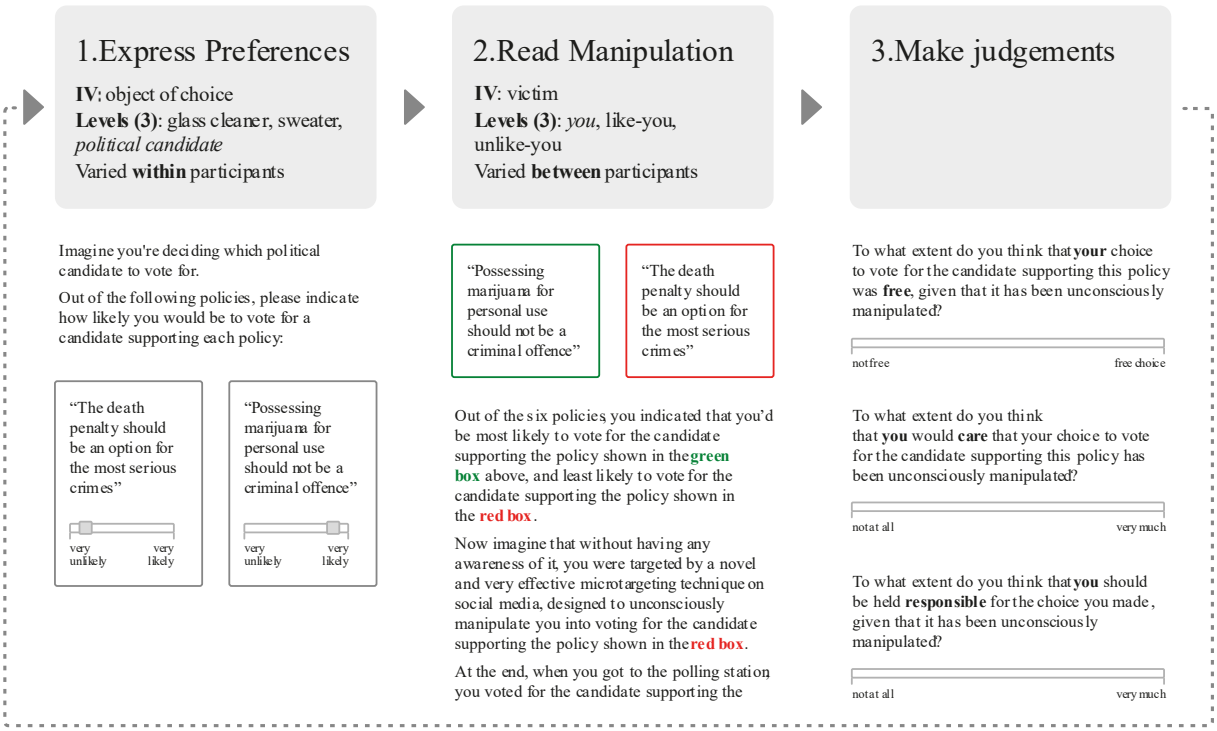
selected to differ in terms of the degree to which they reflect people's values. The assumption here is that people identify more with their political views, followed by their fashion choices, and the cleaning products they buy (Klein III et al., 1993). If so, it is reasonable that they would show a variable degree of concern when each of these choices is being manipulated, and that may also be reflected in their free choice evaluations.

To examine the effect of personalisation observed in previous work (Osman & Bechlivanidis, 2021, 2022, 2024), the victim of the manipulation is experimentally varied. Some participants are asked to imagine that they are victims of the manipulation themselves, while others are asked to imagine another person being the victim. We further distinguish between victims having the same preferences as those expressed by the participant, and victims with opposite preferences. While philosophical theories remain neutral on this point, past work and classical psychological frameworks yield divergent predictions, which will be discussed in the following sections.

All experiments follow an identical three-step procedure. In every case, participants first express their preferences for 6 items (Figure 1, left panel). They are then asked to imagine a scenario where those preferences have been manipulated (middle panel). Finally, they evaluate the extent to which, given the manipulation, the choice is free, whether the victim would be concerned and who would be responsible for the outcome (right panel).

Figure 1

Procedure followed in all experiments, with a screenshot illustrating each step



Note: The experiments differed only in step 2 which described different types of manipulations (see text for more details). The screenshots have been slightly modified for presentation purposes, e.g. step 1 shows only two of the six options shown to participants. The level in italics for each IV is the one shown in the screenshot.

The experiments differed only in the type and direction of the manipulation. In Experiments 1 and 2 the manipulation was described as operating below the victim's awareness (microtargeting), while in Experiment 3, the victim was fully aware (peer pressure; Noggle, 2021). Also, while in Experiments 1 and 3, the manipulation and the outcome were in the opposite direction to the victim's preferences (incongruent), in Experiment 2, the manipulation direction was congruent to the victim's preferences. In the incongruent case the victim got their least preferred option but in the congruent case the victim ended up with the option they indicated as preferred. The otherwise identical nature of

the three experiments was intended to allow direct cross comparisons between the various forms of manipulation.

Since we were interested not only in relative comparisons but also in the degree to which various factors (e.g., awareness, choice subversion) diminished perceived free choice, we conducted a preliminary experiment (N = 79) to estimate baseline metrics, i.e. the degree of perceived free choice and responsibility in the absence of any manipulation. This experiment followed a structure similar to the studies described above, and its full details and results are available on OSF at [insert link]). Critically, it included a condition presented in the second person, where participants imagined a scenario in which "the law bans any attempt by organisations or marketing teams to influence a choice" and the outcome matched their preferences. In this case of no manipulation and a congruent outcome, free choice ratings were close to ceiling (M = 86.43, SD = 21.33), as were responsibility attributions (M = 82.14, SD = 28.10). We will be using these baseline values below when making relevant comparisons.

Experiment 1

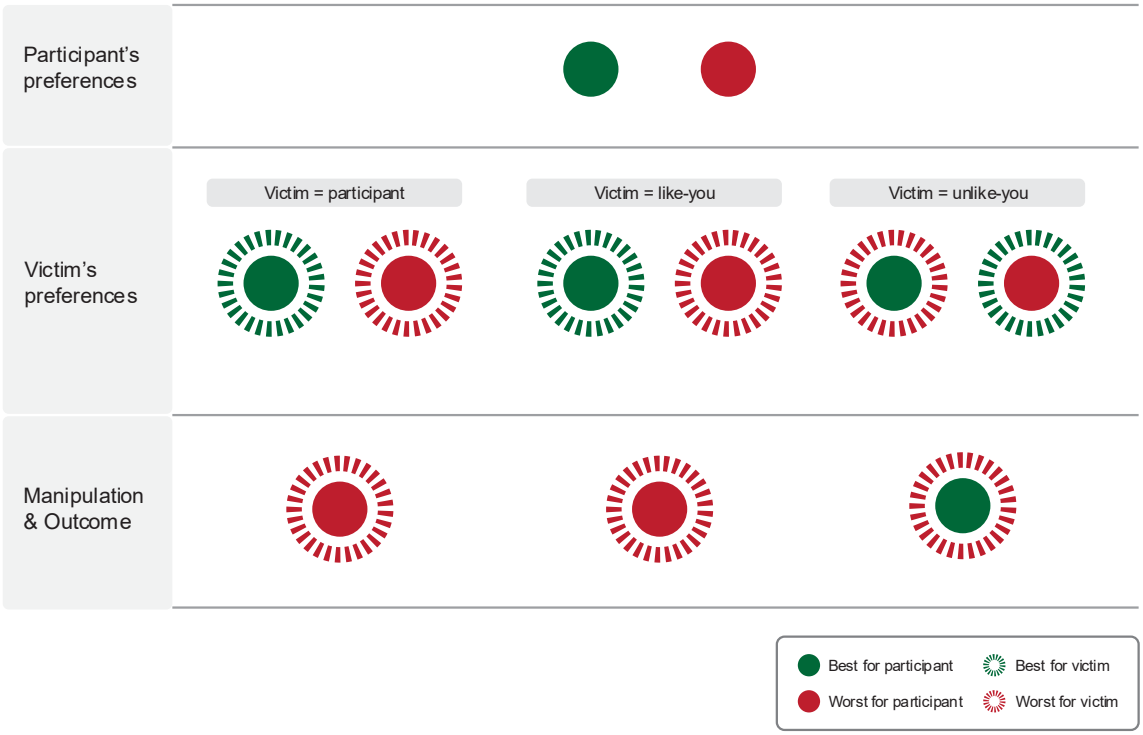
The first experiment aimed at investigating the effects of the object of choice and the identity of the victim on perceptions of free choice, in the context of a manipulation that targets the victim under their conscious awareness. All participants followed the procedure outlined in Figure 1 three times, once per object of choice (glass cleaner, sweater, political candidate). The victim (varied between participants) was framed as either the participant themselves, someone who shared the same preferences as the participant, or someone with the opposite preferences. Regardless of the victim and the object of choice, the experimental set up described the manipulation method (microtargeting) in such a way that it succeeded at completely subverting the victim's preferences, so that the victim in each case was always influenced into choosing the option that they liked the least.

It is important to note that the design of this experiment indirectly generates another independent variable, that of the valence of the outcome from the participant's perspective. In conditions where the victim shared the participant's preferences, the manipulation resulted in an

outcome both parties liked least (Figure 2, middle column). Conversely, when the victim held opposite preferences, they were steered toward their own least-preferred option—which, in this case, happened to be the participant's favorite (right column). This allows us to study whether free choice judgements are affected not only by the intentions of the victim, but also by the attractiveness of the final outcome in the eyes of the participant, even if the victim was manipulated into it.

Figure 2

The three different conditions in Experiment 1



Note: In all conditions the participant expressed their preferences (green circle for their most and red circle for their least preferred options) and then was asked to imagine a scenario of manipulation where the victim was either themselves, someone with the same (like-you) or someone with the opposite preferences (unlike-you) - the victim's preferences are shown here with a green or red ring. The manipulation led always to the least preferred outcome for the victim, which in the "unlike-you" condition coincided with the participant's most preferred option.

Although, to our knowledge, no studies have examined people's views on the relation between the unconscious, its manipulation and the effect on free choice, some predictions can be derived from theoretical arguments and some recent empirical evidence. So, theories that hold that consciousness is a necessary ingredient of free will (Frankfurt, 1999; G. Strawson, 1994), would predict no effect of our independent variables on judgements of free choice. Any successful manipulation that unconsciously influences the victim, also strips them of their free will, independent of what the choice is about, the identity or the preferences of the victim, let alone how palatable the outcome of a manipulated choice appears from the perspective of a third party.

Conversely, for deep self theories (Sripada, 2016; Watson, 1975; Wolf, 1990) that argue that the hallmark of free will actions is the expression of the "true" self, manipulations that affect "important" decisions (e.g. voting) would have a more detrimental effect on free choice compared to less identity-bearing choices (e.g. a glass cleaner). In the case of the victim of the manipulation, although we are not aware of a philosophical theory that would predict any effect, we can derive opposing predictions, based on psychological theories. If participants see themselves as less free under manipulation than other people, as previous work seems to suggest (Osman & Bechlivanidis, 2022), that could be due to their tendency to hold other people more responsible. This would be in line with fundamental attribution error (Ross, 1977) or the actor-observer effect (Jones & Nisbett, 1972), whereby, for negative outcomes, people are more likely to use contextual factors to explain their own behaviour, but refer to internal dispositions to account for other people's behaviour. Alternatively, if participants are reluctant to see themselves as easier to manipulate than other people (cf. third-person effect, Davison, 1983; Perloff, 1999), they will try to preserve a sense of their own freedom despite manipulation, thus judging themselves as freer than others. We also considered whether the effect of personalisation carries over when considering other people, and, specifically, whether participants would judge people with preferences similar to their own to be more, or less free compared to people with the opposite preferences.

Participants

All experiments were preregistered at OSF (https://osf.io/zrvkt/?view_only=4c50ed5a97ce4834ac9456d2539f11ba). Below, we report how we determined our sample size, all data exclusions, all manipulations, and all measures used in our studies.

We aimed for a sample size of 480 for each experiment, decided through a power calculation based on pilot studies[4], using an alpha of .05, and expecting effect sizes ranging from η_p^2=.0114 to η_p^2=.0725 (see preregistration document for more details). For Experiment 1, we recruited a total of 486 participants through Prolific (www.prolific.com), requiring only that they resided in the United States, since some stimuli referred to local issues. Of them, 2 participants were removed from the analysis due to a technical error, and an additional 9 for failing more than one of the three attentional checks (see procedure below). The resulting 475 participants had a mean age of 42.74 years (SD = 13.96), 252 were females (53%), 219 were males (46%) and 4 chose "other"[5].

The experiments were programmed in JavaScript (the source code is available at https://osf.io/64jgb/files/osfstorage) and conducted over the Internet. Participants were paid £0.70 for taking part. All experiments received ethical approval (EP/2023 /002).

Design and procedure

The design was mixed (see Figure 1) with the between-subject condition being the victim type ("you", "like-you, "unlike-you") and the within-subject condition being the object of the decision ("glass cleaner", "sweater", "policy").

Participants were first asked to imagine a situation in which, depending on the within-subject condition, they had to decide which glass cleaner or which sweater to buy, or which political candidate to vote for on the basis of a political statement the candidate stood for. In each case, they were shown

[4] The data for the pilot studies, the results of which are in the same direction as what we present here, can be found at https://osf.io/64jgb/files/osfstorage, together with graphs that mirror those shown below.

[5] A sensitivity power analysis conducted using G*Power 3.1.9.4 and default settings suggested that with this sample size, a minimum effect size of n_p^2= .013 (f= .117) could be detected for between subjects factors (power=0.80, α = 0.05), n_p^2= .003 (f= .058) for within subjects factors and n_p^2= .004 for interactions (f= .065).

six options in random order and were asked how likely they would be to buy or vote for each. The glass cleaners and the sweaters were randomly chosen pictures with the only requirement to look sufficiently different. The political statements were selected from an online political orientation quiz (https://politicalcompass.org/) aiming for a mixture of democrat and republican leaning statements.

After expressing their preferences, participants were shown the option they indicated they liked the most and the option they indicated they liked they least. In the "like-you" condition, participants were asked to imagine that some person called Frankie had exactly the same preferences as them. In the "unlike-you" condition, they were asked to imagine that Frankie had exactly the opposite preferences, i.e. Frankie liked most the option that the participant liked the least, and Frankie liked the least the option that participant liked the most. In the policy condition for example, if the participant was most in favour of legalising marijuana and least in favour of the death penalty, Frankie would have the same preferences in the "like-you" condition, while in the "unlike-you" condition, Frankie's strongest opinions would be in favour of the death penalty and against legalising marijuana. When present, Frankie was described using either masculine or feminine pronouns, randomly chosen for each participant.

Participants were then asked to imagine that "without having any awareness of it" either them ("you" condition) or Frankie ("like-you" and "unlike-you" conditions), "were targeted by a novel and very effective microtargeting technique on social media, designed to unconsciously manipulate" them into buying the item or voting for the candidate they liked the least. Note that the item the victim liked the least coincided with the item the participant liked the least in the "you" and "like-you" conditions. In the "unlike-you" condition, the option that Frankie was manipulated into buying or voting for, was the option that Frankie liked the least but that the participant liked the most (see Figure 2). So, for example, if the participant said they were most strongly in favour of legalising marijuana, in the "unlike-you" condition, Frankie was described as strongly opposing that policy but was then manipulated into voting for the candidate supporting it.

In the next screen, the participant saw the option that they or Frankie ended up with and were asked "to what extent the choice was free, given that it has been unconsciously manipulated", whether the victim "would care that their choice has been unconsciously manipulated", and, finally, to what extent the victim "should be held responsible" for buying the product or voting for the candidate they ended up with given the unconscious manipulation. All responses were given on a sliding scale ranging from 0 (and labelled "not free" or "not at all") to 100 (labelled "free choice" or "very much"). An additional attention check question requested participants to drag the slider to a predefined location. After responding, participants saw the next item (randomised for each participant), read the manipulation scenario, expressed their preferences and answered the same set of questions (including the attention check). After going through this procedure three times, once per item type, participants were asked for some basic demographics (age, sex, highest level of education and political leanings) and were thanked for their participation.

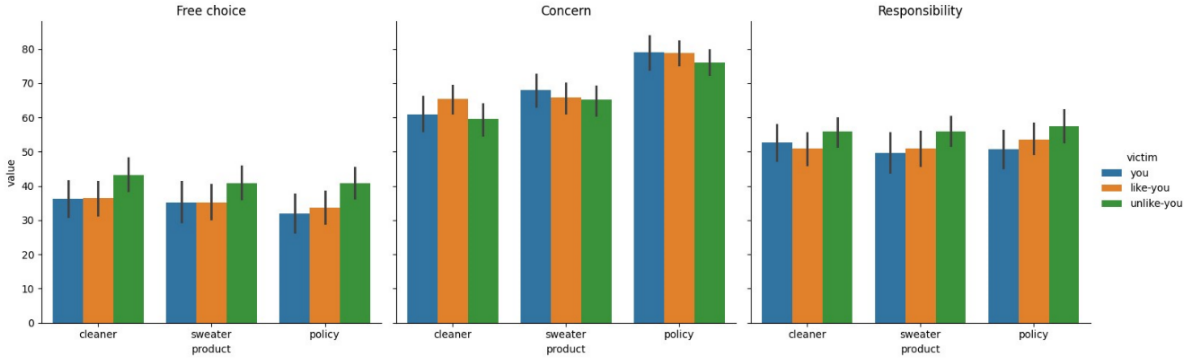
Results

Looking at the leftmost panel in Figure 3 below[6], judgements of free choice reveal two patterns. When participants imagine themselves (blue bars) or someone with the same preferences (orange bars) as the victim of the manipulation, the choice is rated to be less free than when imagining the victim having the opposite preferences to them (green bars). Second, there is an effect of the object of choice, whereby voting choices are deemed less free when manipulated, compared to choices of sweaters or glass cleaners. When participants were asked how concerned the victim of the manipulation would be, (Figure 3, middle panel), there seems to be a more pronounced effect of the object of choice but no discernible effect of victim. Finally, the two independent variables (victim, object of choice) don't seem to have a discernible influence on judgements of responsibility.

[6] The raw data for all experiments, as well as the Python script that was used to produce the plots and conduct the statistical analysis can be found at https://osf.io/64jgb/files/osfstorage

Figure 3

Mean ratings for the 3 questions (free choice, concern and responsibility) in Experiment 1, across objects of choice and victims



Note: Error bars represent 95% confidence intervals.

A set of mixed effects ANOVAs verify these impressions: In the case of free choice, there's a main effect of victim ($F(2,472)=3.30$, $p=.038$, $n_p^2=.014$) and an effect of the object of choice ($F(2,944)=3.33$, $p=.036$, $n_p^2=.007$) but not a significant interaction ($F(4,944)=.32$, $p=.862$, $n_p^2=.001$). Preregistered follow-up pairwise t-tests show that the effect of victim is driven by the "unlike-you" condition, in which free choice ratings are significantly higher than those in "you" ($t(269.90)=2.13$, $p=.034$, $d=.245$) and "like-you" conditions ($t(344.26)=2.20$, $p=.029$, $d=.235$), since the conditions "you" and "like-you" do not differ significantly ($t(259.35)=.16$, $p=.871$, $d=.020$). On the other hand, the object of choice effect is driven by the difference between the glass cleaner and the policy conditions ($t(474.00)=2.58$, $p=.010$, $d=.094$), with no significant difference found in the other two comparison (glass cleaner-sweater: $t(474.00)=1.53$, $p=.127$, $d=.050$, policy-sweater: $t(474.00)=1.10$, $p=.270$, $d=.043$).

Similar tests for concern, show a strong effect of the object of choice ($F(2,944)=83.20$, $p<.001$, $n_p^2=.150$) but no effect of victim ($F(2,472)=.86$, $p=.422$, $n_p^2=.004$) and no interaction effect ($F(4,944)=1.35$, $p=.248$, $n_p^2=.006$). Follow up t-tests for the object of choice show all pairwise comparisons to be significant (cleaner-policy: $t(474.00)=11.55$, $p<.001$, $d=.577$, cleaner-sweater: $t(474.00)=3.54$, $p<.001$, $d=.146$, policy-sweater: $t(474.00)=9.31$, $p<.001$, $d=.428$). Finally, a mixed effect

ANOVA for the effect of either variable on judgements of responsibility showed no significant main effects or interactions (victim: $F(2,472)=1.78$, $p=.169$, $n_p^2=.008$, object :$F(2,944)=1.42$, $p=.242$, $n_p^2=.003$, interaction: $F(4,944)=.90$, $p=.461$, $n_p^2=.004$)).

Ratings of free choice are correlated with both ratings of concern ($r(1423)=-.16$, $p<.001$) and ratings of responsibility ($r(1423)=.53$, $p<.001$) but ratings of concern and responsibility show no significant correlation ($r(1423) =-.05$, $p=.086$).

Discussion

This experiment utilized a concrete scenario to investigate the impact of unconscious manipulation on everyday choices. Based on a context that was suggested by participants in previous work (Osman, 2020), we designed a task that prioritized personal relevance by subverting preferences expressed by participants moments earlier. Our results indicate that the perceived relationship between manipulation and free will is not invariant; rather, it is characterized by a high degree of nuance that depends on the object of choice, the identity of the victim, and the valence of the subverted outcome.

Participants believe that choices are relatively freer when, following unconscious manipulation, the victim ends up buying a cleaning product they least preferred, compared to voting for a candidate whose policies they least preferred. This is in line with deep self theories, whereby free will is, above all, the expression of one's deepest, most stable beliefs, preferences and values theories (Frankfurt, 1971; Sripada, 2016; Watson, 1975; Wolf, 1990). It is reasonable to assume that for most people, such values are reflected more in their political leanings, followed by their fashion choices, and less by the way they choose mundane, everyday consumer items. Should people share similar intuitions with these deep self views then, it stands to reason that free choice judgements would follow the same ranking. In other words, the more the object of choice expresses one's identity (Klein III et al., 1993), one's "true" self, the less free one is if the outcome deviates from the preferred choice. This is further explored in the General Discussion section when considering explanations about the relationship between concern, which showed the opposite pattern, and perceived freedom.

The second notable pattern concerned the type of victim of manipulation. Equivalent levels of free choice were attributed to the self as to victims with the same preferences, but victims with opposite preferences were judged to be freer. This is despite the fact that, across conditions the type of manipulation is identical, and that the victim always ends up with the option they least preferred. One way to explain this effect is by hypothesising that, according to Construal Level Theory (Liberman et al, 2007) people with opposite preferences are seen to be psychologically more distant or outgroup (Tajfel & Turner, 1979). If that is so, then the Actor-Observer asymmetry (Jones & Nisbett, 1987), would predict that more distant people would be held more responsible for their actions and outcomes. The moderate correlation between judgements of free choice and judgements of responsibility lends support to this explanation. It is undermined, however by the lack of significance when estimating the effect of victim on responsibility (although the trend is similar).

An alternative, and simpler explanation for the differentiation of victims based on their preferences, focuses exclusively on the valence of the outcome. As explained, the victim with opposite preferences ended up with what, from the participant's point of view, is the most preferred option. From this position, the victim achieved a positive, albeit manipulated outcome. Thus, from the participant's position, there may be reasons that this victim has actually changed their mind, i.e. they "saw the light". Under this interpretation, participants may be using the outcome of some behaviour as evidence when considering the level of free choice. What a participant deems as a non-preferred outcome, suggests manipulation for them and victims with similar preferences, because neither got what they wanted. On the other hand, the participant's preferred choice as an outcome for a victim with opposite preferences may suggest a deliberate switch to the "right" or "good" option. Rather than the result of manipulation, such outcome would be explained by the victim's own updated preferences, making the choice freer. The General Discussion section further explores this interpretation in an attempt to capture both the effect of victim and the effect of the object of choice discussed here, as well as the results of the experiments to be presented next.

What the above discussion highlights is that the scenarios we presented to participants in the two “other” groups, differ not only in the victim’s preferences (making them similar or dissimilar to the participant’s) but also in the valence of the manipulated outcome (making it desirable or undesirable to the participant). In an effort to distinguish the effect of these two variables, and the associated explanations that are based either on psychological distance and responsibility or the valence of the outcome, in Experiment 2, there will be no difference between the agent’s initial preferences and the manipulated outcome.

Finally, “folk” applications of philosophical concepts as revealed in Experiment 1 do not appear to suggest confusion or incoherence, as has often been concluded, but instead indicate a nuanced calculus for making free choice attributions. Even if the absence of consciousness appears to have a detrimental effect on freedom, consciousness is not the only consideration. Participants in this study judge the victim of manipulation to be more or less free depending on both the object of the choice and the outcome of the decision. Choices with which one doesn’t deeply identify, or outcomes that are favourable are perceived as freer choices, despite unconscious manipulation.

Experiment 2

This experiment examines appraisals of unconscious manipulation and free choice when the manipulation is in line with the victim’s preferences. Typically, manipulation is studied in the context of choice subversion, presumably due to the clear ethical implications. By focusing on this less examined 'aligned' manipulation, we can assess whether concern for manipulation persists regardless of the outcome, providing a critical point of comparison to the results of Experiment 1. This context is more prevalent than might be expected. The hypnosis specialist, for example, may try to make the patient quit smoking, following the patient’s wishes. Or, more commonly, one may come across a social media campaign that tries to sway them towards a political candidate they have already decided to vote for.

How do people make free choice attributions when the agent ends up with the option they initially preferred but were also manipulated towards? If what matters when evaluating freedom and manipulation is the presence or absence of consciousness, as some scholars seem to suggest (Strudler,

2005; Susser et al., 2019), then the direction of manipulation is irrelevant. Conversely, an alternative view focuses on the alignment between an agent's intentions and outcomes. Here, an intervention only undermines freedom if it satisfies the counterfactual - meaning the agent *would* have acted differently without it. When this counterfactual is false because the agent's actual choices match their uninfluenced desires, autonomy is preserved (Frankfurt, 1969).

The second aim of this experiment was to determine for which of two viable hypotheses arising from Experiment 1 there is empirical support. One possible interpretation of the pattern observed there, is that victims with different preferences are perceived to be more psychologically distant (Trope & Liberman, 2010), making them appear more responsible for their actions (Jones & Nisbett, 1972), thus freer. The other explanation is not based on initial preferences, but on final outcomes: the more participants like an outcome, the freer the agent that ends up with that outcome is taken to be, even if, in every case, the agent was manipulated against their preferences. In this second experiment, preferences and outcomes will be in the same direction, hopefully shedding some light to this dilemma.

Participants were asked for their preferences on the same items as before and saw a scenario of unconscious manipulation against themselves or others, depending on condition. This time, however, the manipulation and the resulting outcome were in the same direction as the victim's preferences. In other words, compared to Experiment 1, the only change was that the direction of the manipulation and the outcome were congruent rather than incongruent with the victim's preferences, meaning that regardless of conditions, all victims got the outcome they wanted.

Participants

We recruited 489 participants from the United States through Prolific and paid them £0.70 for taking part. 15 participants were later removed from the analysis for failing more than 1/3 attentional

checks. Their mean age of the remaining 474 participants was 40.88 years (SD = 13.16), and there were 260 females (54%), 207 males (44%) and 7 choosing "other"[7].

Design and procedure

The design was identical to Experiment 1, i.e. the victim ("you", "like-you, "unlike-you") was varied between participants and the object of the choice ("glass cleaner", "sweater", "policy") varied within. The procedure was also very similar with the only difference being the direction of the manipulation. Participants first expressed their preferences for the six items for each of the three choice scenarios (glass cleaners, sweaters, political statements). In the "you" condition, they were then reminded of the option they liked the most and the option the liked they least. If they were in the "like-you" condition they were asked to imagine another person, Frankie, having the same preferences as them, or, in the "unlike-you" condition, they were asked to imagine Frankie having the opposite preferences, same as in Experiment 1. The manipulation was then described in identical terms, but this time participants were told that it happened to be in the direction of the option that the victim liked the most. In the "you" and "like-you" condition that was the item the participant liked the most, but in the "unlike-you" condition it was the item the participant liked the least (but again it was congruent with the victim's preferences). Finally, participants were told that the victim ended up buying or voting in the direction of their preferences and the manipulation, before seeing the same set of questions as in Experiment 1 (free choice, concern, responsibility) and proceeding to the next set of options. After going through all three items, the same demographics questions were administered, and participants were thanked for taking part.

Results

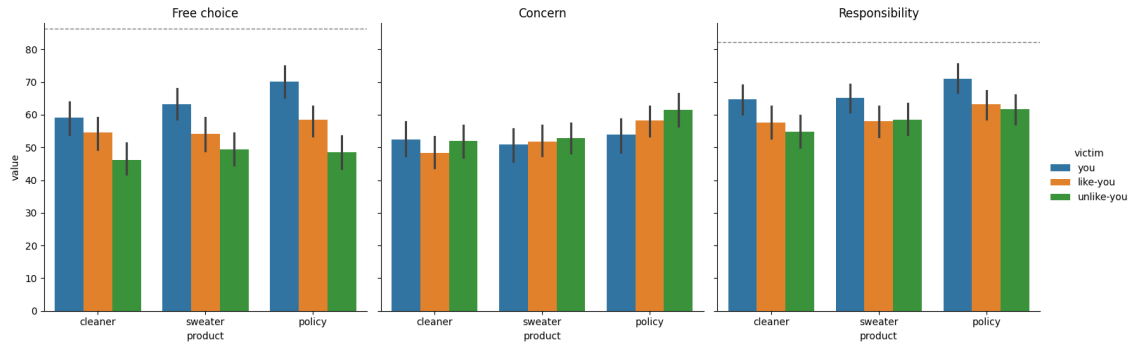
Figure 4 summarises the main results of Experiment 2. Free choice judgements are noticeably higher than in Experiment 1, where the manipulation was in the opposite direction of the victim's

[7] A sensitivity power analysis conducted using G*Power 3.1.9.4 and default settings suggested that with this sample size, a minimum effect size of n_p^2= .013 (f= .117) could be detected for between subjects factors (power=0.80, α = 0.05), n_p^2= .003 (f= .058) for within subjects factors and n_p^2= .004 for interactions (f= .065).

preferences. Still, we observe no ceiling effect, even in the "you" condition where the mean (M=64.16, SD=30.31) was much lower than the baseline free choice judgement observed in the preliminary experiment (M = 86.43, SD = 21.33). This indicates that participants still allow for the manipulation to play some role, making a dent on the level of free choice they assign to themselves. Both the object of choice and the victim show the opposite pattern to what was observed in Experiment 1. The pattern suggests that in the glass cleaner scenario, choices are judged to be less free compared to choices about sweaters, which are less free than choices about political candidates. Exactly the same pattern can be seen for both judgements of concern (same as in Experiment 1) and judgments of responsibility. Regarding the victim, the opposite pattern to Experiment 1 was found, though this time participants also distinguish between themselves and a person with identical preferences ("like-you" condition). So, as is clearly shown in the left panel of Figure 4, across objects of choice, participants judge themselves (blue bars) to be freer than other people with same preferences (orange bars), which are judged to be freer than people with opposite preferences (green bars). A similar pattern applies to responsibility (right panel), while judgements of concern don't seem as clearly affected by the identity and preferences of the victim (middle panel).

Figure 4

Mean ratings for free choice, concern and responsibility questions in Experiment 2 across objects of choice and victims.



Note: Error bars represent 95% confidence intervals. The horizontal dotted line represents the baseline value, i.e. the perceived degrees of free choice and responsibility in the absence of manipulation, as observed in the baseline experiment described above.

Mixed effects ANOVAs verify many of these observations: In the case of the freedom of choice, we find significant effects for the object of choice ($F(2,942)=11.31$, $p<.001$, $n_p^2=.023$) and victim ($F(2,471)=12.85$, $p<.001$, $n_p^2=.052$) and their interaction ($F(4,942)=3.26$, $p=.012$, $n_p^2=.014$). Follow-up t-tests for the object of choice show that the policy choice was deemed to be significantly more free than both the choice of glass cleaner ($t(473.00)=4.68$, $p<.001$, $d=.175$) and the choice of sweater ($t(473.00)=2.82$, $p=.005$, $d=.104$), but the other two didn't differ significantly from each other ($t(473.00)=1.90$, $p=.058$, $d=.071$). All pairwise comparisons between victims were significant (like-you -- unlike-you: $t(323.33)=2.44$, $p=.015$, $d=.270$, like-you -- you: $t(317.94)=2.74$, $p=.006$, $d=.304$, unlike-you -- you: $t(299.70)=5.27$, $p<.001$, $d=.606$).

A similar analysis for concern shows a significant main effect of the object of choice ($F(2,942)=18.06$, $p<.001$, $n_p^2=.037$) and a significant interaction effect ($F(4,942)=2.49$, $p=.042$, $n_p^2=.010$) but no significant effect of victim ($F(2,471)=.55$, $p=.578$, $n_p^2=.002$). Follow-up t-tests show that manipulation in policy choices is more concerning than both choices of glass cleaners ($t(473.00)=5.32$, $p<.001$, $d=.225$) and sweaters ($t(473.00)=4.54$, $p=.001$, $d=.192$) but again participants saw no difference in concern between sweaters and glass cleaners ($t(473.00)=.92$, $p=.360$, $d=.034$).

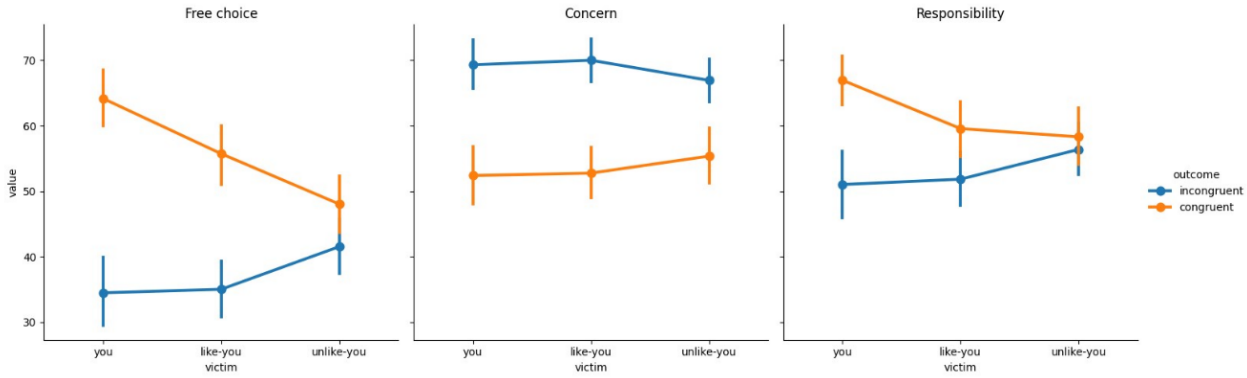
Finally, in the case of responsibility judgements there were main effects of victim ($F(2,471)=4.75$, $p=.009$, $n_p^2=.020$) and object of choice ($F(2,942)=18.87$, $p<.001$, $n_p^2=.039$) but no significant interaction ($F(4,942)=.59$, $p=.670$, $n_p^2=.003$). Follow-up t-tests for the object of choice show, as before, that the victim was deemed more responsible when choosing political candidates than both glass cleaners ($t(473.00)=5.42$, $p<.001$, $d=.212$) and sweaters ($t(473.00)=4.32$, $p<.001$, $d=.160$), but again choosing glass cleaners or sweaters didn't result in any significant difference ($t(473.00)=1.66$, $p=.098$, $d=.051$). All three dependent variables were significantly correlated with each other: free x concern: $r(1422)=-.26$, $p<.001$, free x responsibility: $r(1422)=.51$, $p<.001$, concern x responsibility: $r(1422)=-.12$, $p<.001$).

By averaging judgements across the three objects for each participant, a direct comparison can be made between Experiments 1 and 2, i.e. between the case when the manipulation and the outcome

are incongruent (Experiment 1) or congruent (Experiment 2) with the victim's preferences (Figure 5). A two-way ANOVA for free choice shows that the effect of the victim is not significant ($F(2,943)=2.41$, $p=.091$, $n_p^2=.005$) as the two opposite patterns cancel each other out, but both the direction of the outcome ($F(1,943)=100.97$, $p<.001$, $n_p^2=.097$) and the interaction ($F(2,943)=13.61$, $p<.001$, $n_p^2=.028$) are significant. Regarding concern, the main effect of outcome is significant ($F(1,943)=85.86$, $p<.001$, $n_p^2=.083$) but not that of victim ($F(2,943)=.04$, $p=.960$, $n_p^2=.000$) or the interaction ($F(2,943)=1.32$, $p=.268$, $n_p^2=.003$). For responsibility there was a significant main effect of outcome ($F(1.0)=20.63$, $p<.001$, $n_p^2=.021$) and of the interaction with victim ($F(2.0)=5.04$, $p=.007$, $n_p^2=.011$) but not a main effect of victim ($F(2.0)=1.33$, $p=.265$, $n_p^2=.003$).

Figure 5

Ratings of free choice, concern and responsibility by victim and by outcome in Experiment 1 (incongruent) and Experiment 2 (congruent)



Note: Error bars represent 95% confidence intervals.

Discussion

In motivating the second experiment, we suggested two extreme predictions about how participants would judge the impact of manipulations to free choice when the influence is in the same direction as the victim's preferences. We conjectured, that either consciousness is necessary, meaning that agents who are unaware of factors that influence their choices would be judged as not free, or the counterfactual is necessary, meaning that only when the stated preferences and the outcome are at

odds would participants judge those choices as not free. Here, we find some evidence for the latter. Across the board, free choice judgements were lower for incongruent (Experiment 1) compared to congruent outcomes (Experiment 2, see Figure 5). At the same time, it appears that consciousness still matters: The mean degree of perceived freedom found here is far below the values observed in the baseline experiment, where the outcome was the same, but no manipulation had taken place. This difference also addresses the possibility of participants conjuring that the manipulation was ineffective since there was no observed change in the outcome. Rather, it is likely that participants thought that the manipulation possibly acted as a double prevention (Lombrozo, 2010), stopping the victim from changing their mind, thus guaranteeing the outcome.

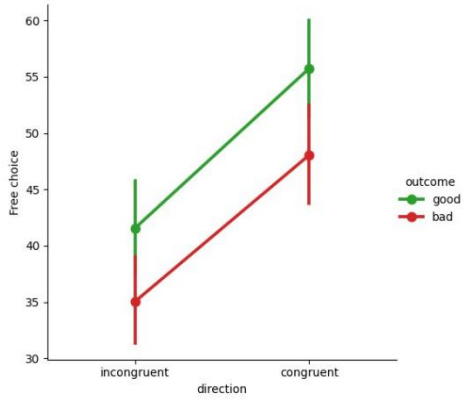
Overall, the findings indicate that both knowing what influences one's decisions, and the subversion of the "normal" course of events matter to participants, although the exact way these two interact remains unclear. To further explore the role of awareness, it is necessary to compare the present findings against a case where the agent is conscious of the external influence. Accordingly, Experiment 3 utilizes the same series of judgments but introduces a scenario where the manipulation—while leading to the same incongruent outcomes as Experiment 1—is transparent to the victim. We operationalise this conscious influence through a scenario of peer pressure.

The second major aim of the Experiment 2 was to evaluate the two possible explanations put forward regarding the difference in free choice ratings for the two types of "other" victims. In Experiment 1, the victim with opposite preferences who was manipulated into what participants saw as "positive" outcomes, was deemed freer. One possible explanation was that this was due to the victim's preferences, making them more psychologically distant to the participant and deemed more responsible and freer. The alternative explanation was that due to the valence of the outcomes, that is, because the outcome was desirable to the participant, the participant conjectured that the victim had a conscious change of mind, towards the "right" option. Figure 6, reframes the results of Experiments 1 and 2, focusing only on the conditions where the victim is some other person (not the participant). The outcomes are labelled from the perspective of the participant, such that "good" outcomes are scenarios

in which the victim ended up with what the participant liked independent of the victim's original preferences. The pattern suggests that agents are most free when their own preferences, good or bad, match the outcomes (i.e. are congruent), showing as before, that lack of freedom hinges on the counterfactual case being true. Beyond that, however, when we keep congruence (or the counterfactual truth value) constant, the victim is deemed freer when they end up with "good" outcomes, i.e. options that the participant likes the most (green line in Figure 6), compared to "bad" options that participant likes the least (red line in Figure 6).

Figure 6

The effect of the desirability of the outcome from the participant's perspective



Note: The figure shows free choice ratings for good vs bad outcomes (from the participant's perspective) following incongruent (exp1) or congruent (exp2) manipulation, focusing only on "others", i.e. victims with same (incongruent-bad/congruent-good) or different (congruent-bad/incongruent-good) preferences to the participant. Error bars represent 95% Cis.

Therefore, it appears that what might be driving free choice judgements in this case is less the psychological distance to the victim (although there might be some evidence in that direction in Experiment 2), and more the degree to which the participant approves of the outcome, even if the victim's actual preferences are subverted. Despite the non-rational undertones of this behaviour, in the General Discussion section we will offer an explanation on the basis of largely reasonable causal reasoning.

Experiment 3

Experiment 3 is designed to investigate the effect of outcome valence through judgements about other people's freedom, but the focus this time is to explore the role of consciousness, which has featured prominently in philosophical and policy debates. In Experiments 1 and 2, unconscious manipulation was found to reduce free will whether or not it is in line with the victim's preferences, albeit to different degrees. It is not clear, however, whether what matters for participants is that some manipulation takes place, or that the manipulation targets the unconscious, i.e. it is an influence that the victim is unaware of. Do people think that one's freedom is diminished even when they are aware of the manipulation tactic at the time when it happens, or does knowing the factors that influence one's decisions and actions suffice to ensure one's freedom?

While, as discussed earlier, many philosophers have focused on the relationship between free will and consciousness, others have been looking at the relationship between manipulation and autonomy irrespective of consciousness (Klenk, 2022; Noggle, 1996). Although often associated with deception (Cohen, 2018; Goodin, 1980), there are multiple examples of manipulation where the victim is aware of both the intentions and the techniques of the manipulator, e.g. guilt tripping, peer pressure, emotional blackmailing (Noggle, 2022). And although commonly assumed to be morally reprehensible for restricting one's autonomy, some have raised doubts about whether all cases of manipulation affect one's autonomy. Buss (2005) for example, describes a case where the aim of the manipulation is actually to increase a person's autonomy and freedom. What has not been investigated and will be the focus of Experiment 3, is how people judge this relationship. More specifically, the aim is to examine the extent to which people believe that any kind of manipulation threatens one's autonomy, or whether, the absence of awareness is a major factor explaining the judgements of diminished free choice observed so far. Again, the aim here is also to see whether the effects of the victim's identity and object of choice are observed in a scenario of peer-pressure, or whether consciousness is relevant for those effects as well.

Participants

We recruited 493 participants from the United States through Prolific and paid them £0.70 for taking part. 20 participants were later removed from the analysis for failing one or more of the three attentional checks. Their mean age of the remaining 473 participants was 42.37 years (SD = 13.29), and there were 249 females (53%), 219 males (46%) and 5 choosing "other"[8].

Design and procedure

The design was identical to Experiments 1 and 2, i.e. the victim ("you", "like-you", "unlike-you") was varied between participants and the object of the decision ("glass cleaner", "sweater", "policy") varied within.

The procedure was also very similar to previous experiments, differing only in the type of manipulation participants were asked to consider. So, after participants expressed their preferences, they were shown their most and least preferred options, as before. If they were in the "like-you" condition they were asked to imagine Frankie having the same preferences as them, or in the "unlike-you" condition they were asked to imagine Frankie having the opposite preferences. Rather than reading a microtargeting scenario though, this time participants were asked to imagine that the victim's friends informed the victim that they'd certainly buy or vote for the option the victim liked the least, and some of the victim's "closest friends laughed at the suggestion that one would even consider buying" or voting for the option the victim liked the most. Then, as in Experiment 1, participants were told that the victim, (which was themselves in the "you" condition) ended up buying or voting for the option most liked by the friends, but least liked by the victim. Finally, participants were asked for the same set of judgements as before and proceeded to the next set of objects.

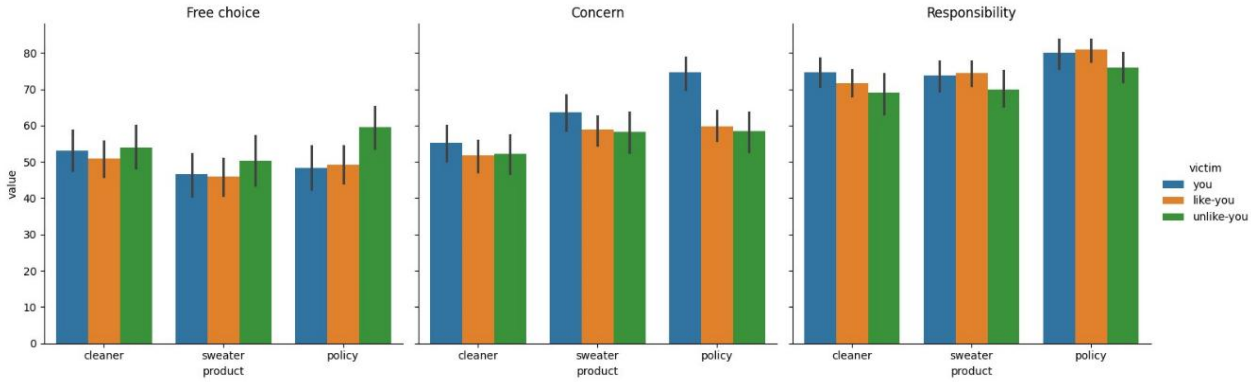
[8] A sensitivity power analysis conducted using G*Power 3.1.9.4 and default settings suggested that with this sample size, a minimum effect size of n_p^2= .013 (f= .117) could be detected for between subjects factors (power=0.80, α = 0.05), n_p^2= .003 (f= .058) for within subjects factors and n_p^2= .004 for interactions (f= .065).

Results

What is clear in Figure 7, is that the differences between victims and objects of choice are less pronounced this time, though not entirely absent. Free choice judgements are higher than in Experiment 1 (see also Figure 7 left panel for a comparison), but similar to Experiment 2, there are no ceiling effects, indicating that both the presence of manipulation and targeting the unconscious matter.

Figure 7

Mean ratings for the 3 questions (free choice, concern and responsibility) in Experiment 3, across objects of choice and victims.



Note: Error bars represent 95% confidence intervals.

For judgements of free choice, there was only a main effect of the object of choice ($F(2,940)=7.94$, $p<.001$, $n_p^2=.017$), this time driven by the sweater which received significantly lower ratings than both the glass cleaner choice ($t(472.00)=4.11$, $p<.001$, $d=.147$) and the choice of the policy to vote for ($t(472.00)=2.73$, $p=.007$, $d=.110$), since the policy didn't differ from the cleaner ($t(472.00)=.85$, $p=.393$, $d=.033$). Although the effect of victim was not significant across objects of choice ($F(2,470)=1.32$, $p=.268$, $n_p^2=.006$), the interaction did approach significance ($F(4,940)=2.12$, $p=.077$, $n_p^2=.009$) and the pattern was remarkably similar to what we found in Experiment 1 (see Figure 8, left panel for comparison). A closer look to the policy question in particular[9] (Figure 7), reveals the same

[9] We have preregistered a special analysis of the policy condition on the basis that it raises more concern across the board, and that the repeated nature of our design may introduce noise (https://osf.io/zrvkt). In the interest of

differences we observed in Experiment 1. Specifically, when what is manipulated is the choice of a political candidate, free choice ratings in the "unlike-you" condition are significantly higher than the ratings in both "you" ($t(266.74)=2.57$, $p=.011$, $d=.298$), and "like-you" ($t(250.10)=2.54$, $p=.012$, $d=.295$) conditions, while "you" and "like-you" do not differ significantly ($t(348.77)=.21$, $p=.832$, $d=.022$).

Regarding concern, there was a significant main effect of victim ($F(2,470)=5.45$, $p=.005$, $n_p^2=.023$), driven by participants reporting that they themselves would be more concerned than other people with the same ($t(352.66)=2.88$, $p=.004$, $d=.304$) or the opposite preferences ($t(248.37)=2.69$, $p=.008$, $d=.321$). There was an even stronger effect of the object of choice ($F(2,940)=41.95$, $p<.001$, $n_p^2=.082$) with concern about voting choices being significantly higher than choices of sweaters ($t(472.00)=3.45$, $p=.001$, $d=.148$), and glass cleaners ($t(472.00)=8.15$, $p<.001$, $d=.390$), while sweater choices were significantly higher than choices of glass cleaners ($t(472.00)=6.14$, $p<.001$, $d=.243$).

Finally, the mixed effects ANOVA for responsibility revealed only a main effect of object of choice ($F(2,940)=31.83$, $p<.001$, $n_p^2=.063$), driven by the policy choice being significantly higher than both the choice of sweater ($t(472.00)=6.28$, $p<.001$, $d=.250$) and the choice of glass cleaner ($t(472.00)=6.86$, $p<.001$, $d=.286$), while there was no significant difference between the sweater and the glass cleaner ($t(472.00)=1.11$, $p=.269$, $d=.037$). There was no main effect of victim ($F(2,470)=1.53$, $p=.217$, $n_p^2=.006$) and no significant interaction ($F(4,940)=.93$, $p=.445$, $n_p^2=.004$). Again the three ratings were significantly correlated with each other (free x concern: $r(1419)=.14$, $p<.001$, free x responsible: $r(1419)=.24$, $p<.001$, concern x responsible: $r(1419)=.26$, $p<.001$), although note that unlike the previous two experiments, concern ratings and responsibility ratings were positively correlated (Experiment 1: negative non-significant, Experiment 2: negative significant).

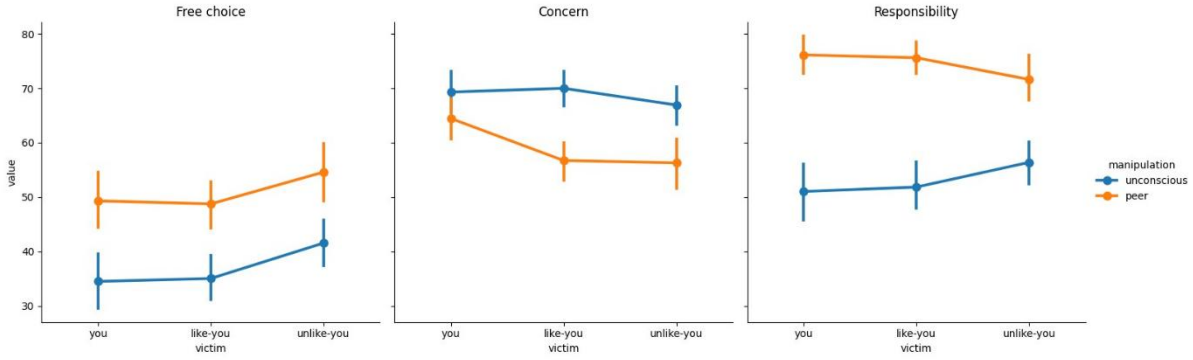
Figure 8 compares the two types of manipulation described in Experiments 1 and 3, across items. We can see that compared to unconscious manipulation, choices are judged to be freer and less concerning, while the victim is seen to be more responsible when they are aware of the manipulation. A

brevity, no special mention was made when discussing the other two experiments, since the results were in the same direction as the omnibus tests.

two-way ANOVA on free choice judgements shows significant main effects for both victim F(2.0)=4.31, p=.014, n_p^2=.009) and type of manipulation (F(1.0)=49.26, p<.001, n_p^2=.050) but no significant interaction (F(2.0)=.06, p=.938, n_p^2=.000). The same was the case for judgements of concern, showing significant main effects (victim: F(2.0)=4.22, p=.015, n_p^2=.009), manipulation: F(1.0)=38.18, p<.001, n_p^2=.039) but not a significant interaction (F(2.0)=2.51, p=.082, n_p^2=.005). Finally, there was a main effect of manipulation type for judgements of responsibility (F(1.0)=167.56, p<.001, n_p^2=.151) and a significant interaction effect (F(2.0)=3.22, p=.041, n_p^2=.007) but no significant main effect of victim (F(2.0)=.15, p=.858, n_p^2=.000).

Figure 8

Comparison of Experiments 1 and 3, where the manipulation was respectively unconscious or not, across victims and judgements.



Note: Error bars represent 95% confidence intervals.

Table 1

Mean ratings (SD) for the three DVs across all three experiments, collapsed across objects of choice

Outcome	Manipulation	Victim	Free choice	Responsibility	Concern
Congruent	None (Baseline)	You	86.43 (21.33)	82.14 (28.10)	-
	Unconscious	You	64.16 (30.31)	66.96 (27.51)	52.41 (32.41)
		Like-you	55.73 (31.82)	59.58 (30.71)	52.77 (31)
		Unlike-you	48.04 (32.14)	58.31 (30.47)	55.37 (31.56)
Incongruent	Unconscious	You	34.50 (32.27)	51.03 (31.21)	69.31 (29.15)
		Like-you	35.04 (30.30)	51.85 (30.4)	70.01 (27.29)
		Unlike-you	41.55 (31.93)	56.39 (30.56)	66.93 (29.56)
	Conscious	You	49.31 (38.99)	76.15 (28.2)	64.46 (32.9)
		Like-you	48.75 (35.55)	75.63 (23.76)	56.73 (29.16)
		Unlike-you	54.60 (32.63)	71.63 (25.28)	56.3 (28.27)

Discussion

The results of Experiment 3 (conscious awareness of manipulation) taken together with the findings from Experiment 1 (unconscious manipulation) and 2 (ambiguous influence of unconscious manipulation) point towards two possible conclusions. First, a lack of awareness is perceived to diminish free choice. Second, even when the victim is aware of the influence, any outcome that subverts their stated preferences is still seen as a threat to autonomy. This leaves open for future research the question whether a consciously accessible manipulation (as in Experiment 3) that aligns with a victim's preferences (as in Experiment 2) would be viewed as having no impact on free choice. Or, conversely, whether free will is perceived as inherently threatened by any form of manipulation, regardless of its transparency or its congruence with one’s desires.

Regarding the effects of the victim and the object of choice in free choice judgements, we observed some differences compared to the previous experiments. First, the statistical analysis showed no difference between victims that are aware of the possible influences on their behaviour. Still, the

trend was in the same direction as in previous experiments, and the analysis particularly for the political choice condition turned out to be significant. One possible explanation for the weaker effect is that awareness allows for deliberation, and given the chance to deliberate, the general perspective people take is that everyone is equally free irrespective of their preferences. In the section that follows we will return to this point, to compare the role of victim identity and valence across all three experiments.

In the case of the object of choice, although there is a strong effect on judgements of free choice, Experiment 3 is different from both Experiments 1 and 2, where we observed a gradual increase (Experiment 1) or decrease (Experiment 2) of free choice ratings following the "importance" of the choice (cleaner<sweater<policy). In Experiment 3, when the victim is aware of the manipulation, it is the choice of sweater that leads to the least free choices compared to both political and glass cleaner choices. One tentative explanation is that for the least concerning choices or choices that least reflect the identity of the agent (glass cleaner), the victim can allow themselves to be swayed while remaining free, while for choices that carry the most responsibility (political), a person should not allow themselves to be influenced, thus also remaining free. On this interpretation, choices about fashion items hit the sweet spot, where participants expect peer pressure to have the stronger influence on them, and to reduce their freedom the most. What is most noteworthy of all, is that conscious awareness of a manipulation leads to very different considerations when evaluating choice freedom.

General Discussion

Across three studies, we investigated people's views regarding freedom of choice in the face of manipulation, and especially when it takes place below conscious awareness. We used examples drawn from those volunteered by participants (see Osman, 2020, Osman & Bechlivandis, 2024) and designed the judgement tasks to be as relatable as possible. Participants were asked to express their actual preferences before imagining scenarios in which these preferences have been manipulated. The present set of experiments stand in contrast to previous folk belief x-phi work that often relies on extreme unrealistic situations disconnected from people's lives and their day to day understanding of free choice and manipulation. We instead presented situations in which participants could bring to bear their

intuitions, to respond to relatable situations, cases of manipulations they are likely to have personal experience of[10].

Overall, the findings suggest that concrete details such as the object of the manipulated choice and the identity of the victim of manipulation affect people's judgements. Starting from the former, we find an effect in the direction that deep self theories (Frankfurt, 1971; Sripada, 2016; Watson, 1975; Wolf, 1990), or theories that expect a strong link between free will and responsibility (Dennett & Caruso, 2021; Fischer & Ravizza, 1998) would predict. Specifically, choices that are closer to personal values are judged as less free when the outcome contradicts stated preferences and could have been the result of unconscious manipulation. The trend is reversed when the unconsciously manipulated outcome is in the direction of one's preferences, with people judging congruent outcomes of more important choices indicative of greater freedom. Participants consistently expressed higher degrees of concern, when more important choices are manipulated, while judgements of responsibility were less influenced by the type of choice.

The observed negative correlation between ratings of free choice and ratings of concern (i.e. the more concerned, the less free when manipulated), suggests a pragmatic explanation about the effect of the object of choice in judgements of freedom under unconscious manipulation. First, it can be argued that one is expected to attend less to choices one cares less about. If so, then for the least consequential choices, one might implicitly consent by not paying attention to whom and what may be influencing those choices. In this case one can remain free by consciously deciding to allow themselves to be unconsciously manipulated (e.g. influenced by marketing jingles in a supermarket; see also Sunstein, 2014b). Whether decisions earlier in the causal chain would be associated with assuming responsibility (due to the conscious decision) or relinquishing it (due to the unconscious manipulation), was not tested here. We nevertheless believe that the relationship between the timing of manipulation and deliberation, and their effect of judgements of freedom and responsibility is a very fertile area for

[10] Of course, an experiment, whether conducted in the lab or online, almost always departs from the real world in some sense. We still believe that there are degrees to this departure and hope to see even more ecologically valid approaches in this area, e.g. through field experiments.

future investigation. Still, we will return to the effect of the object of choice to suggest another, more parsimonious, though not necessarily competing explanation, after summarising the rest of our findings.

In regard to the victim, participants judge that the level of concern and responsibility would be largely similar for all victims across manipulations[11]. However, participants judge that the degree to which a manipulated choice is free depends on who is making it and what the outcome is. The point to emphasis here is that the victim is seen to be freer when the manipulated outcome is desirable for the participant, despite going against the victim's preferences. On that basis, for example, an opponent of the death penalty would view a supporter as freer when finding out they eventually voted against the death penalty, despite knowing that they were unconsciously manipulated. Though weakened for choices that do not concern politics, the same effects are still detectable when the victim is aware of the manipulation. We proposed two alternatives to explain this effect, either based on initial preferences, psychological distance, and attributions of responsibility, or based on outcome valence from the participant's perspective. The results of Experiment 2 where the preference, the manipulation and the outcome were in the same direction, lend support to the valence-based explanation. Despite appearances, we think that a rational explanation can account for the observed behaviour, as discussed below.

This research sought to investigate people's intuitions for their inherent value and for their potential to provide evidence that could be used to inform public debate and public policy making (Council Directive 2018/1808/EC, 2018; Halpern & Costa, 2019). However, some of our hypotheses were inspired by philosophical theorising, and, therefore, some of our conclusions are relevant to philosophical debates. We found that, as broadly suggested (Nahmias et al., 2020), consciousness is relevant to free will judgements, while being neither necessary nor sufficient to it. When manipulations bypass consciousness but are in the same direction as the victim's preferences (Experiment 2), the victim is less free, than in the case of the victim who is aware of the manipulation but ends up with the

[11] Two exceptions: participants judge that themselves would be more concerned than others in peer manipulation (Experiment 3) and more responsible in congruent manipulation (Experiment 2)

option they don't like (Experiment 3). In that sense, the counterfactual, i.e. knowing that the victim would have done differently had there been no manipulation seems more important than consciousness when attributing free will (Frankfurt, 1969). It is interesting though, that even in the case where the counterfactual is false, when the manipulated outcome is in the same direction as the victim's preferences (Experiment 2), we still didn't observe a ceiling effect in free choice judgements. Since our aim was not to inform philosophical debates, more targeted work is needed to address such issues. We do underline, however, the importance of keeping the scenarios relatable for the people whose intuitions we seek.

In summary, we find that freedom of choice is affected by the importance and the outcome of a choice under manipulation. The outcome matters not only in terms of whether it is in line with the victim's preferences but, critically, in terms of whether it accords with the preferences of the participant /observer. Although we suggested ways to explain each of these effects separately, we now turn to a more parsimonious explanation, based on diagnostic causal reasoning under uncertainty (Meder et al., 2014).

In order to make free choice attributions in the presented scenarios, participants are deciding the extent to which the outcome is caused by the agent (free choice) or the manipulator (unfree choice). Despite the scenarios trying to make a convincing case for the role of the manipulation, which in the vignettes is described as "very effective" and inaccessible (thus arguably undefendable), it is clear that participants are taking other factors into account. Since the factors that influence decision-making are generally latent, it makes sense to rely predominantly on diagnostic reasoning, i.e. to infer the causes of some behaviour based on the observable outcome. Still, if it were this simple, then the congruence between the outcome and the stated preferences would be all that mattered, with congruent choices deemed free and incongruent unfree.

The variability in the observed judgements of freedom suggests that rather than simply deciding between the agent and the manipulator, participants are considering a third causal pathway, the possibility that between the original preferences and the final decision, one may change their mind. In

that case, the task of estimating the degree of choice freedom, becomes the task of estimating the probability that one has consciously changed their mind, or in the congruent case (Experiment 2), the probability that one was prevented from doing so. On this basis, it makes sense that a person is more likely to change their mind when they are aware of the manipulation (explaining higher free choice judgements in Experiment 3 than Experiment 1). The same would apply when making an important decision (explaining the effect of object of choice on free choice judgements in Experiments 1 and Experiment 2). Finally, it is not necessarily irrational for an observer to think that one is more likely to change their mind towards the option that is "superior", in the observer's view. Although in matters of morality or aesthetics, normatively this should not be the case, for more objective properties it is easier to see why that would be reasonable. If, for example, the choice was between two identical glass cleaners, with one sold at a higher price, we would expect participants to judge that the victim manipulated into buying the cheap option is freer, because the estimated probability that they changed their mind would be higher. Even so, the perception that one's political or moral views are objectively superior is, of course still questionable. Future work should investigate this effect by examining, for example, the degree to which people would insist on these attributions when asked to directly compare them (e.g. in a within groups design) and possibly attempt to directly gauge the estimated probability that a victim changed their mind in each of the presented scenarios.

Why do people bring other considerations to bear in this task? After all, we asked them to imagine a situation which was as clear, direct, and, in some sense, as extreme as possible: Seconds after expressing their preferences, participants are told that themselves, or someone in very similar circumstances, was manipulated without knowing, not merely into a suboptimal option (e.g. their 2nd preference), but into the option they liked the least, one that they wouldn't conceivably go for. Why then do participants look for alternative possibilities instead of accepting that the outcome indicates the success of the manipulation, and therefore their lack of freedom? The present findings suggest a reluctance to attribute behaviours to unconscious influences (see also Genschow et al., 2023). Despite the current consensus that is based either on tacit assumptions or on empirical findings in response to

imaginative and abstract scenarios, participants strongly resisted experimental attempts to encourage them to assign a decisive role to the unconscious and accept the absence of free choice. Instead, participants appeared to embark on a thorough search for causal factors in order to preserve freedom of choice. Only when that search failed (e.g. when they are told that themselves have voted in support of the policy they disagreed most with), do they accept an explanation where freedom is absent or seriously diminished.

On the basis of parsimony, the proposed explanation that is presented here and is based on diagnostic causal reasoning and the estimation of the probability of a conscious change of mind, is better placed to account for the set of results we have obtained. At its core is the possibility that rather than being irrational, participants entertain a different causal model than the experimenters (i.e. ourselves) have intended. As such, it highlights a common danger in experimental psychology and philosophy, namely, to diagnose suboptimal reasoning capacities, biases or motivated reasoning, whenever participants do not perform in accordance with some normative, but often inadequate standard. Irrespective of the particular proposal we've put forward, the nuanced way in which participants make attributions of free choice in our experiments, shows that the strategy of presenting participants with imaginative thought experiments and expecting to reach useful generalisable conclusions is suboptimal. In response to issues around ecological validity highlighted in the introduction, we show here that people take a variety of factors into account when reasoning about concrete cases, and sometimes they bring their own motivations to bear, e.g. to preserve free choice. Such considerations would remain latent if we insisted on probing people for intuitions on what, from their perspective, are purely theoretical matters.

The present study examined the impact of manipulative tactics designed to influence choices through both unconscious targeting and conscious pressure. By varying the victim of the manipulation and the nature of the choice—ranging from low-stakes consumer items to identity-relevant political voting—we found that participants make highly nuanced judgments regarding free choice, concern, and responsibility. What informs these judgments is an attempt to identify the causal determinants of the

choice, in which the alignment between subjective or 'normative' preferences and the final outcome weighs heavily. Overall, these findings suggest that people evaluate agency through a rational, causal-analytic process, seeking explanations for why an outcome does or does not match the expectations of the individual making the choice.

References

Baumeister, R. F., Masicampo, E. J., & Dewall, C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, *35*(2), 260–268. https://doi.org/10.1177/0146167208327217

Block, N. (1995). Some Concepts of Consciousness. *Behavioral and Brain Sciences*, *18*(2), 272–287.

Bodó, B., Helberger, N., & De Vreese, C. H. (2017). Political micro-targeting: A manchurian candidate or just a dark horse? *Internet Policy Review*, *6*(4), 0–13. https://doi.org/10.14763/2017.4.776

Buss, S. (2005). Valuing autonomy and respecting persons: Manipulation, seduction, and the basis of moral constraints. *Ethics*, *115*(2), 195–235. https://doi.org/10.1086/426304

Cohen, S. (2018). Manipulation and Deception. *Australasian Journal of Philosophy*, *96*(3), 483–497. https://doi.org/10.1080/00048402.2017.1386692

Council Directive 2018/1808/EC on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities (2018), Official Journal L303 p. 69

Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly*, *47*(1), 1–15.

Dennett, D. C., & Caruso, G. D. (2021). *Just deserts: Debating free will*. John Wiley & Sons.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. Psychological bulletin, 130(6), 959.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility.* Cambridge University Press.

Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy*, *66*(23), 823–839. https://doi.org/10.2307/2023833

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, *68*(1), 5–20.

Frankfurt, H. G. (1999). *Necessity, Volition, and Love*. Cambridge University Press.

Genschow, O., Cracco, E., Schneider, J., Protzko, J., Wisniewski, D., Brass, M., & Schooler, J. W. (2023). Manipulating Belief in Free Will and Its Downstream Consequences: A Meta-Analysis. *Personality and Social Psychology Review*, *27*(1), 52–82. https://doi.org/10.1177/10888683221087527

Genschow, O., Rigoni, D., & Brass, M. (2017). Belief in free will affects causal attributions when judging others' behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(38), 10071–10076. https://doi.org/10.1073/pnas.1701916114

Goodin, R. E. (1980). *Manuipulatory Politics*. Yale University Press.

Halpern, D., & Costa, E. (2019). *The behavioural science of online harm and manipulation, and what to do about it: an exploratory paper to spark ideas and debate.* https://www.bi.team/wp-content/uploads/2019/04/BIT_The-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it_Single.pdf

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiveing the causes of behavior* (pp. 79–94). General Learning.

Kane, R. (1985). *Free will and values*. State University of New York Press.

Klein III, R. E., Kleine, S. S., & Kernan, J. B. (1993). Mundane Consuption and the Self: A Social-Identity Perspective. *Journal of Consumer Psychology*, *2*(3), 209–235. http://www.sciencedirect.com/science/article/pii/S1057740808800150

Klenk, M. (2022). (Online) manipulation: sometimes hidden, always careless. *Review of Social Economy*, *80*(1), 85–105. https://doi.org/10.1080/00346764.2021.1894350

Knobe, J., & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and*

the Cognitive Sciences, *7*(1), 67–83. https://doi.org/10.1007/s11097-007-9066-y

Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive psychology*, *61*(4), 303-332.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121*(3), 277–301. https://doi.org/10.1037/a0035944

Murray, D., & Lombrozo, T. (2017). Effects of Manipulation on Attributions of Causation, Free Will, and Moral Responsibility. *Cognitive Science*, *41*(2), 447–481. https://doi.org/10.1111/cogs.12338

Murray, D., & Nahmias, E. (2014). Explaining away incompatibilist intuitions. *Philosophy and Phenomenological Research*, *88*(2), 434–467. https://doi.org/10.1111/j.1933-1592.2012.00609.x

Nahmias, E., Allen, C. H., & Loveall, B. (2020). When Do Robots have Free Will? Exploring the Relationships between (Attributions of) Consciousness and Free Will. In *Value Inquiry Book Series* (Vol. 338, pp. 57–80). https://doi.org/10.1163/9789004409965_005

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, *18*(5), 561–584. https://doi.org/10.1080/09515080500264180

Nichols, S., & Knobe, J. (2007). *Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions*. *41*(4), 663–685.

Noggle, R. (1996). Manipulative actions: A conceptual and moral analysis. *American Philosophical Quarterly*, *33*(1), 43–55. http://www.jstor.org/stable/20009846

Noggle, R. (2021). Manipulation in Politics. In *Oxford Research Encyclopedia of Politics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228637.013.2012

Noggle, R. (2022). The Ethics of Manipulation. In Zalta Edward N. (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2022/entries/ethics-manipulation/

Osman, M. (2020). Overstepping the boundaries of free choice: Folk beliefs on free will and determinism in real world contexts. *Consciousness and Cognition*, *77*(September 2019), 102860.

https://doi.org/10.1016/j.concog.2019.102860

Osman, M., & Bechlivanidis, C. (2021). Public Perceptions of Manipulations on Behavior Outside of Awareness. *Psychology of Consciousness: Theory Research, and Practice*. https://doi.org/10.1037/cns0000308

Osman, M., & Bechlivanidis, C. (2022). Impact of Personalizing Experiences of Manipulation Outside of Awareness on Autonomy. *Psychology of Consciousness: Theory Research, and Practice*. https://doi.org/10.1037/cns0000343

Osman, M., & Bechlivanidis, C. (2023). Folk Beliefs About Where Manipulation Outside of Awareness Occurs, and How Much Awareness and Free Choice Is Still Maintained. *Psychology of Consciousness: Theory, Research, and Practice*. https://doi.org/https://doi.org/10.1037/cns0000379

Pereboom, D. (2006). *Living Without Free Will*. Cambridge University Press.

Perloff, R. M. (1999). The Third-Person Effect: A Critical Review and Synthesis. *Media Psychology*, *1*(4), 353–378. https://doi.org/10.1207/s1532785xmep0104_4

Phillips, J., & Shaw, A. (2015). Manipulating Morality: Third-Party Intentions Alter Moral Judgments by Changing Causal Reasoning. *Cognitive Science*, *39*(6), 1320–1347. https://doi.org/10.1111/cogs.12194

Polman, E., & Maglio, S. J. (2023). Improving the Generalizability of Behavioral Science by Using Reality Checks: A Tool for Assessing Heterogeneity in Participants' Consumership of Study Stimuli. *Perspectives on Psychological Science*, *18*(4), 955–975. https://doi.org/10.1177/17456916221134575

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 173–220). Academic Press.

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is belief in free will a cultural universal?. *Mind & Language*, *25*(3), 346-358.

Schmaltz, R. M. (2016). Bang Your Head: Using Heavy Metal Music to Promote Scientific Thinking in the

Classroom. *Frontiers in Psychology*, *7*(February), 1–4. https://doi.org/10.3389/fpsyg.2016.00146

Shepherd, J. (2012). Free will and consciousness: Experimental studies. *Consciousness and Cognition*, *21*(2), 915–927. https://doi.org/10.1016/j.concog.2012.03.004

Shepherd, J. (2015). Consciousness, free will, and moral responsibility: Taking the folk seriously. *Philosophical Psychology*, *28*(7), 929–946. https://doi.org/10.1080/09515089.2014.962018

Sripada, C. (2012). What Makes a Manipulated Agent Unfree? *Philosophy and Phenomenological Research*, *85*(3), 563–593. https://doi.org/10.1111/j.1933-1592.2011.00527.x

Sripada, C. (2016). Self-expression: a deep self theory of moral responsibility. *Philosophical Studies*, *173*(5), 1203–1232. https://doi.org/10.1007/s11098-015-0527-9

Stillman, T. F., Baumeister, R. F., & Mele, A. R. (2011). Free will in everyday life: Autobiographical accounts of free and unfree actions. *Philosophical Psychology*, *24*(3), 381–394. https://doi.org/10.1080/09515089.2011.556607

Stracqualursi, V. (2020). *Trump campaign microtargeted Black Americans disproportionally 'to deter' them from voting in 2016 election, Channel 4 reports*. CNN. https://edition.cnn.com/2020/09/29/politics/trump-2016-campaign-voter-deterrence/index.html

Strawson, G. (1994). The Impossibility of Moral Responsibility. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *75*(1/2), 5–24.

Strawson, P. (1962). Freedom and Resentment. *Proceedings of the British Academy*, *48*, 187–211.

Strudler, A. (2005). Deception Unraveled. *The Journal of Philosophy*, *102*(9), 458–473. https://www.jstor.org/stable/3655633

Sunstein, C. R. (2014a). The Ethics of Nudging. *SSRN Electronic Journal*, *30*(413). https://doi.org/10.2139/ssrn.2526341

Sunstein, C. R. (2014b). Choosing not to choose. *Duke Law Journal*, *64*(1), 1–52.

Susser, D., Roessler, B., & Nissenbaum, H. F. (2019). Online Manipulation: Hidden Influences in a Digital World. *Georgetown Law Technology Review*, *4*. https://heinonline.org/HOL/Page?handle=hein.journals/gtltr4&div=6&g_sent=1&casa_token=&col

lection=journals

Sytsma, J., & Machery, E. (2009). How to study folk intuitions about phenomenal consciousness. *Philosophical Psychology*, *22*(1), 21–35. https://doi.org/10.1080/09515080802703653

Sytsma, J., & Snater, M. (2023). Consciousness, Phenomenal Consciousness, and Free Will. *Advances in Experimental Philosophy of Action*, 1–26. https://doi.org/10.5040/9781350266353.0007

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

Tallis, F. (2002). *Hidden minds: A history of the unconscious*. Arcade Publishing.

Tappin, M. Ben, Wittenberg, C., Hewitt, B. L., Berinsky, J. A., & Rand, G. D. (2023). Quantifying the potential persuasive returns to political microtargeting. *Proceedings of the National Academy of Sciences*, *120*(25). https://doi.org/https://doi.org/10.1073/pnas.2216261120

Trope, Y., & Liberman, N. (2010). Construal-Level Theory of Psychological Distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Watson, G. (1975). Free Agency. *The Journal of Philosophy*, *72*(8), 205–220.

Weaver, M. (2018). *Social media "micro-targeting" of voters on the increase, MPs told*. The Guardian. https://www.theguardian.com/media/2018/jan/23/social-media-micro-targeting-of-voters-on-the-increase-mps-told

Wolf, S. (1990). *Freedom within reason*. Oxford University Press.