UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

EDUCATED DECISION-MAKERS ARE HARDER TO BIAS: COMPARING EDUCATION AND NUDGES ON RELIABLE DELIBERATION ABOUT RECYCLED WATER

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

BRADEN TANNER

Norman, Oklahoma

2023

EDUCATED DECISION-MAKERS ARE HARDER TO BIAS: COMPARING EDUCATION AND NUDGES ON RELIABLE DELIBERATION ABOUT RECYCLED WATER

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF PSYCHOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Adam Feltz, Chair

Dr. Edward Cokely

Dr. Scott Gronlund

Dr. Dingjing Shi

Dr. David Sabatini

Dr. Jason Vogel

Acknowledgments

I want to express my sincere gratitude to all those who have helped me get to this point and who have made this project possible. First and foremost, to Dr. Adam Feltz, thank you for being a fantastic mentor. Your enthusiasm for knowledge and quality research is both inspiring and contagious, and your feedback, encouragement, and guidance, have helped me grow in ways I did not think would be possible five years ago. I truly could not have asked for a better advisor. To Dr. Edward Cokely, thank you for pushing me to be my best. Your ability to push me out of my comfort zone while still providing consistent encouragement was fundamental in helping me develop the skills I needed to be successful. To Dr. Scott Gronlund and Dr. David Sabatini, thank you for the encouragement and feedback you have provided over the years. Your calm and caring natures have always uplifted me, and your truly exemplary careers have given me great hope for what I may be able to accomplish in the future. To Dr. Dingjing Shi and Dr. Jason Vogel, thank you for joining us on this journey, always being willing to ask important questions, and providing resources that have helped me along the way. To my wife, Bridget, thank you for being my biggest cheerleader and for working all the crazy hours and jobs that you have to support us throughout this process. And finally, thank you to all my family, friends, and colleagues—especially the members of the Ethical Interactions Lab. Your encouragement and support have meant the world to me, and I cannot imagine where I'd be without you.

Table of Contents

Abstract

Recently, it has become important to define ethical rules governing the most appropriate ways to interact with those that decision interventions seek to influence. However, philosophical theories and current methods of comparing interventions in the decision sciences are likely inadequate to fully demonstrate ethical differences among various types of interventions. Here, I seek to address this problem by first proposing a framework (based on the American Psychological Association's ethical code of conduct) for criteria that should likely be considered when making ethical comparisons among interventions. Then, I propose a method through which one could identify the extent to which nudges and educational interventions promote reliable deliberation—a condition that has been argued to be central to ensuring autonomous decision making—when making decisions regarding recycled water. In four experiments, I show that nudges and educational interventions can both be used to shift individual preferences regarding recycled water, but that only educational interventions result in greater choice consistency, a factor that I propose is integral to demonstrating one has deliberated reliably. These results are likely of important practical benefit, as they might guide policymakers and water practitioners towards interventions that are likely to result in consistent and stable public support of recycled water, which could help avoid costly consequences such as protests and legal challenges. They likely also have important ethical implications, as they demonstrate a condition on which some nudges might fail to respect autonomy to the same extent as educational interventions. These results help us move one step closer to being able to empirically quantify ethical risks and benefits of using different intervention strategies and represent an important stepping-stone in defining an integrated ethical interaction theory.

Introduction

Insights from the behavioral sciences seeking to inform interventions in a way that helps people make "better" decisions have become more pronounced in governments, corporations, and organizations (D. Halpern & Sanders, 2016; OECD, 2017). However, what constitutes a "better" decision—and, by extension, what constitutes an effective intervention—has recently come under substantial philosophical and empirical scrutiny. While good decisions are often characterized by exogenously identified criteria (e.g., if the decision-maker makes a choice that was normatively correct or consistent with logical rules; see Baron, 2004; Tversky & Kahneman, 1974)—and thus effective interventions are often characterized by the extent to which they efficiently help people make a decision in line with those rules (e.g., Benartzi et al., 2017)—other factors can (and likely should) be considered. Among these factors, there is a growing consensus that it has become important to define ethical rules governing the most appropriate ways to interact with those that decision interventions seek to influence. Various professional organizations acknowledge this need. For example, the American Psychological Association (APA) lists respect and promotion of human rights as one of its guiding principles (American Psychological Association & American Psychological Association Services, Inc., 2019). Accordingly, the APA has issued and continually amends an ethical code of conduct, with the goal of ensuring the welfare and protection of all those with whom psychologists work (American Psychological Association, 2017).

While many theoretical discussions regarding intervention ethics in the decision sciences have occurred (e.g., Blumenthal-Barby & Burroughs, 2012; Bovens, 2009; Hausman & Welch, 2010; Selinger & Whyte, 2011; Sunstein, 2015a, 2015b), current methods suffer from at least two notable limitations. First, most discussions seem to focus on ethics in a broad sense without defining relevant criteria on which ethics might be judged. And second, very little effort has been made to empirically demonstrate and quantify potential ethical costs along those specific criteria. Here, I begin to address these limitations by laying the groundwork for a theory of ethical interactions. First, using the APA ethical code of conduct as a guide, I suggest a set of foundational principles that could be used as criteria for ethical comparisons. Then, I propose and demonstrate a method that could be used to compare interventions along one of these criteria—the reliable deliberation condition of autonomy—as it relates to decisions and preferences concerning recycled water.

Overall, in a series of four experiments, I provide evidence that many kinds of decision interventions (e.g., many nudges) can be used to bias participants toward choosing to accept recycled water. However, I also show that compared to libertarian paternalistic interventions like nudges, informing people about recycled potable water is better able to promote consistent preferences regarding recycled water. In addition to the practical benefit that such preference consistency could provide in this domain, these results suggest that some libertarian paternalistic nudges may not support reliable deliberation as well as some educational interventions. This could have important ethical implications related to respect for individual autonomy.

Criteria for Ethical Interactions

In order to make empirical comparisons among different kinds of decision interventions, it is first necessary to identify relevant ethical criteria on which those comparisons are to be made. One problem in identifying these criteria is that there are many current philosophical theories about what makes right actions right (Copp, 2006), none of which are likely completely suitable for the purposes of comparing interventions. An exhaustive list of each of these views is outside the scope of this dissertation; however, three main theories of note are consequentialism, deontology, and virtue ethics.

Consequentialism is a theory of ethics that proposes that the rightness or wrongness of an act depends on its consequences (see Scheffler, 1988). Among the early consequentialist views were classic views of utilitarianism, which holds that an act can only be considered morally good if its net good consequences outweigh its net bad consequences for everyone affected and out of all options available (Bentham, 1825; Mill, 1863; Sidgwick, 1874). While on the surface, consequentialism seemingly has direct applications to decision intervention ethics (i.e., one could use models of expected utility to estimate net good and net bad consequences of a decision, and all things being equal, the intervention that leads the most people to a choice where utility is maximized could be considered more ethical), there are limitations to this theory that would likely be problematic.

First, there are many different ways that good and bad consequences can be conceptualized, and by extension, many different ways that utility could be measured. For example, consequences could refer only to *actual, present* consequences of a decision, but it could also refer to *foreseeable future* consequences. Likewise, they could refer to consequences to an *individual* decision-maker, but they could also refer to consequences for *all* people or beings (Sinnott-Armstrong, 2022). Similarly, it is not clear exactly what utility is (e.g., multiple characterizations of utility exist, including maximization of pleasure, minimization of pain, levels of happiness, etc.; see Driver, 2022). Thus, it is not clear what characterization of utility or method of measurement would be most appropriate to use.

Moreover, consequentialism and utilitarianism are not typically viewed as decision procedures. Philosophers have often distinguished between theories that provide standards of rightness (i.e., the properties through which objects can be judged as right or justified) and theories that offer decision procedures (i.e., criteria that dictate how people should deliberate, reason, and act; see Bales, 1971; Kymlicka, 2002; Railton, 1984). Because utilitarian reasoning would often dictate that individuals make decisions that maximize total welfare for all rather than decisions that protect their own interests, many philosophers have argued that even if consequences are the things that make actions right or wrong, utilitarianism and consequentialism cannot be viewed as universal decision procedures (Brink, 1986; Stark, 1997). For this reason, it would likely be inappropriate to view utility calculations as action-guiding for individuals and policymakers.

Finally, even if consequentialism was action-guiding, use of this model would often risk focusing on the making of a specific, discrete choice as the main criteria for judging ethics (i.e., the choice that maximized utility the most). This is of concern, as it necessarily assumes that one of the choices is normatively superior to the other. While some decisions may have a normatively superior option, for many decisions, this is unlikely to be the case. For example, it has been demonstrated that pre-marking either a comfort or life-extending option on a form can influence real people's decisions about their end-of-life care (S. D. Halpern et al., 2013); however, there is little reason to believe as an objective truth about the world that either option is normatively superior to the other. Some people value their comfort over living longer and vice versa. Thus, in absence of a basic, intrinsic value that is shared by most decision makers, it would likely be inappropriate to base ethical comparisons on how well the intervention influenced individuals to make a specific choice.

In contrast to consequentialism, deontology is an ethical theory that defines the rightness of an act by the extent to which it conforms with the correct set moral norms (Alexander &

Moore, 2021). Rather than judging rightness by consequences or outcomes, deontology holds that the feature of an action that makes it right is whether or not it falls in accordance with the correct set of moral rules or duties, some of which may not be utility maximizing (R. Johnson & Cureton, 2022). For example, one might be viewed as having a duty to be honest with their spouse even if that honesty causes the spouse pain. Again, this theory seemingly could have direct applications in decision ethics (i.e., if there were universal duties that needed to be adhered to in any given circumstance, a decision intervention could be judged based on the extent to which it upheld those duties). However, there are also several problematic limitations to this theory. First, some have argued that because the moral duties one could select as universal are more or less arbitrary (i.e., there is often no unifying rationale for what duties are considered good), deontological ethics may not be action-guiding either (see Joseph, 1933; McNaughton, 1996; Raphael, 1994).

Moreover, even if deontological principles were action-guiding, several significant problems would still exist. For example, we would need to answer the question of who decides which principles matter for decision ethics. People across many religions, political parties, and backgrounds have different conceptions of what duties are most important to uphold (e.g., Graham et al., 2009), so the potential arbitrariness of what could be considered right would pose significant problems for defining a set of criteria for ethical interactions. Furthermore, even if we agreed on who should decide, according to the philosophical personality argument (PPA; see Feltz & Cokely, 2012), we still might have cause to doubt those principles that are selected. The PPA argues, based on a growing body of empirical literature, that basic philosophical intuitions are often affected by diverse factors that are irrelevant to the task (e.g., extraversion affects people's judgements about whether determinism and free will are compatible; Feltz & Cokely, 2013, 2016; Schulz et al., 2011; Stich & Tobia, 2015). Thus, any selection of deontological principles that are assumed to be *the* correct set of rules would likely be suspect (and perhaps be inappropriate) for ethical comparisons in this domain.

Finally, virtue ethics is a third ethical theory, which defines the rightness or wrongness of actions based on motives and character traits. According to this theory, actions are good if they exemplify virtuous traits. For example, if an agent has good motives for an action, virtue ethics would typically view that action as “right,” because the good motives reflect the virtue of the individual (Slote, 1995). Virtue ethics has sometimes been thought to focus on developing good habits and good character in people so that they can make actions that are true to themselves (e.g., McCracken et al., 1998; Shanahan & Hyman, 2003; Williams & Murphy, 1990). On this conceptualization, virtue ethics could hold some promise for evaluating decision interventions (i.e., if an intervention is able to help individuals develop some skill or habit that allows them to make decisions that are true to themselves, and if it does so to a greater extent than another intervention, all things being equal, that intervention could be considered more ethical).

However, there are limitations to this method that would likely render virtue ethics unable to be a standalone guide to ethical interactions. For example, while the notion above (i.e., good habits and character traits that help people take actions that are true to themselves) may serve as a useful guide, virtue ethics itself does not necessarily define which virtues, motives, habits, etc. matter the most. There are many virtues that could be considered (e.g., honesty, unselfishness, objectivity, realism; see Hursthouse & Pettigrove, 2022), some of which may not matter for comparing intervention ethics. Thus, as with previous theories, virtue ethics by itself is unlikely to provide *the* correct set of virtues and character traits that should be evaluated. Moreover, once the virtues and traits needing to be developed are decided upon, on my conceptualization, it would likely not be enough merely for the intervention designer to *intend* for the intervention to support that development. Rather, it would need to be empirically demonstrated that the intervention actually *does* support the development of important habits and character. Accordingly, a standalone appeal to virtue ethics (e.g., determining whether the intervention designer had good motives that were relevant to decision-makers' character development when they created the intervention) would likely be insufficient to make ethical comparisons between interventions. However, in this dissertation, I will propose the beginnings of a framework and demonstrate a method that could be used to overcome these limitations.

Absent an overarching, standalone philosophical theory to guide ethical interactions, behavioral economists and psychologists have sometimes discussed and attempted to use other independent forms of ethical criteria. For example, principles that have received some discussion in the past are welfare (i.e., whether the intervention helped the decision-maker choose something that made their life "longer, healthier, [or] better;" Thaler & Sunstein, 2008, p. 5), freedom of choice (i.e., whether the intervention formally left open the option to choose among various alternatives; see Sunstein, 2015a), and judged acceptability (i.e., whether decision-makers, on average, viewed the interventions as acceptable to use; Felsen et al., 2013; Hagman et al., 2015; Yan & Yates, 2019). While all of these criteria are important, they represent only small parts of the larger picture of ethics (e.g., welfare likely represents a small part of the "net good" as characterized in consequentialist theories) and are likely not sufficient to guarantee an intervention is ethical. For example, if someone were to give a decision-maker a choice between two options but also threaten violence against the individual if they choose one of the options over the other, most people would likely feel that ethical rules were still broken, notwithstanding the fact that freedom of choice technically exists in this context. Likewise, judged acceptability could be subject to the argument of the PPA, meaning that judged acceptability may be affected by diverse, irrelevant factors to the decision task.

Accordingly, a more comprehensive set of criteria would likely be beneficial in order to accurately compare interventions on ethical merit. In order to avoid conflicts arising from the PPA (i.e., the view that we are unlikely to come to know the mind-independent, non-conceptual truth of many philosophical issues—including ethics), one way to select these criteria might be to look to principles that *most* people generally value. One promising set of criteria that could begin to form this framework for a larger theory of ethical interactions can be found in the APA's code of ethical conduct. Specifically, the APA has specified five aspirational principles that are intended to guide psychologists to uphold the highest ideals of the profession (American Psychological Association, 2017). Of note, these principles are accepted by a wide variety of professional organizations and businesses (Holt, 2023). These principles are:

1. Beneficence and nonmaleficence
2. Fidelity and responsibility
3. Integrity
4. Justice
5. Respect for people's rights and dignity

In brief, these principles represent some commonly held values that have been empirically demonstrated to have wide acceptance (e.g., Schwartz, 2012; Schwartz et al., 2012). Beneficence and nonmaleficence refers to an obligation to maximize benefits while minimizing harm to individuals. Fidelity and responsibility encompass the need to fulfill responsibilities by acting in the best interests of the individual being served. Integrity involves a duty to foster truthful and honest relationships. Justice means ensuring that individuals are treated fairly and equitably. Respect for people's rights and dignity embraces the rights of individuals to govern their own lives. In this context, we might think of an ethical decision intervention as one that not only leaves open the option to choose, but also which demonstrably minimizes the risk to these ideals to fullest extent possible (e.g., by promoting autonomy, building trust between the choice architect and the decision-maker, etc.).

To be clear, these principles do not necessarily represent a necessary set of conditions for ethical interactions. In fact, given the arguments presented above, it is likely that we could never know if they constitute *the only* correct set of criteria. However, given their importance in the decision sciences and their acceptance by people generally (J. A. Lee et al., 2011; S. H. Schwartz & Bardi, 2001), I propose that by measuring the extent to which various decision interventions respect these principles, one could reasonably approximate the extent to which those interventions have upheld the ethical standards of the field. Thus, they likely serve as a suitable set of criteria for defining ethical interactions in this domain. In this paper, I will demonstrate how one could go about empirically comparing interventions on these ethical principles. Specifically, I will focus on how one could compare interventions on the extent to which they respect reliable deliberation—a condition that has been argued to be central to respecting one's right to autonomy.

Autonomy

While there are many rights that are important to protect, one of the rights that is likely most widely agreed upon is autonomy (J. A. Lee et al., 2011; S. H. Schwartz, 2012; S. H. Schwartz & Bardi, 2001). Respect for autonomy is mentioned in numerous ethics codes (e.g., American Counseling Association, 2014; American Nurses Association, 2015; American Occupational Therapy Association, 2020). Autonomy has been argued to have intrinsic value (i.e., it is often thought that autonomy is valuable in and of itself, because individuals often view it as important that they be the authors of their own lives; see Benn, 1975; Dworkin, 1988). It is also thought to have instrumental value (i.e., it is thought to bring about other good things, such as positive affect, self-acceptance, improved health outcomes, etc.; see Vedam et al., 2017; Weinstein et al., 2012; Wichmann, 2011). Thus, taking into account its value across a wide variety of professional and personal settings, respecting or promoting autonomy while implementing decision interventions would likely be valuable.

In order to understand how one could measure the extent to which an intervention respects an individual's autonomy, it is first important to define autonomy. Unfortunately, there is no widespread consensus on what autonomy is. At the very base level, most conceptions of autonomy seem to refer to the ability of individuals to make self-governed or self-determined decisions according to their own coherent set of beliefs and values (Benn, 1975; Buchanan & Brock, 1989; Dworkin, 1981, 1988; Mele, 2001). However, within that definition, there are still many ways in which autonomy could be characterized. For example, Feinberg (1986) identifies at least four connotations that could be associated with the idea of self-governance: (1) the *capacity* to govern oneself, (2) the *actual experienced condition* of governing oneself, (3) the *ideal character* that emerges when governing oneself, and (4) the *right* to govern oneself. For this research, I define autonomy using the notion set forth by Mele (2001), who argues that autonomy is the actual condition of agents governing themselves. Mele further argues that one way to ensure one has acted autonomously is if:

1. The agent has no compelled motivational states, nor any coercively produced motivational states.

2. The agent's beliefs are conducive to informed deliberation about all matters that concern him.

3. The agent is a reliable deliberator. (p. 187)

In other words, one way for an individual to be autonomous is to be free from coercion (Condition 1), to have access to whatever correct information is required to deliberate accurately (Condition 2), and to be able—whether through executive skills, or through characteristics of the choice environment—to meaningfully incorporate relevant values, beliefs, emotions, etc. into their overall decision (Condition 3). Taking this definition into account, it could be argued that by demonstrating the extent to which an intervention respects these three conditions, one could sufficiently show the relative ability of an intervention to actually protect (and perhaps promote) autonomy.

While many attempts have been made to measure autonomy in the past, to my knowledge, no measure of autonomy currently exists that is suitable for this purpose. Some measures of autonomy measure trait levels of autonomy (i.e., measures of one's *general* disposition and preferences regarding autonomy; e.g., Anderson et al., 1994; Weinstein et al., 2012). These measures are likely insufficient for two reasons. First, traits have typically been found to be relatively stable over time (e.g., Anusic & Schimmack, 2016; Cobb-Clark & Schurer, 2012; Gustavsson et al., 1997). Thus, at least in the short-term, we would likely not expect autonomy measured using a dispositional scale to change in response to one-time interventions. Second, even if a trait measurement were to change over time, a dispositional account of autonomy would likely not fit with the current definition of autonomy (i.e., a trait measure would seemingly be more relevant to a characterization of autonomy as the *capacity* to govern oneself generally—i.e., Feinberg's first listed connotation—rather than the extent to which they *actually did* govern themselves in a given situation).

Some measures of autonomy have been developed which seemingly do focus on the actual experience of autonomy (e.g., Broeck et al., 2010; Upadhyay et al., 2014; Vedam et al., 2017), but these scales are likely problematic as well. None of the scales seems to measure all three conditions for autonomy cited previously. Most focus only on the coercion condition of autonomy. For example, scales developed by Broeck et al. (2010) and Upadhyay et al. (2014) only measured the extent to which workers felt coerced to do tasks a specific way and the extent to which women felt their partners forced them to feel and act in certain ways concerning pregnancy/family planning, respectively. Vedam et al. (2017) measured the extent to which women felt their doctors provided them adequate information and space to make their own maternity care options (i.e., the coercion and informed deliberation conditions for autonomy) but did not explore the reliable deliberation condition. Of course, there could be other sets of sufficient conditions for autonomy. However, given that Mele's conditions for autonomy are a collective set of sufficient conditions, all three conditions need to be established in order for an individual to behave autonomously on Mele's specific conceptualization. Moreover, these measures are all subjective measures of autonomy (i.e., the extent to which one *feels* they made an autonomous decision). While subjective appraisals are important and in some cases may be adequate measures of autonomy, subjective estimates of many variables are often imperfectly related to objective measures of the same constructs (e.g., Fagerlin et al., 2007; Mahmoud-Elhaj et al., 2020).

In light of these issues, to most effectively and accurately compare the extent to which an intervention promotes autonomy, I would argue that *objective* measures of each of the three sufficient conditions need to be developed. Previous work by Tanner (2021) has examined one of these conditions: informed deliberation. Using an objective measure of recycled water knowledge developed by Mahmoud-Elhaj et al. (2020), Tanner found that objective knowledge mediated the relationship between an educational intervention and recycled potable water acceptance while the same relationship did not hold for libertarian paternalistic default nudges. This suggests that some educational interventions are better able to respect the condition of informed deliberation than some default nudges when making decisions about recycled water. However, the other conditions (i.e., being free from coercion and reliable deliberation) represent important factors that have yet to adequately be explored empirically. In this dissertation, I will attempt to empirically explore one of those factors: reliable deliberation.

Measuring Reliable Deliberation

To understand how one could objectively measure reliable deliberation, it is important to first define reliable deliberation. Similar to autonomy, there is no widespread consensus on what it means to be a reliable deliberator. Reliable deliberation has sometimes been characterized in terms of competency (i.e., whether the individual possesses the executive qualities needed to properly deliberate about a topic). For example, some have proposed that in order to be capable of proper deliberation, individuals need to be of full age, sound mind, and literate (Schramme, 2021; see also Harvey, 2001). However, defining reliable deliberation in terms of competence or executive skills would likely be problematic. The selection of executive skills that would qualify one as a competent deliberator would likely be suspect due to the PPA. Likewise, past research has shown that the lack of a skill does not preclude the possibility of individuals acting as if they had that skill. For example, Garcia-Retamero and Galesic (2010) showed that effective visual aids were able to help individuals with low numerical skills understand risk at the same level as those who were already high in that skill. Thus, an effective decision intervention could likely help individuals to be competent, reliable deliberators in at least some domains.

Others have defined reliable deliberation using characteristics and outcomes of the deliberation itself. This characterization of deliberation is based around the argument that there are some ways of deliberating that are better than others and typically assumes that proper deliberation is a process that follows logical rules and arrives at rational conclusions (e.g., decisions that maximize utility; see Arpaly & Schroeder, 2012; Danielson, 1998). This conceptualization is also likely to be problematic, given that it likely requires there to be both a normatively superior way to arrive at a decision and a normatively superior option to choose. As previously discussed, there are likely many cases where a normatively superior option does not exist. Moreover, even in cases where there are normative standards, there are likely many processes through which one can arrive at them. For example, Cokely & Kelley (2009) found that when making decisions about hypothetical lotteries, many high-ability participants arrived at decisions that were normatively correct according to expected value calculations; however, many did so without actually making those calculations, and instead arrived at their choice using simple yet elaborative heuristic processes. Thus, defining reliable deliberation by the extent to which individuals use logical rules or arrive at normatively superior choices would likely not be adequate.

I propose that one way we could instead define reliable deliberation is to break the term down into its component parts (i.e., reliable and deliberation). To define deliberation, one might again look to Mele (2001; also see Baron, 2005; Blacksher et al., 2012; Dietz, 2013; Elwyn & Miron-Shatz, 2010; Mansbridge, 2015). Mele describes deliberative processes as methods where (1) one has some psychological basis for evaluative reasoning (e.g., values, desires, and beliefs);

(2) an evaluative judgment is made on the basis of that reasoning which recommends a particular course of action; (3) an intention is formed in line with the evaluative judgement; and (4) an action is carried out to execute the intention. In other words, one might describe deliberation as the process of taking one's own values, desires, beliefs, etc., and translating them into appropriate intentions and actions.

For this research, I define "reliable" in the same sense that reliability is often referred to in statistics. Reliability typically refers to how consistent results are when using the same or similar measures. For example, one introductory statistics textbook refers to a reliable measurement as "one that will give you or anyone else approximately the same result time after time when taken on the same object or individual" (Utts, 2014, p. 55). Combining these two terms, we might define reliable deliberation as the condition of making consistent judgements and decisions when deliberating about the same set of values and beliefs. That is, in instances where an individual's values, desires, and beliefs about a topic do not change (all else being equal), we would expect them to translate through into relatively stable decisions and behavioral intentions across similar, short-term decision-making trials. Moreover, we might expect this consistency even when one is exposed to additional random information that should be irrelevant to one's understanding of a topic.

For example, applied to recycled water, an ideal decision-making intervention might help one take their values and beliefs relevant to recycled water (e.g., values related to health, environmental responsibilities, etc.) into account, and lead to an intention or decision consistent with those values (e.g., because research shows that recycled water can be as healthy as normal tap water and can reduce waste discharged back into the environment, they might see that recycled water is consistent with their values and report intentions to vote to implement recycled water). Then, unless something happens to change one's relevant values or beliefs, we would expect these individuals to make similar decisions and report similar intentions consistently across many trials, even when exposed to potentially biasing, but ultimately irrelevant information (e.g., information that someone seeking to implement recycled water is a member of an opposing political party). Theoretically, this information should hold little relevance to one's representative understanding of recycled water. For example, in line with some factors identified in Cokely et al.'s (2018) framework for skilled decisions, the political identity of one advocating for recycled water should have little to no effect on the extent to which one comprehends relevant facts about recycled water, nor should it affect most affective responses to recycled water (e.g., we would not necessarily expect it to predict the level of disgust one feels about recycled water). Thus, if one has reliably deliberated about recycled water, we would likely expect very little change in their acceptance even after they have received this extra information.

Notably, similar definitions of reliable deliberation have been presented before. For example, Gastil and Dillard (1999) and Rosenberg (2022) argue that one likely outcome of proper deliberation should be increased coherence of thinking on an issue. Likewise, Mele (2001) asserts that deliberative practices that are based on common decision biases like the anchoring effect may be unreliable. However, to my knowledge, no studies have attempted to *measure* reliable deliberation based on this characterization. Studies on debiasing techniques have sometimes made related measurements; however, there are at least two key differences that would make debiasing measurements inadequate for the purposes of reliable deliberation. First, in many cases, the debiasing literature seems to center around topics and decisions where there is a normatively correct option. Thus, measurements related to debiasing are typically based on the making of accurate, normatively correct decisions (e.g., Jenkins & Youngstrom, 2016; Lambe et al., 2016; Sellier et al., 2019). As discussed previously, such a measurement is likely not appropriate for my characterization of reliable deliberation, especially given that my characterization focuses on the *consistency* of one's choices rather than the choice itself. Second, most debiasing literature seems to focus on either one-off decisions or the same decision made across a small number of longer-term timepoints (e.g., Dai et al., 2021; Hinds, 1999; Morewedge et al., 2015). Beyond the aforementioned limitation (i.e., such tasks usually still focus on the decision itself as the main outcomes of interest), my characterization of reliable deliberation will use iterative decision tasks across a short period of time, where additional pieces of irrelevant, biasing information will be added for each successive trial, and consistency within those trials will be analyzed. To my knowledge, no studies of this kind currently exist in the debiasing literature, so this method likely represents a unique contribution to the field.

Interventions and Hypotheses

Using the previously discussed operationalization (i.e., consistency across trials, even when presented with irrelevant, but biasing information), in this dissertation, I will seek to compare two kinds of decision interventions on the extent to which they support reliable deliberation: educational interventions and libertarian paternalistic nudges. These kinds of interventions have both been used extensively in past behavior change research (e.g., Jachimowicz et al., 2019; Zelezny, 1999). However, due to differences in their proposed mechanisms for behavior change, it is likely that these interventions might differ with respect to their ability to promote autonomy.

In particular, educational interventions represent a class of behavioral interventions that typically seek to promote change by increasing one's domain-specific knowledge of a topic. Studies across many scientific disciplines—including the domain of recycled potable water—

suggest that low support for scientific innovations is often at least partially attributable to misunderstanding or ignorance of key domain-specific facts (Allum et al., 2008; Glick et al., 2019; Nisbet, 2005; Nisbet & Goidel, 2007; Sturgis & Allum, 2004). Cokely et al. (2018) have further suggested that the effects of education and general decision decision-making skills on choices often occur by increasing representative understanding. Representative understanding can be defined as an understanding of a topic that is nuanced enough that additional information received on that topic is unlikely to change tendencies made on the basis of that understanding. The skilled decision framework posits that representative understanding is achieved by having enough high-quality information about to topic to prompt deliberation, which leads to well-calibrated levels of confidence in one's knowledge about a topic and helps individuals to calibrate their own affective reactions to the topic (see Figure 1). The integration of knowledge and appropriate affective reactions is then thought to drive—at least in part—the decision one eventually makes, and lead to a high-quality choice (e.g., one that is in line with expert consensus on a topic).

Past research has found some educational interventions to be effective at increasing both simple binary choices to accept recycled potable water, as well as a complex of factors related to recycled water acceptance (e.g., related behavioral intentions, lower reported worry about recycled water, etc.). Likewise, research has also found that decisions about recycled potable water acceptance likely proceed along a pathway consistent with the framework for skilled decisions, such that effective educational videos increase knowledge, which has both a direct effect on recycled water acceptance, as well as a strong indirect effect through reduced worry about health risks associated with recycled water (Tanner, 2021; Tanner & Feltz, 2022). Taking these factors into account, I hypothesize:

H_1: Individuals who are educated about recycled water will show higher acceptance of recycled water across a series of similar, but biasing, decision trials than those who are not educated about recycled water.

Meanwhile, libertarian paternalistic nudges are a class of intervention that typically seeks to change behavior by targeting automatic decision processes (i.e., thinking that is quick, reflexive, and is typically heavily influenced by features of the environment; see Evans, 2008; Thaler & Sunstein, 2008). In particular, I define a libertarian paternalistic intervention as one that seeks to predictably influence behavior while still leaving open the ability to choose differently (i.e., libertarian), but that also involves the violation of some moral rule, such as withholding relevant information (i.e., paternalistic; see Gert & Culver, 1979; Tanner, 2021). For example, one type of intervention thought to fall in this category is default nudging. Default nudges are typically designed to target automatic decision processes through the setting of default rules (i.e., making clear the option that participants would receive if they were to do nothing else). Past research has found that because of factors such as decision inertia (i.e., the tendency to choose the default because it would take more effort to actively switch to another option) and social sensemaking (i.e., the tendency of decision-makers to see defaults as tacit recommendations by the intervention designer), people will often stay with default options, even when they are free to choose differently (Blumenthal-Barby & Burroughs, 2012; Ghesla et al., 2019; E. J. Johnson & Goldstein, 2013; Krijnen et al., 2017; Madrian & Shea, 2001; Thaler & Sunstein, 2008; Venema et al., 2018).

For example, Egebark and Ekström (2016) found that by changing the default printing settings at a large public university to double-sided, the sheets of paper used per day at the university dropped by an average of 15 percent, relative to the previous five weeks where the default setting was single-sided. Similar effects have been observed across multiple decision domains (e.g., saving for retirement, consenting to be an organ donor, end-of-life decisions, and various consumer domains; Feltz, 2015, 2016; Jachimowicz et al., 2019; Johnson & Goldstein, 2013; Madrian & Shea, 2001), and default options have even been shown to affect choices to accept recycled water (Tanner, 2021; Tanner & Feltz, 2022). Thus, I hypothesize:

H_2: Individuals who are nudged into recycled water use (e.g., using default rules) will show higher acceptance of recycled water than those who are not defaulted into recycled water use.

However, unlike educational interventions, research has suggested that libertarian paternalistic nudges typically do not affect any of the other factors related to recycled water acceptance (e.g., similar behavioral intentions, worry about risks, etc.), and they likely do not engage representative understanding pathways (Tanner, 2021). Because they are unlikely to increase representative understanding, I hypothesize:

H_3: Those exposed to an educational intervention will exhibit higher levels of reliable deliberation (operationalized through more consistency in choice) across a series of similar, but biasing, decision trials than those who only receive a series of libertarian paternalistic nudges.

I tested these hypotheses in a series of four experiments. In Experiments 1a and 1b, I developed and tested a number of nudge scenarios designed to bias participants towards accepting or rejecting recycled water and found that a number of common biases can be used to consistently sway opinions on recycled water. In Experiment 2, I conducted an experiment comparing participants who received an educational intervention prior to receiving the best set of nudge scenarios from Experiments 1a and 1b to those who only received the set of nudge scenarios. I found that those who received an educational intervention displayed more consistent choices across the series of nudge scenarios than those who only received the nudges. In Experiment 3, I replicated these results using a larger sample and found that for many people, the consistency observed for the educational intervention likely proceeded along a pathway that was consistent with the model for skilled decisions.

Experiment 1a

The purpose of Experiment 1a was to develop a set of scenarios that could be used to bias participants to accept or reject recycled water to be used in later experiments.

Method

Participants. Participants were recruited from Amazon's Mechanical Turk (MTurk), in exchange for $0.75 USD. A filter was applied through CloudResearch's MTurk Toolkit that restricted the MTurk subject pool to only those who have been vetted and approved by CloudResearch. CloudResearch-approved participants have been found to provide higher quality data than many other online participant recruitment platforms (Hauser et al., 2022; Peer et al., 2022). To determine the number of participants needed for this experiment, an a priori power analysis was conducted based on effect sizes for recycled potable water acceptance observed in previous studies. The power analysis suggested a need for about 100 participants in each bias condition (*power* = .8, *alpha* = .05, *d* = .40), resulting in an overall need for about 800 participants.

To account for participants who would likely need to be excluded from analyses, a total of 838 participants were recruited to complete the experiment. However, a total of 86 participants were excluded for failing to complete the survey, and an additional nine participants were excluded for answering one or more attention check questions incorrectly. Among the final sample of 743 participants, 404 identified as female, 328 identified as male, and 11 declined to specify their gender. The ages of participants ranged from 19 to 84 (M = 43.04, SD = 13.51).

Procedure. Prior to the start of the experiment, I developed a set of unique scenarios meant to bias participants towards choosing to accept or reject recycled water. These scenarios were created based on a list of more than 100 cognitive biases provided by The Decision Lab (https://thedecisionlab.com/biases). I examined each bias on the list, evaluated each one on its ability to be incorporated into a scenario about recycled water, and identified a total of 25 candidates[1]. For each candidate, I then developed a set of two logically equivalent scenarios: one scenario meant to nudge the participant towards preferring recycled water and one meant to nudge the participant towards preferring "conventional" water.

For example, one bias identified is known as the reactive devaluation bias, which holds that people tend to favor proposals made by those they view as similar to themselves and devalue proposals made by those they view as opposite, regardless of the content of the proposal (e.g. Maoz et al., 2002). Accordingly, I developed a set of scenarios where participants were asked to imagine they attended a city council meeting where the mayor proposed a recycled water scheme. In the scenario meant to bias participants toward preferring recycled water, the mayor was described as being "[a] close personal friend, shares many of the same political views…and who is known to be well-respected among much of the community." Meanwhile, in the scenario meant to bias participants toward preferring conventional water, the mayor was described as

[1] Note, because these scenarios attempt to influence people to express a desired preference (while still leaving open the formal option to choose otherwise) and were built using cognitive biases (and thus are unlikely to proceed through rational agency, representative knowledge pathways; see Tanner, 2021), I consider each of these scenarios to likely represent a libertarian paternalistic nudge. Accordingly, hereafter, I sometimes use the terms "bias scenarios" and "nudges" interchangeably.

being "[someone] you have never met, holds many of the opposite political views…and who is not known to be well-respected among much of the community." See Table 1 for a list of all biases for which scenarios were developed.

Prior to testing, three of the biases (bottom dollar effect, IKEA effect, and mere-exposure effect were eliminated due to the hypothetical nature of the scenarios [i.e., I viewed it as unlikely that these biases would produce reliable effects in a solely hypothetical environment]). To test the remaining pool of 22 biases (44 scenarios), participants viewed six randomly selected biases from the candidate list. Within each of the six randomly selected biases, participants were further randomly assigned to view *either* the conventional bias scenario *or* the recycled bias scenario. Thus, participants could not be assigned to both conditions of the same nudge, but they could be assigned to view a mix of conventional and recycled bias scenarios throughout the experiment (see Figure 2).

For each scenario, responses were on a 6-point Likert scale (1 = *would definitely prefer conventional water*, 6 = *would definitely prefer recycled water*). Additionally, after each scenario, participants responded to a battery of questions developed by Tanner (2021) aimed at measuring several additional indicators of recycled water. Specifically, these questions measured the extent to which participants felt they would be satisfied with policymakers, their relevant affective reactions (worry), the extent to which they would endorse engaging in related behaviors in protest to recycled water, and the extent to which they would make negative recommendations to others if recycled water were to be implemented in their city:

1) I would be upset with [my local government] if they asked people to use recycled water for drinking.

2) I would trust [my local government] if they decided to ask people to use recycled water for drinking.

3) I would be worried about my health if [my local government] decided to ask people to use recycled water for drinking.

4) I would likely protest if [my local government] decided to ask people to use recycled water for drinking.

5) I would consider moving if [my local government] decided to ask people to use recycled water for drinking.

6) I would tell my friends and family not to move here if [my local government] decided to ask people to use recycled water for drinking.

Language for these questions was adjusted to be consistent with each individual scenario (e.g., for the reactive devaluation bias "my local government" was changed to "the mayor"). Responses options were on a 6-point Likert scale (*1 = strongly disagree, 6 = strongly agree*). These indicators were included in order to serve as further indicators of effectiveness. Finally, participants responded to basic demographic questions (e.g., age, gender).

Results and Discussion

I conducted an independent samples t-test for each individual bias to see whether condition (recycled water bias vs. conventional water bias) affected average preference for water type. In total, 15 of the biases displayed significant effects on water preference (see Table 1), supporting my second hypothesis that many nudges might be able to increase recycled water acceptance. Effects for the conventionally significant biases varied from conventionally small sizes (d = .29) to large sizes (d = .96).

In order to further reduce the number of candidates for later experiments, I also examined the effects of bias condition on each of the additional indicators. Three of the fifteen significant biases (omission bias, hyperbolic discounting, and the bandwagon effect) displayed no reliably significant effects on any of the additional indicators and thus were eliminated from use in later experiments (see Table 1).

Experiment 1b

The purpose of Experiment 1b was to narrow down the conventionally significant biases from Experiment 1a into a smaller set of items for use in later experiments.

Method

Participants. Participants for Experiment 1b were recruited from MTurk, again using the CloudResearch filter, in exchange for $0.75 USD. To be consistent with the power analysis reported in Experiment 1a, here, I aimed to recruit about 400 participants total (100 participants for each the four between-subjects conditions used within this experiment). In total, responses were collected from 422 participants; however, 18 participants were excluded from analyses for incorrect answers on attention check questions, and an additional 18 participants were excluded for failing to complete the survey. This resulted in a final sample of 386 participants. Ages ranged from 20 to 84 (M = 43.94, SD = 13.74). Two hundred thirty-seven participants identified as female, 145 identified as male, two were non-binary, and two declined to specify their gender.

Procedure. Prior to the start of the experiment, I split the 12 remaining bias scenarios into two unique sequences of biases. This was done in an effort to reduce the amount of time participants would need to complete the experiment. The sequence and order each bias was assigned alternated based on effect sizes observed in Experiment 1a (e.g., the bias with the largest observed effect size—the noble edge effect, d = .96—was assigned to be the first bias participants viewing Sequence 1 would see; the next largest effect—the ambiguity effect, *d* = .91—was assigned to be the first bias participants viewing Sequence 2 would see; the third largest effect (action bias, *d* = .80) was assigned to be the second bias participants viewing Sequence 1 would see, etc.). This resulted in two sequences of six biases, sorted in descending order by effect size.

For Experiment 1b, participants were randomly assigned to one of four conditions, where they viewed Sequence 1 or 2 as previously described, or one of the two sequences with bias positions reversed (i.e., Sequence 1R or 2R, where biases were displayed in ascending order based on effect size). Sequences 1R and 2R were included in an effort to guard against potential carryover effects. It is possible that viewing the bias with the largest effect size first could have a strong enough influence to override effects of all subsequent biases, thus examining the sequences in reverse might provide a better idea of what set of biases would be most appropriate for shifting preferences regarding recycled water.

For each scenario presented, participants responded using a 6-point Likert scale (1 = *would definitely prefer conventional water*, 6 = *would definitely prefer recycled water*). The sequences were programmed such that the order in which the biases were received was fixed; however, the version of each scenario that participants received (i.e., the version meant to bias towards recycled water vs. the version meant to bias towards conventional water) was adaptive, based on their most recent response. Specifically, participants always received the version of the scenario meant to bias them towards the opposite option of their previously stated preference. For example, if someone in Sequence 1 received the noble-edge scenario and answered with a three (would slightly prefer conventional water), for the next bias (action bias), they would receive the version meant to bias them towards preferring recycled water.

After participants responded to each of the scenarios in their assigned sequence, all participants responded to a scale gauging objective knowledge of recycled water (Mahmoud-Elhaj et al., 2020). This scale contains 39 *true/false/I don't know* questions. Responses were coded for correctness (*I don't know* answers were coded as incorrect) and summed to form a composite score of recycled water knowledge. Finally, participants responded to basic demographic questions.

Results and Discussion

Analytic strategy. The main outcome of interest for this experiment was the consistency of participants' preferences regarding recycled water. I used two methods to gauge consistency. First, I calculated a standard deviation for each participant across their responses to the six bias scenarios they received. Standard deviations are a measure of how spread out numbers are from the mean in a set of data (Utts, 2014). Thus, if someone were to give the same response across all six scenarios (e.g., if they responded with 4 [*slightly prefer recycled water*] for each question), we would expect a standard deviation of zero, because each of their responses would be identical to their mean of four. Conversely, if someone were to alternate between responses of 1 (*strongly prefer conventional water*) and 6 (*strongly prefer recycled water*) on every scenario, we would expect a higher standard deviation, as each of their responses would be the maximum distance away from their own mean of three. On this logic, because a standard deviation was calculated for each individual across their responses to the six scenarios, smaller standard deviations should indicate less spread in each individual's responses and thus higher consistency in their own choices.

For the second method, I operationalized consistency by the number of qualitative switches participants made between preferences for recycled and conventional water throughout the experiment (i.e., the number of times participants changed from a 1-3 rating to a 4-6 rating—or vice versa—in each successive scenario). This operationalization was included because of potential limitations that could be associated with using standard deviations. Especially when using small sample sizes, standard deviations can be unstable and highly influenced by outliers (e.g., Utts, 2014). Given that standard deviations in this experiment were calculated using only six indicators, it could easily be the case that someone could display high consistency in choice but still have a higher standard deviation than an inconsistent decision-maker if any of the choices were outliers (e.g., an individual who responded 1, 1, 1, 1, 1, 1, 6 across seven choices would have a standard deviation ($SD = 1.89$) that is higher than an individual who responded 2, 5, 2, 5, 2, 5, 2 ($SD = 1.60$), notwithstanding the fact that the latter individual consistently changed their preference while the former did not).

Consistency results: standard deviations. Using a one-way ANOVA, I examined average differences in standard deviations across the four sequences. The omnibus test suggested a marginally significant effect (F (3, 382) = 2.1, $p = .10$, $\eta^2 = .02$). Post-hoc tests conducted using Tukey's Honestly Significant Difference (HSD) tests (a robust test for making comparisons among multiple groups; see Midway et al., 2020) suggested a marginally significant difference in variance between Sequences 1 and 2 (t (192) = 2.35, $p = .09$, $d = .33$), where participants in Sequence 2 ($M_{SD} = 1.21$, $SD_{SD} = .59$) were more consistent (i.e., had lower average standard deviations) than those in Sequence 1 ($M_{SD} = 1.40$, $SD_{SD} = .57$). Notably, variation in Sequence 2R was very similar, though the comparison between Sequences 1 and 2R did not reach conventional levels of significance ($M_{SD} = 1.25$, $SD_{SD} = .54$; t (204) = 1.90, $p = .23$, $d = .27$; see Figure 3).

To further probe any differences between the sequences, I split participant knowledge scores into quartiles and conducted a series of one-way ANOVAs to see if consistency differed as a function of a participant's native knowledge in each individual sequence. The logic behind these tests was that if average standard deviations differed as a function of native knowledge, there would be some evidence that the order of scenarios was not so overpowering that it would drown out all other effects (i.e., because it would still be sensitive to individual differences such as knowledge). There were no reliably significant differences in consistency among knowledge groups for the sequences where larger effect biases were displayed first (Sequence 1: F (3, 101) = 1.28, p = .29, η^2 = .04; Sequence 2: F (3, 85) = 1.04, p = .38, η^2 = .04), suggesting that starting with the most effective biases might have overpowered the other biases. However, both reverse-ordered sequences displayed significant or marginally significant effects (Sequence 1R: F (3, 87) = 2.89, p = .04, η^2 = .09; Sequence 2R: F (3, 97) = 2.29, p = .08, η^2 = .07), indicating that one of these sequences would likely be most appropriate for shifting water preferences.

To differentiate between the two sequences, I conducted post-hoc tests using Tukey's HSD tests. Specifically, I examined the sequences to see which results aligned best with theory discussed in the introduction (i.e., that those who knew more about recycled water would likely exhibit more consistency in preferences). Patterns for Sequence 1R were opposite what would be expected given theories of representative understanding (e.g., there was a marginally significant difference where participants in the lowest knowledge quartile displayed *more* consistency in their choices (M_{SD} = 1.1, SD_{SD} = .71) than those who were in the highest knowledge quartile (M_{SD} = 1.5, SD_{SD} = .42; t (57) = 2.55, p = .06, d = .66; see Figure 4)). Meanwhile, patterns for Sequence 2R were more consistent with what would be expected (e.g., there was a marginally significant effect where participants in the second knowledge quartile displayed less consistency in their choices (M_{SD} = 1.40, SD_{SD} = .55) than those who were in the highest knowledge quartile (M_{SD} = 1.0, SD_{SD} = .45; t (40) = 2.37, p = .09, d = .78; see Figure 5)).

Consistency results: switching. Using a one-way ANOVA, I also examined the number of switches participants made across the six scenarios. The omnibus test was not reliably significant (F (3, 382) = 1.29, p = .28, η^2 = .01), indicating that there were no reliable differences in consistency among the four sequences.

I also examined the number of switches made as a function of knowledge quartile within each individual sequence. Similar to standard deviations, Sequence 1 had no reliable effects on switches (F (3, 101) = 1.29, p = .28, η^2 = .04). In both sequences 1R and 2, there were reliably significant or marginally significant differences in the number of switches made by participants in knowledge quartile (Sequence 1R: F (3, 87) = 2.50, p = .07, η^2 = .08; Sequence 2: F (3, 85) = 3.09, p = .03, η^2 = .10). However, both sequences displayed effects that were inconsistent with theory. For example, in Sequence 1R, those in the lowest knowledge quartile switched preferences less ($M_{Switches}$ = 2.52, $SD_{Switches}$ = 1.65) than those who were in the highest knowledge quartile ($M_{Switches}$ = 3.5, $SD_{Switches}$ = .42; t (57) = 2.55, p = .06, d = .70; see Figures 6 and 7). Meanwhile, in Sequence 2R, no reliably significant differences were observed (F (3, 97) = .49, p = .69, η^2 = .02).

Discussion. While results from this study were mixed, with respect to native knowledge, Sequences 1, 1R, and 2 either displayed no reliable differences or differences inconsistent with theory across the two operationalizations of choice consistency. Meanwhile, there was some indication that Sequence 2R might be the best candidate for ensuring that the bias scenarios would be able to shift participants' preferences regarding recycled water in a way that is consistent with theory (e.g., when comparing standard deviations among the knowledge quartiles). Thus, this sequence was selected for use in the final two experiments. See the Appendix for the wording and order of the retained items.

Experiment 2

The goal of Experiment 2 was to conduct a small-scale pilot test using the sequence of items retained after Experiment 1b. In addition, educational interventions were implemented to see if there would be a difference in consistency between the educational intervention and libertarian paternalistic nudges.

Method

Participants. Participants for Experiment 2 were recruited from the student participant pool at a large university in the United States. Participants completed the survey in exchange for partial course credit. Responses were collected from a total of 128 participants. Fourteen responses were excluded from analyses for failure to complete the survey, one response was excluded for not meeting the minimum age requirement, and an additional 31 responses were discarded for answering one or more attention check questions incorrectly. Of the 82 responses that were analyzed, 64 of the respondents identified as female, 16 identified as male, one was non-binary, and one declined to specify their gender. Ages ranged from 18 to 38 (*M* = 18.78, *SD* = 2.32). A post-hoc sensitivity analysis suggested that the current sample could reliably detect a medium sized effect (*power* = .80, *alpha* = .05, one-tailed; *d* = .55).

Procedure. At the start of survey, participants were first randomly assigned to either the *education condition* or the *control condition*. Participants in the education condition watched a 5-minute video produced in collaboration with water experts that highlights several key facts about the need for recycled water, the recycling process, areas already using recycled water, etc. (https://www.youtube.com/watch?v=eGcZjnwQy2w). Past research (Tanner, 2021; Tanner, McDonald, et al., 2023; Tanner & Feltz, 2022) has found this video to be very effective at increasing knowledge and acceptance of recycled water. Participants in the control condition watched a similar length video describing how the internet works (https://www.youtube.com/watch?v=7_LPdttKXPc).

After participants watched their assigned video, participants were further randomly assigned to a *default-in* or *default-out* condition. Participants in the *default-in* condition were presented with the following hypothetical scenario:

> Imagine you work for a small business. As part of your job, you are in charge of ordering water for the office water dispenser. In the past, your business has ordered from Company A, which is listed as the business's default provider of water. One day, Company A announces it will officially supply only recycled water to customers. You notify your boss of this change, and they tell you that you can continue ordering from Company A, or you can choose to switch providers and instead start ordering water from Company B. Assuming Company A (recycled water) and Company B (traditional water) offer the same prices for their water and you have no outside pressure to choose one option over the other, which company would you choose?

Participants in the *default-out* condition received the same hypothetical scenario, except Company B (the non-default provider) is the one selling recycled water. Participants were then asked to rate how likely they would be to stay with their default water company or switch to the opposite one on a 6-point Likert scale (1 = *definitely stay with the default and order from Company A*, 6 = *definitely switch and order from Company B*). Responses were reverse-coded in the default-in condition to ensure that the scales were consistent (i.e., 1 = *definite preference for conventional water*, 6 = *definite preference for recycled water*). Previous research (Tanner, Cokely, et al., 2023) has indicated that this particular default nudge oftentimes influences the choice to accept recycled water.

Next, participants completed the sequence of six biases retained from Experiment 1b. Similar to Experiment 1b, the order of the biases was fixed, however, the version of each bias the participants received depended on their response to the most recent scenario, such that those who reported preferring conventional water in the previous scenario received the version meant to bias one towards recycled water, and those who reported preferring recycled water in the previous scenario received the version meant to bias one towards conventional water. Responses were again on 6-point Likert scales, where 1-3 represented a preference for conventional water and 4-6 represented a preference for recycled water.

Additionally, after each scenario, participants responded to the same battery of additional acceptance indicator questions reported in Experiment 1a. After finishing the scenarios, participants completed the same objective knowledge of recycled water scale reported in Experiment 1b. Then, they were asked questions gauging their level of statistical numeracy. This was measured using the 7-item version of the Berlin Numeracy Test (Cokely et al., 2012; L. M. Schwartz et al., 1997). Statistical numeracy represents the ability to understand and apply basic probabilistic information, which has been shown to predict better understanding and general decision-making skill across many domains (e.g., Garcia-Retamero et al., 2019; Peters, 2012), including recycled water (e.g., Tanner, 2021 has found significant positive associations between numeracy and knowledge, as well as significant negative associations between numeracy and worry about recycled water). For these reasons, numeracy has often been thought to be an influential skill in helping individuals achieve representative understanding of a topic (e.g., Cho et al., 2021). Finally, participants provided basic demographic information.

Results and Discussion

Manipulation checks. To ensure the interventions worked as intended, I first checked to ensure that the educational intervention and default nudge increased acceptance of recycled water. Consistent with my first hypothesis, those in the education condition were more accepting (i.e., they exhibited higher preferences for recycled water) on average across all the scenarios combined (*M* = 3.95, *SD* = .75) than those who were in the control condition (*M* = 3.54, *SD* = .71; *t* (80) = 2.50, *p* = .01, *d* = .55; see Figure 8). There was no reliably significant difference in acceptance across all scenarios for the first nudge (i.e., default condition; (*t* (80) = 1.10, *p* = .28, *d* = .24)). However, consistent with my second hypothesis, many of the nudges increased acceptance within their own scenarios (see Table 2)[2]. For example, within the default scenario itself, being defaulted into buying from the recycled water provider resulted in significantly higher acceptance of recycled water (*M* = 4.53, *SD* = 1.47) than being defaulted out (*M* = 3.61, *SD* = 1.48; *t* (80) = 2.80, *p* < .01, *d* = .62; see Figure 9). These results suggest that both interventions likely successfully increased acceptance in some way.

---

[2] In some cases (across both Experiments 2 and 3), the nudges produced results that were opposite of expectations (i.e., significantly higher acceptance in the conventional bias condition than the recycled bias condition). I hypothesize that this might be an artifact of the methods used in this experiment. That is, because participants were always placed in the condition opposite of their most recently stated preference, if a participant consistently stayed high in acceptance, they would consistently be placed in the conventional water bias condition. Thus, for example, if the education intervention increased both acceptance and choice consistency, it would not be surprising to observe systematically higher acceptance in the conventional bias condition. It is also worth noting that when analyzing average acceptance across all seven scenarios (see Table 3 for Experiment 2 and Table 6 for Experiment 3), in almost every case, each individual nudge was associated with large, reliably significant differences that were opposite what would be expected (i.e., significantly higher average acceptance in the conventional bias condition than the recycled bias condition). Consistent with the notion above, this might provide some evidence of the ability of education to increase reliability across the scenarios (i.e., because the educational intervention was shown to increase average acceptance across all the scenarios, the consistent effects suggest that participants might have started and stayed high in acceptance, regardless of nudges).

Analytic strategy and assumption checks. The primary result of interest for this experiment was the extent to which consistency would differ for those in the education condition as opposed to those in the control condition (i.e., the participants who only received the series of nudges). Because of the low sample size and because previous research has never found systematic interactions between educational interventions and nudges (e.g., defaults) when measuring recycled water acceptance (Tanner, 2021; Tanner, Cokely, et al., 2023; Tanner & Feltz, 2022), I analyzed results using independent samples t-tests.

For this experiment, I used the same two methods of operationalizing consistency from Experiment 1b. For the first method, I calculated a standard deviation for each individual's responses to the default scenario, as well as the six additional nudges (thus, consistent with Experiment 1b, lower standard deviations should indicate less spread among a participant's own responses and more consistency in acceptance). I then averaged those individual standard deviations in the education group and control group. The average standard deviation in the education group was then compared to the average standard deviation in the control group to compare consistency.

Before running those t-tests, I checked to see if the normality and homogeneity of variance assumptions for conducting t-tests were met. First, I checked skew and kurtosis statistics for both groups to ensure they were not substantially non-normal. While there are many different standards for what skew and kurtosis values should be considered non-normal (e.g., some have suggested absolute skew values greater than two or absolute kurtosis values larger than seven are substantially non-normal; see Kim, 2013), I used a more conservative criteria (absolute skew values greater than one and absolute kurtosis values larger than three; e.g., Blanca et al., 2013). On these criteria, neither group was substantially non-normal (education group: skewness = -.09, kurtosis = -.43; control group: skewness = -.14, kurtosis = -.92). Then, I conducted Levene's test (Levene, 1960; see also Lim & Loh, 1996) to check for homogeneity of variance. Levene's test was not significant (p = .83) suggesting that the assumption of homogeneity of variance was also met.

I also calculated individual standard deviations for each of the additional indicators (being upset, trust, worry, intentions to protest, intentions to move, and intentions to give negative recommendations) and averaged each by education group. The assumptions of normality held for each of these indicators (education group: skewness range [.18, .94], kurtosis range [-.95, .78]; control group: skewness range [-.31, .72], kurtosis range [-.56, .03]). Likewise, Levene's test was not significant for any of the additional indicators (all *ps* > .29), thus analyses for additional indicators were conducted using independent samples t-tests as well.

The logic of the second operationalization of consistency (i.e., qualitative switches between preferences) was identical to that reported in Experiment 1b. Assumptions for t-tests held for this variable (education group: skewness = -.24, kurtosis = -.70; control group: skewness = -.33, kurtosis = -.53; Levene's test: p = .62).

Consistency results: standard deviations. The independent samples t-test for the primary acceptance indicator revealed a significant difference between consistency in the education condition (M_{SD} = 1.21, SD_{SD} = .44) and control condition (M_{SD} = 1.40, SD_{SD} = .40; t (80) = 2.07, p = .04, d = .46; see Figure 10). Specifically, in support of my third hypothesis, consistency was higher in the education condition than it was in the control condition. Notably, the same pattern of results held for all of the additional indicators as well.

In particular, those in the education condition were more consistent in their levels of being upset (M_{SD} = .55, SD_{SD} = .48) than were those in the control condition (M_{SD} = .87, SD_{SD} =

.52; t (80) = 2.96, p < .01, d = .66; see Figure 11) and were also more consistent in their rated likelihood of moving in response to recycled water (M_{SD} = .40, SD_{SD} = .45) than the control participants (M_{SD} = .65, SD_{SD} = .50; t (80) = 2.33, p = .02, d = .52; see Figure 12). Those who were in the education condition were marginally more consistent in their levels of trust (M_{SD} = .60, SD_{SD} = .47) than those who were not (M_{SD} = .77, SD_{SD} = .41; t (80) = 1.76, p = .08, d = .39; see Figure 13). Those in the education condition were marginally more consistent in how worried they would be about recycled water (M_{SD} = .55, SD_{SD} = .45) than those who were not (M_{SD} = .72, SD_{SD} = .42; t (80) = 1.78, p = .08, d = .39; see Figure 14), Finally, educated decision-makers were marginally more consistent in their rated likelihood of protesting recycled water (M_{SD} = .46, SD_{SD} = .46) than those in the control condition (M_{SD} = .62, SD_{SD} = .41; t (80) = 1.70, p = .09, d = .38; see Figure 15) and were also marginally more consistent in their intentions to give negative recommendations (education condition: M_{SD} = .41, SD_{SD} = .46; control condition: M_{SD} = .61, SD_{SD} = .55; t (80) = 1.77, p = .08, d = .39; see Figure 16).

Notably, numeracy was also significantly correlated with acceptance variation (r (80) = -.30, p < .01; see Table 4 for correlations from Experiments 2 and 3). Those who were higher in numeracy exhibited higher levels of consistency in their acceptance across the scenarios. These results seem consistent with prior research on numeracy (e.g., Allan et al., 2017 has found that numeracy predicts susceptibility to biasing tornado-related myths), and thus may point to the skilled decision framework and representative understanding pathways as a potential model to explain these effects.

Consistency results: qualitative switches. The independent samples t-test comparing switches in the education group (M = 3.39, SD = 1.52) to switches in the control group (M = 3.61, SD = 1.51) did not yield reliably significant results (t (80) = .66, p = .51, d = .15). However further analyses suggested that this could be a function of averaging across all individuals. For example, when examining patterns related to the extremity of participants' response to the first scenario (i.e., the default nudge), it seemed that among individuals with more extreme answers, the educational decision aid made a difference. For example, examining only those who responded with a 1 or 6, there was a medium-sized effect where those in the control group switched more (N = 11, M = 3.91, SD = 1.64) than those in the education condition (N = 9, M = 3.11, SD = 1.83; t (18) = 1.03, p = .32, d = .46). Notably, these results did not reach conventional levels of statistical significance, but the effect size suggests this may be function of the very small groups. Similarly, when including those who were next most extreme in their response (i.e., those who answered with 1, 2, 5, or 6), a nearly marginally significant effect occurred where those in the education condition made fewer switches throughout the experiment (N = 29, M = 3.68, SD = 2.02) than those in the control condition (N = 28, M = 4.45, SD = 1.55; t (55) = 1.62, p = .11, d = .43). Only when including those who were the least extreme (i.e., those who responded with a 3 or 4) did the effect seem to shrink, as reported above. This likely suggests that many people were being pushed around in their preferences, but effects may have been dampened by a smaller group of participants who were not shifting at all, regardless of other factors.

Thus, I hypothesized that the lack of reliably significant results might be due to systematic differences between some of the participants in the study. For example, some participants might generally have been more resistant to change than others (see Hoang et al., 2023). To test this, I conducted a latent profile analysis (LPA) on participants' responses to the seven scenarios. Results suggested that a two-profile solution likely had adequate fit to the data (*entropy* = .77, *AIC* = 2046.73, *BIC* = 2099.95, *SABIC* = 2030.55) and the Lo-Mendell-Rubin adjusted likelihood ratio test (Lo et al., 2001) suggested that a two-profile solution would likely fit the data better than a one-profile solution (*AIC* = 2092.79, *BIC* = 2126.66, *SABIC* = 2082.50; *p* = .06). I also tested a three-profile solution (*entropy* = .87, *AIC* = 2015.63, *BIC* = 2088.19, *SABIC* = 1993.57); however, the Lo-Mendell-Rubin test suggested it would not provide significantly better statistical fit (*p* = .12). Thus, I retained a two-profile solution.

The two-profile solution revealed one class (Class 1; *N* = 35) that was generally lower in preference for recycled water compared to the other class (Class 2; *N* = 47; see Figure 17 for estimated means from the LPA). Accordingly, I conducted a 2x2 ANOVA comparing education condition (control vs. education) and latent class (Class 1 vs. Class 2) on the number of qualitative switches made throughout the experiment. There was no main effect of education condition on switches (F (1, 78) = .02, p = .89, $\eta^2 < .01$). There was a main effect of latent class on switches (F (1, 78) = 7.99, p = .01, $\eta^2 = .09$), such that those in Class 1 (the lower acceptance group) switched more (*M* = 4.66, *SD* = 1.42) than those in Class 2 (the higher acceptance group; *M* = 3.62, *SD* = 1.82). However, this was qualified by a significant interaction (F (1, 78) = 5.67, p = .02, $\eta^2 = .06$), where those in the control condition had a similar number of switches regardless of class (Class 1: *M* = 4.25, *SD* = 1.37; Class 2: *M* = 4.09, *SD* = 1.68), but those in the education condition switched far less when they were in Class 2 (*M* = 3.17, *SD* = 1.88) as opposed to Class 1 (*M* = 5.07, *SD* = 1.39; see Figure 18). These results support the idea that averaging across participants might have muted effects on qualitative switching.

Discussion. In sum, these results likely all lend support to Hypothesis 3. Across many different scenarios and indicators of acceptance, there seems to be a pattern where educating participants about recycled water helps them become more consistent in their choices and preferences than those who are only exposed to a series of nudges. Moreover, these results seem to hold both when consistency is operationalized using standard deviations, as well as when consistency is operationalized using qualitative switches (at least for those for those in Class 2, who were generally higher in acceptance). These results might suggest that properly tested educational interventions may be able to support reliable deliberation to a greater extent than a series of nudge interventions.

Experiment 3

Experiment 3 aimed to accomplish two main goals. First, while results from Experiment 2 were promising, its small sample size was a notable limitation. Thus, Experiment 3 aimed to replicate the results using a larger group of participants. Second, Experiment 2 did little to examine the cognitive mechanisms that might explain differences in consistency. Therefore, in Experiment 3, I also aimed to use structural equation modeling techniques (SEM) to explore the potential decision pathways through which effects might occur.

Method

Participants. Participants for Experiment 3 were recruited from the student participant pool at a large university in the United States in exchange for partial course credit. To estimate the number of participants needed for this experiment, I first conducted an a priori power analysis using the acceptance consistency effect size observed in Experiment 2 (d = .46). This suggested a need for 120 participants (power = .80, alpha = .05). However, in order to ensure accuracy when using SEM techniques, Kline (2015) suggests an ideal sample size to parameter ratio of 20:1. To be consistent with models reported by Tanner (2021), I hypothesized that a total of 12 parameters would need to be estimated for my path model. Thus, I aimed to recruit 240 participants (20 participants * 12 parameters) for this experiment.

Responses were collected from 265 participants, but prior to analyses, 58 participants were excluded for failing to complete the survey or meet the minimum age requirement, and 20 participants were excluded for failing attention check questions. Of the remaining 187 participants, ages ranged from 18 to 60 (M = 19.39, SD = 3.27). There were 152 females in the sample and 35 males.

Procedure. The procedure and materials for Experiment 3 were identical to those reported in Experiment 2, with two exceptions. First, two items were added measuring subjective knowledge of recycled water:

1. How much do you know about recycled water for drinking purposes?
2. How confident would you feel about providing someone else with information about water recycling?

These items were drawn from previous literature (Fielding & Roiko, 2014; Schultz & Fielding, 2014) and measure how much participants *feel* they know about recycled water. Responses to both questions were measured using a 7-point Likert scale (1 = *nothing at all/not at all confident*; 7 = *a lot/extremely confident*). Responses to these questions were highly correlated (r (185) = .73, p < .001), thus responses were averaged to form a composite subjective knowledge score.

Additionally, the subjective and objective measures of recycled water knowledge were delivered both pre- and post-intervention (as opposed to Experiment 2, where objective knowledge was only measured post-intervention). Specifically, pre-subjective knowledge was the first measure participants completed in the experiment, followed by pre-objective knowledge. The post-measures were delivered in the same order and were completed after participants finished all of the nudge scenarios. Consistent with Tanner (2021), I calculated difference scores by subtracting pre-scores from post-scores to gauge the amount of subjective and objective knowledge gained after the interventions.

Results

Manipulation checks. In line with past research (Mahmoud-Elhaj et al., 2020; Tanner, 2021; Tanner, McDonald, et al., 2023), those who were shown the educational video had significantly higher difference scores for both objective knowledge (M = 9.12, SD = 6.94) and subjective knowledge (M = 1.10, SD = 1.30) than those shown the control video (objective knowledge: M = 3.75, SD = 5.90, t (185) = 5.69, $p < .001$, d = .83, see Figure 19; subjective knowledge: M = .56, SD = 1.03, t (185) = 3.17, $p < .01$, d = .46, see Figure 20), suggesting that the educational intervention likely did help participants gain knowledge. Likewise, consistent with results from Experiment 2 and my first hypothesis, those in the education condition were more accepting (i.e., exhibited higher preferences for recycled water) on average across all seven scenarios (M = 3.77, SD = .66) than those who were in the control condition (M = 3.32, SD = .70; t (185) = 4.54, $p < .001$, d = .66; see Figure 21). There was no reliably significant difference in acceptance across all scenarios for the first nudge (i.e., defaults; t (185) = .91, p = .36, d = .13). However, consistent with Experiment 2 and with my second hypothesis, within many of the individual nudge scenarios themselves, those who were nudged into recycled water use had significantly higher levels of acceptance than those who were not (see Table 5). For example, in the default scenario alone, those who were defaulted in were significantly higher in acceptance (M = 4.14, SD = 1.37) than those who were defaulted out (M = 3.07, SD = 1.39; t (185) = 5.25, p

< .001, *d* = .77; see Figure 22)[3]. These results suggest that both kinds of interventions successfully increased acceptance.

Assumption checks. I used the same analytic strategy from Experiment 2 in this experiment. Thus, I checked normality and homogeneity of variance assumptions for each of the indicators. For standard deviations across the primary acceptance indicator, neither the education group nor the control group was substantially non-normal (education group: skewness = .50, kurtosis = .30; control group: skewness = .01, kurtosis = -.57). Levene's test was not significant (p = .70) suggesting that the assumption of homogeneity of variance was also met.

Standard deviations for the additional acceptance indicators also met both assumptions. For the assumption of normality, skewness in the education groups ranged from .34 to .93, with kurtosis ranging from -.28 to 1.16. For the control group, skewness ranged from .12 to .36 and kurtosis ranged from -.82 to .26. Likewise, Levene's test was not significant for any indicators (all *ps* > .09). Finally, for qualitative switches, normality was maintained in both groups (education group: skewness = .00, kurtosis = -.81; control group: skewness = -.58, kurtosis = .11), and Levene's test was not significant (p = .24). Because assumptions were met for all indicators, I used independent samples t-tests to analyze all consistency measures in this experiment.

Consistency results: standard deviations. Contrary to results from Experiment 2, the independent samples t-test for the primary acceptance indicator did not detect a reliably significant difference between consistency in the education condition (M_{SD} = 1.22, SD_{SD} = .41)

[3] Also note that results for the additional indicators administered after the default nudge were consistent with prior research, where the educational intervention was always associated with expected significant / marginally significant effects for each the indicators, while the default condition never exhibited any reliable effects (see Table 7).

and control condition (M_{SD} = 1.20, SD_{SD} = .38; *t* (185) = .39, *p* = .70, *d* = .06; see Figure 23). Additionally, for the other indicators, only one significant difference was detected, specifically for the intention to protest indicator (education condition: M_{SD} = .47, SD_{SD} = .39); control condition: M_{SD} = .62, SD_{SD} = .46; *t* (185) = 2.31, *p* = .02, *d* = .34; see Figure 24). None of the other indicators exhibited reliably significant effects (all *ps* > .14; see Table 8).

Consistency results: qualitative switches. For the full sample, the independent samples t-test comparing switches in the education group (*M* = 3.76, *SD* = 1.63) to switches in the control group (*M* = 4.02, *SD* = 1.57) did not yield reliably significant results (*t* (185) = 1.12, *p* = .27, *d* = .16). However, as in Experiment 2, results again suggested that this might be a function of averaging across participants. For example, among those who were the most extreme in acceptance during the default acceptance (i.e., those who answered with a 1 or 6), a large difference occurred, where those in the education condition switched preferences significantly fewer times (*N* = 20, *M* = 2.95, *SD* = 1.19) than those in the control condition (*N* = 15, *M* = 4.40, *SD* = 1.55; *t* (33) = 3.13, *p* < .01, *d* = 1.07). Likewise, when including those who were next most extreme in responses (i.e., those who answered with 1, 2, 5, or 6), a marginally significant effect occurred where those in the education condition made fewer switches throughout the experiment (*N* = 55, *M* = 3.73, *SD* = 1.57) than those in the control condition (*N* = 56, *M* = 4.25, *SD* = 1.60; *t* (109) = 1.74, *p* = .09, *d* = .33). Only when including all participants was the effect not reliably significant, again suggesting that results might have been dampened by a group of people who were not having the preferences moved, regardless of outside factors.

Thus, I again conducted a latent profile analysis (LPA) on participants' responses to the seven scenarios to identify any systematic differences between participants. Fit statistics suggested that a two-, three-, or four-profile solution would likely fit the data best (two-profile:

entropy = .78, *AIC* = 4431.95, *BIC* = 4503.03, *SABIC* = 4433.35; three-profile: *entropy* = .77, *AIC* = 4400.46, *BIC* = 4497.39, *SABIC* = 4402.37); four-profile: *entropy* = .83, *AIC* = 4354.50, *BIC* = 4477.29, *SABIC* = 4356.92). Lo-Mendell-Rubin adjusted likelihood ratio tests supported the three-profile solution. Specifically, the two-profile solution was found to fit the data better than a one-profile solution (*AIC* = 4514.71, *BIC* = 4559.94, *SABIC* = 4515.60; $p < .001$) and the three-profile was suggested to be statistically preferable to a two-factor solution ($p = .04$). Meanwhile, the four-profile solution was not found to provide a better statistical fit than the three-profile solution ($p = .24$).

Despite the 3-profile solution being statistically preferred, I used the 2-profile solution for analyses. This was done for several conceptual and statistical reasons. For example, use of the 3-profile solution would have resulted in one class—Class 3—being disproportionately smaller ($N = 37$, than Class 2 ($N = 64$) and Class 1 ($N = 86$). Likewise, there were no consistent features that made Class 2 easily interpretable (e.g., for several of the nudges, acceptance was very similar to Class 1, while for others, acceptance was either lower or higher than all other classes). Moreover the entropy did not increase for the three-profile solution over the two-profile solution. Thus, to be consistent with Experiment 2, I again used the two-profile solution in this experiment.

Similar to Experiment 2, among the two identified classes, one class (Class 1; $N = 72$) was lower in preference for recycled water compared to the other class (Class 2; $N = 115$; see Figure 25 for estimated means from this LPA). Accordingly, I again conducted a 2x2 ANOVA comparing education condition (control vs. education) and latent class (Class 1 vs. Class 2) on the number of qualitative switches made throughout the experiment. There was no main effect of education condition on switches (F (1, 183) = .24, $p = .63$, $\eta^2 < .01$). Likewise, there was no main effect of latent class on switches (F (1, 183) = .94, $p = .34$, $\eta^2 = .01$. However, there was a marginally significant interaction (F (1, 183) = 3.47, p = .06, η^2 = .02), where in the control condition, those in Class 1 (M = 3.91, SD = 1.69) and those in Class 2 (M = 4.13 SD = 1.47) had similar numbers of switches, whereas in the education condition, those in Class 2 switched less (M = 3.55, SD = 1.64) than Class 1 (M = 4.24, SD = 1.55; see Figure 26). These results again support the idea that averaging across participants might have muted effects on qualitative switching.

Decision model. To model the potential cognitive mechanisms explaining these effects on consistency, I constructed a structural equation model based on Cokely et al.'s (2018) model of skilled decision-making. As discussed previously, this model suggests that factors such as domain-general skills (e.g., numeracy) and educational interventions typically influence choice by helping individuals integrate knowledge with their values and emotions, which are thought to then primarily drive the decisions one makes. Accordingly, after examining relations among variables in this experiment, I hypothesized that the educational intervention would likely increase consistency (operationalized by number of switches) indirectly by increasing knowledge, which in turn would decrease negative affect (operationalized by a participant's average reported worry across the seven scenarios), leading to a lower number of switches. Because statistical numeracy has been found to predict appropriate affective reactions and risk perceptions (e.g., Ramasubramanian, 2022; Tanner, 2021) I also hypothesized that statistical numeracy would have an indirect effect on consistency through worry.

Because of the two latent profiles identified through LPA, I tested a multigroup structural equation model, with class as the grouping variable. The fully constrained model had poor fit to the data ($\chi 2$ (14) = 50.14, p < .001; $RMSEA$ = .17, 90% CI [.12, .22]; CFI = .57; TLI = .45, $SRMR$ = .14), so I systematically freed path estimates individually to gauge differences between

Class 1 and Class 2. Chi-square difference tests suggested that freeing the path from worry to number of switches and the path from knowledge to worry each individually led to a statistically better fit (worry→switch: $\Delta\chi2$ (1) = 29.92, $p < .001$; knowledge→worry: $\Delta\chi2$ (1) = 6.95, $p =$ .01). Thus, I tested a final model with these two paths unconstrained. The unconstrained model had good fit to the data ($\chi2$ (12) = 13.27, $p = .35$; *RMSEA* = .03, 90% CI [.00, .11]; *CFI* = .99; *TLI* = .98, *SRMR* = .06)

In particular, for Class 2 (i.e., the higher acceptance group), results were consistent with the framework for skilled decisions (see Figure 27). Average worry was a strong predictor of the number of times one switched their preferences regarding recycled water ($\beta = .52$, $p < .001$). Numeracy had a significant effect on worry, such that those who were higher in numeracy exhibited less worry ($\beta = -.20$, $p = .001$). Likewise, the educational intervention also exhibited an indirect effect on consistency through increased knowledge ($\beta = .34$, $p < .001$), which led to lower levels of worry ($\beta = -.32$, $p < .001$).

Conversely, for Class 1 (the lower acceptance group), knowledge was not related to worry ($\beta = .04$, $p = .71$). In turn, worry was negatively associated with switching (worry→switch: $\beta = -.28$, $p = .01$; see Figure 28), meaning that those who were more worried switched less than those who were more worried. These results might be attributable to the lower overall acceptance exhibited in Class 1. Because participants in this class were already lower in acceptance, those who were more worried likely consistently displayed preferences on the low end of the scale (i.e., they likely consistently answered within the 1-3 range, meaning fewer switches between conventional and recycled water preferences). Meanwhile, those who were less worried were probably slightly higher in acceptance (e.g., they might have had more responses of 3 or 4, which might have resulted in more qualitative switches). The opposite may have been true for those in Class 2. That is, because they were generally higher in acceptance, those in Class 2 who were less worried might have been more consistently in the 4-6 response range, while those who were more worried might have fallen closer the middle and triggered more qualitative switches.

Discussion. While results from this experiment were not as clear as results from Experiment 2, patterns here seem to also support Hypothesis 3. Unlike results observed previously, consistency measured using standard deviations displayed no significant associations with the interventions used. However, when measured using qualitative switches between preferences among those who were higher in acceptance, the same pattern observed previously emerged, where educating participants about recycled water helped them become more consistent in their choices and preferences than those who were only exposed to the series of default nudges and biases. These results might suggest the null results observed for consistency when using standard deviations could be at least partially attributable to the instability of variance calculations when using a small sample size.

These results also add to the prior experiments by showing a potential cognitive mechanism through which choice consistency might occur. The structural equation model suggests that for those who were naturally higher in acceptance, these effects might proceed along a route consistent with the model for skilled decision. Specifically, those participants seemed more likely to achieve reliable deliberation if they had the skills and/or tools necessary to integrate correct knowledge with appropriate emotions, such as lower worry about health risks.

General Discussion

The purpose of these experiments was to test the extent to which educational interventions and libertarian paternalistic nudges would support the reliable deliberation condition for autonomy when making decisions regarding recycled water. Across both Experiments 2 and 3, and consistent with prior research (e.g., Tanner, 2021; Tanner & Feltz, 2022), results converged to suggest that both education and nudges can affect acceptance of recycled water in various circumstances. However, results also revealed important differences between the interventions. For example, in both experiments, nudges typically only increased acceptance within their own individual scenario (and even within individual scenarios, only some of the nudges had reliably significant effects in the hypothesized direction), while the educational intervention increased acceptance on average across all scenarios. Likewise, in both experiments, the educational intervention helped people be more consistent in their preferences. This was evidenced by the educated decision-makers tending to make fewer qualitative switches between preferring recycled or conventional water throughout the course of the seven scenarios (as opposed to those in the control condition, who only received the nudges).

These results hold important theoretical, ethical, and practical value. On a theoretical level, results concerning consistency and the Class 2 path model presented in Experiment 3 seem consistent with Cokely et al.'s (2018) framework for skilled decision-making. Previous research has shown that the decision to accept recycled water likely primarily proceeds along a pathway through which transparent decision aids increase knowledge, which then decreases worry about recycled water and, in turn, increases acceptance (Tanner, 2021; Tanner & Feltz, 2022). However, to my knowledge, this is the first research to demonstrate not only that the skilled decision framework can often describe and predict decisions, but that it also could potentially describe and predict the *consistency* of one's decisions across iterative, biasing decision tasks.

Moreover, on a broad scale, these results provide some support for the idea of some transparent decision aids primarily working by increasing representative understanding. As stated in the introduction, representative understanding has sometimes been defined as an understanding of a topic that is nuanced enough that additional information received on that topic is unlikely to change tendencies made on the basis of that understanding. Because these results show that education helped people reach more stable decisions—even when multiple pieces of biasing information were presented thereafter—these results might be thought to provide direct empirical evidence of the ability of educational decision aids to increase representative understanding of recycled water.

Of note, the results might also provide theoretical evidence that some individuals might be naturally better at reliably deliberating than others. In both experiments, statistical numeracy was significantly associated with lower standard deviations across the scenarios, and the path models also suggest a potential indirect effect through decreased worry. This seems consistent with previously expressed notions (e.g., Mele, 2001) that some executive skills or dispositions might be able to help individuals avoid irrelevant cognitive biases. Effects of education and knowledge through worry were stronger than the indirect effects of numeracy through worry. This is understandable, given that numeracy could be thought to be domain general (i.e., higher numeracy is generally associated with lower risk perceptions across many domains; see Ramasubramanian, 2020, 2022). Thus, it is perhaps unsurprising that numeracy would predict slightly lower worry and higher choice consistency in this domain, while domain-specific knowledge would have a stronger relationship with recycled water-specific worry. Overall, however, these results might point to statistical numeracy as an important individual difference to consider when evaluating differences in choice consistency.

Finally, these results might lend some credence to virtue ethics as theoretical model capable of supporting ethical comparisons among interventions. As discussed in the introduction, as a standalone theory, virtue ethics might not have *the* answer to what kinds of virtues and character development matter when evaluating choice interventions. However, virtue ethics, combined with the principles presented in the APA code of ethics, may serve as a valuable guide to determining the virtues to consider (e.g., by identifying potential habits and character traits that people in general might view as essential to make decisions that are true to themselves). Then, by following the pattern developed in this dissertation (e.g., developing innovative techniques to measure the extent to which they respect—and perhaps promote—those values), a more integrated, empirically-backed theory of ethical interventions might be achieved.

Practically, these results might be important in many domains. In many cases, it is important that decision-makers are able to make consistent, reliable decisions. This is likely especially true within the applied decision domain of recycled potable water. Despite recycled water representing one promising method to help overcome concerning water scarcity issues (e.g., Chamberlain & Sabatini, 2022; Elimelech, 2006), psychological barriers such as disgust, low trust in authority, and low knowledge of recycled water often result in low public acceptance of recycled potable water (e.g., Glick et al., 2019; Rozin et al., 2015; Wester et al., 2016). In some cases, these factors have sometimes resulted in proposed water recycling projects being delayed or struck down altogether (e.g., San Diego, United States and Toowoomba, Australia; (Hurlimann & Dolnicar, 2010; Po et al., 2003; Uhlmann & Head, 2011). Accordingly, knowledge of interventions that can consistently help individuals overcome these significant hurdles may be especially important to policymakers and water practitioners as they seek to design and implement the best possible interventions.

Together with past research, results from these experiments suggest that when done correctly, effective educational interventions might be especially fit for this task in the domain of recycled water. Notably, not all educational interventions are equally likely to have desired effects. Factors such as the medium through which information is presented and the areas of educational content that are covered are likely to have an impact on the extent to which educational interventions are able produce reliable results. For example, Tanner and Feltz (2022) found infographics about recycled water to be ineffective at changing recycled water acceptance, and Leong and Lebel (2020) found that text descriptions of the technology used to purify recycled water had no effect on participants' odds of choosing a bottle of recycled drinking water over a bottle of mineral water. Meanwhile, Tanner, McDonald, et al. (2023) found that educational videos that emphasized specific areas of educational content (e.g., approaches to implementing recycled water through indirect, direct, or de-facto reuse methods and information about locations successfully using recycled water) were better able to increase acceptance of recycled water than other areas of educational content (e.g., the need for recycled water and the technology used for purification).

By using empirically validated interventions like the educational intervention used in these experiments, however, policymakers and water practitioners may be able to increase acceptance and help individuals be more consistent in their preferences regarding recycled water. Especially in this domain, consistency is likely very important. For example, if individuals are nudged into making a specific choice regarding recycled water, but later realize the opposite choice is more aligned with their values, emotions, and beliefs, consequences might be more severe (e.g., individuals might protest, threaten legal action, use more bottled water, and contribute more plastic waste to landfills, etc.). These results could also be of practical use to decision scientists, as they could provide a valuable template for measuring effectiveness and evaluating ethical merits of various interventions in other applied decision domains.

Finally, from an ethical perspective, these results show that there could be a potential ethical cost in terms of reliable deliberation when opting to use libertarian paternalistic nudges over educational interventions. Given that libertarian paternalistic interventions have been argued to help decision-makers' lives better "as judged by their own preferences, not those of some bureaucrat" (Thaler & Sunstein, 2008, p. 10), it is concerning that by using the nudges in this context, individuals likely had their preferences moved around to a greater extent than those who were educated beforehand. Perhaps even more concerning, across both experiments, the participants who expressed the most extreme opinions after receiving the first nudge were the ones who were most likely to switch their qualitative preferences throughout the experiment if they did not receive the educational intervention. Likewise, in both experiments, it was only those classes that were higher in acceptance that experienced significantly higher amounts of qualitative switches in absence of the educational intervention. Of course, these results only apply to qualitative switches. It is possible that if finer-grained switches were to be examined (e.g., switches from a response of 1 (*strongly prefer conventional water*) to a response of 2 (*prefer conventional water*), these results might not generalize). However, were this pattern to hold, these results would seem to suggest that rather than helping individuals make choices consistent with their own preferences, the nudges might instead shift their preferences to resemble whatever preference was salient in each individual scenario (or, at the very least, not help people stay as consistent in their preference as education does).

As stated previously, we would not expect one's values relevant to recycled water to change in a short period of time, even when exposed to additional pieces of irrelevant, biasing information. Thus, these results might suggest that nudges are likely unable to respect reliable deliberation to the same extent as educational interventions (i.e., they are likely unable to safeguard individuals against the kinds of biases that would make one more likely to make decisions that are inconsistent with their own values, emotions, and beliefs). Combined with results from Tanner (2021), this demonstrates another condition for autonomy on which libertarian paternalistic nudges may be inferior to educational interventions. This is concerning, given the intrinsic, instrumental, and professional value of autonomy discussed previously.

However, as stated by Tanner (2021), such ethical costs do not imply that libertarian paternalistic nudges should never be used. Nudges have some notable benefits that cannot be overlooked (e.g., they are typically easy and cost-effective to implement and often offer more immediate results than other interventions). Thus, this underscores the need for cost-benefit analyses examining tradeoffs between ethical costs to autonomy and implementation benefits (see Hertwig, 2017; Trout, 2005). In order to most successfully accomplish this goal, future research should continue to estimate potential ethical costs by examining the coercion condition of autonomy.

Notably, the method developed here to measure preference consistency may be of value in such cost-benefit analyses. One potential reason why ethical considerations are often not weighted in cost-benefit analyses might be because ethics are typically discussed on a theoretical level, rather than an empirical level. Given the empirical nature of this ethical investigation, it is perhaps possible that information such as the size of effects observed in these experiments (e.g., comparing switching in the education condition to the control condition) could be used to assign actual *values* to the cost of using a libertarian paternalistic nudge over an educational intervention. Then, if policymakers were to provide guidance regarding how much an ethical cost should be weighted versus the benefits (e.g., monetary savings) of using a nudge intervention, these figures could be integrated into a proper cost-benefit analysis (see Gronlund et al., 2015 for a similar discussion concerning weighing costs and benefits of false eyewitness identifications).

Encouragingly, given appropriate time and resources to develop and test suitable nudges, educational interventions, etc., these methods could likely be replicated in many applied decision domains (e.g., plastic recycling, end-of-life care decisions, retirement savings). Especially if measurement techniques are refined, using the technique demonstrated here to iteratively push and pull participants toward different preferences could provide a tool to measure ethical costs to reliable deliberation across a host of domains. Likewise, this method may be of practical use in other fields of research. For example, in the field of debiasing, the iterative push/pull technique used to bias participants toward specific option would likely represent a novel approach to testing the effectiveness of various techniques. Applied properly, this method could be helpful in extending debiasing research to problems wherein a normatively superior option or solution might not exist and could help researchers further explore various debiasing techniques that were not employed here.

Of course, some limitations were present in these experiments. For example, as previously discussed, the measurement of consistency using standard deviations might have been unstable. Accordingly, it is possible that the relationships from Experiment 2 observed using those measurements may have been spurious. This seems unlikely, given that all the variation indicators in that experiment were significant and the qualitative switches variable pointed to the same relationship. However, future research should likely seek to address the variation measurement problems we observed here (e.g., by increasing the number of scenarios presented to participants). Next, the scenarios participants responded to were all hypothetical. Thus, it is possible that results from this study would not generalize well to real-world choices. Future research might wish to replicate these findings using behavioral measures of recycled water acceptance (e.g., by having participants actually consume recycled water during the study; Fielding & Roiko, 2014). Finally, because of factors like lower enrollment in the study and inattention during the survey, sample sizes were generally lower than desired. Especially given that latent profiling techniques are often considered to be large sample techniques (e.g., common rules of thumb suggest that they should typically include at least 300 participants; Nylund-Gibson & Choi, 2018), Experiments 2 and 3 should likely be replicated using more substantial (and perhaps more demographically diverse) samples.

In summary, this research suggests that while multiple techniques can be used to influence individuals' preferences and choices in desired directions, there are important ethical (and practical) factors that must be considered when choosing which interventions to use. Using a novel technique aimed at measuring reliable deliberation, I have shown that educational interventions are often more likely than libertarian paternalistic nudges to encourage stable, consistent preferences regarding recycled water. This represents an important finding for those seeking to find ethically defensible interventions that can best help individuals accept recycled water. Continuing to refine these measurements and developing techniques to empirically measure risks on other ethical criteria represents a difficult but important step in better defining the most appropriate ways of interacting with those we seek to help.

References

Alexander, L., & Moore, M. (2021). Deontological Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/

Allan, J., Ripberger, J. T., Ybarra, V., & Cokely, E. T. (2017). Tornado Risk Literacy: Beliefs, Biases, and Vulnerability. In J. Gore & P. Ward (Eds.), *Naturalistic Decision Making and Uncertainty. Proceedings of the 13th Bi-Annual International Conference on Naturalistic Decision Making* (pp. 284–290). The University of Bath.

Allum, N., Sturgis, P., Tabourazi, I., & Brunton-Smith, I. (2008). Science knowledge and attitudes across cultures: A meta-analysis. *Public Understanding of Science*, *17*(1), 35–54.

American Counseling Association. (2014). *ACA Code of Ethics*. American Counseling Association. https://www.counseling.org/resources/aca-code-of-ethics.pdf

American Nurses Association. (2015). *Code of Ethics for Nurses With Interpretive Statements*. American Nurses Association. https://www.nursingworld.org/practice-policy/nursing-excellence/ethics/code-of-ethics-for-nurses/coe-view-only/

American Occupational Therapy Association. (2020). AOTA 2020 Occupational Therapy Code of Ethics. *The American Journal of Occupational Therapy*, *74*(Supplement_3), 7413410005p1-7413410005p13. https://doi.org/10.5014/ajot.2020.74S3006

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct (2002, amended effective June 1, 2010, and January 1, 2017)*. http://www.apa.org/ethics/code/index.html

American Psychological Association, & American Psychological Association Services, Inc. (2019). *Impact APA: American Psychological Association Strategic Plan*. https://www.apa.org/about/apa/strategic-plan/impact-apa-strategic-plan.pdf

Anderson, R. A., Worthington, L., Anderson, W. T., & Jennings, G. (1994). The development of an autonomy scale. *Contemporary Family Therapy*, *16*(4), 329–345. https://doi.org/10.1007/BF02196884

Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology*, *110*(5), 766–781. https://doi.org/10.1037/pspp0000066

Arpaly, N., & Schroeder, T. (2012). Deliberation and Acting for Reasons. *The Philosophical Review*, *121*(2), 209–239. https://doi.org/10.1215/00318108-1539089

Bales, R. E. (1971). Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure? *American Philosophical Quarterly*, *8*(3), 257–265.

Bar-Eli, M., Azar, O. H., Ritov, I., Keidar-Levin, Y., & Schein, G. (2007). Action bias among elite soccer goalkeepers: The case of penalty kicks. *Journal of Economic Psychology*, *28*(5), 606–621. https://doi.org/10.1016/j.joep.2006.12.001

Baron, J. (2004). Normative Models of Judgment and Decision Making. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (1st ed). Blackwell Pub.

Baron, J. (2005). *Rationality and Intelligence*. Cambridge University Press.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569–579. https://doi.org/10.1037/0022-3514.54.4.569

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should Governments Invest More in Nudging? *Psychological Science*, *28*(8), 1041–1055. https://doi.org/10.1177/0956797617702501

Benn, S. I. (1975). Freedom, Autonomy and the Concept of a Person. *Proceedings of the Aristotelian Society*, *76*, 109–130. https://www.jstor.org/stable/4544883

Bentham, J. (1825). *The Rationale of Reward*. John and H. L. Hunt.

Blacksher, E., Diebel, A., Forest, P.-G., Goold, S. D., & Abelson, J. (2012). What Is Public Deliberation? *Hastings Center Report*, *42*(2), 14–16. https://doi.org/10.1002/hast.26

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and Kurtosis in Real Data Samples. *Methodology*, *9*(2), 78–84. https://doi.org/10.1027/1614-2241/a000057

Blumenthal-Barby, J. S., & Burroughs, H. (2012). Seeking Better Health Care Outcomes: The Ethics of Using the "Nudge." *The American Journal of Bioethics*, *12*(2), 1–10. https://doi.org/10.1080/15265161.2011.634481

Bovens, L. (2009). The Ethics of Nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology* (pp. 207–219). Springer Netherlands. https://doi.org/10.1007/978-90-481-2593-7_10

Brink, D. O. (1986). Utilitarian Morality and the Personal Point of View. *The Journal of Philosophy*, *83*(8), 417–438. https://doi.org/10.2307/2026328

Broeck, A. V. den, Vansteenkiste, M., Witte, H. D., Soenens, B., & Lens, W. (2010). Capturing autonomy, competence, and relatedness at work: Construction and initial validation of the

Work-related Basic Need Satisfaction scale. *Journal of Occupational and Organizational Psychology*, *83*(4), 981–1002. https://doi.org/10.1348/096317909X481382

Buchanan, A. E., & Brock, D. W. (1989). *Deciding for Others: The Ethics of Surrogate Decision Making*. Cambridge University Press.

Chamberlain, J. F., & Sabatini, D. A. (2022). *Fundamentals of Water Security: Quantity, Quality, and Equity in a Changing Climate* (1st ed.). John Wiley & Sons, Inc.

Chernev, A., & Blair, S. (2015). Doing Well by Doing Good: The Benevolent Halo of Corporate Social Responsibility. *Journal of Consumer Research*, *41*(6), 1412–1425. https://doi.org/10.1086/680089

Cho, J., Cokely, E., Ramasubramanian, M., Allan, J., Feltz, A., & Garcia-Retamero, R. (2021). *Risk Literacy Promotes Representative Understanding: Numerate People are Less Biased, More Knowledgeable, and More Concerned about Climate Change* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-420681/v1

Cobb-Clark, D. A., & Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters*, *115*(1), 11–15. https://doi.org/10.1016/j.econlet.2011.11.015

Cokely, E. T., Feltz, A., Ghazal, S., Allan, J., Petrova, D., & Garcia-Retamero, R. (2018). Decision Making Skill: From Intelligence to Numeracy and Expertise. *Cambridge Handbook of Expertise and Expert Performance*, 476–505.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7*(1), 25–47. http://journal.sjdm.org/11/11808/jdm11808.pdf

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*(1), 14.

Copp, D. (Ed.). (2006). *The Oxford handbook of ethical theory*. Oxford University Press.

Croson, R., & Sundali, J. (2005). The Gambler's Fallacy and the Hot Hand: Empirical Data from Casinos. *Journal of Risk and Uncertainty*, *30*(3), 195–209. https://doi.org/10.1007/s11166-005-1153-2

Dai, Y., Yu, W., & Shen, F. (2021). The Effects of Message Order and Debiasing Information in Misinformation Correction. *International Journal of Communication*, *15*, Article 0.

Danielson, P. (1998). *Modeling Rationality, Morality, and Evolution*. Oxford University Press.

Dietz, T. (2013). Bringing values and deliberation to science communication. *Proceedings of the National Academy of Sciences*, *110*(supplement_3), 14081–14087. https://doi.org/10.1073/pnas.1212740110

Driver, J. (2022). The History of Utilitarianism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2022/entries/utilitarianism-history/

Dworkin, G. (1981). THE CONCEPT OF AUTONOMY. *Grazer Philosophische Studien*, *12*(1), 203–213. https://doi.org/10.1163/18756735-90000122

Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press.

Egebark, J., & Ekström, M. (2016). Can indifference make the world greener? *Journal of Environmental Economics and Management*, *76*, 1–13. https://doi.org/10.1016/j.jeem.2015.11.004

Elimelech, M. (2006). The global challenge for adequate and safe water. *Journal of Water Supply: Research and Technology - Aqua*, *55*(1), 3–10. https://doi.org/10.2166/aqua.2005.064

Elwyn, G., & Miron-Shatz, T. (2010). Deliberation before determination: The definition and evaluation of good decision making. *Health Expectations*, *13*(2), 139–147. https://doi.org/10.1111/j.1369-7625.2009.00572.x

Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989X07304449

Feinberg, J. (1986). *Harm to Self: The Moral Limits of the Criminal Law* (Vol. 3). Oxford University Press.

Felsen, G., Castelo, N., & Reiner, P. B. (2013). Decisional enhancement and autonomy: Public attitudes towards overt and covert nudges. *Judgment and Decision Making*, *8*(3), 202–213. https://journal.sjdm.org/12/12823/jdm12823.pdf

Feltz, A. (2015). Ethical Information Transparency and Sexually Transmitted Infections. *Current HIV Research*, *13*(5), 421–431.

Feltz, A. (2016). Financial Surrogate Decision Making: Lessons from Applied Experimental Philosophy. *The Spanish Journal of Psychology*, *19*. https://doi.org/10.1017/sjp.2016.54

Feltz, A., & Cokely, E. (2013). Predicting Philosophical Disagreement. *Philosophy Compass*, *8*(10), 978–989. https://doi.org/10.1111/phc3.12076

Feltz, A., & Cokely, E. T. (2012). The Philosophical Personality Argument. *Philosophical Studies*, *161*(2), 227–246. https://doi.org/10.1007/s11098-011-9731-4

Feltz, A., & Cokely, E. T. (2016). Personality and Philosophical Bias. In *A Companion to Experimental Philosophy* (pp. 578–589). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118661666.ch41

Fielding, K. S., & Roiko, A. H. (2014). Providing information promotes greater public support for potable recycled water. *Water Research*, *61*, 86–96. https://doi.org/10.1016/j.watres.2014.05.002

Fišar, M., Reggiani, T., Sabatini, F., & Špalek, J. (2022). Media negativity bias and tax compliance: Experimental evidence. *International Tax and Public Finance*, *29*(5), 1160–1212. https://doi.org/10.1007/s10797-021-09706-w

Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: Overcoming challenges in people's understanding of risks. *Social Science & Medicine*, *70*(7), 1019–1025. https://doi.org/10.1016/j.socscimed.2009.11.031

Garcia-Retamero, R., Sobkow, A., Petrova, D., Garrido, D., & Traczyk, J. (2019). Numeracy and Risk Literacy: What Have We Learned so Far? *The Spanish Journal of Psychology*, *22*. https://doi.org/10.1017/sjp.2019.16

Gastil, J., & Dillard, J. P. (1999). Increasing Political Sophistication Through Public Deliberation. *Political Communication*, *16*(1), 3–23. https://doi.org/10.1080/105846099198749

Gert, B., & Culver, C. M. (1979). The Justification of Paternalism. *Ethics*, *89*(2), 199–210. https://doi.org/10.1086/292097

Ghesla, C., Grieder, M., & Schmitz, J. (2019). Nudge for Good? Choice Defaults and Spillover Effects. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00178

Glick, D. M., Goldfarb, J. L., Heiger-Bernays, W., & Kriner, D. L. (2019). Public knowledge, contaminant concerns, and support for recycled Water in the United States. *Resources, Conservation and Recycling*, *150*, 104419. https://doi.org/10.1016/j.resconrec.2019.104419

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Chapter One - Conducting an Eyewitness Lineup: How the Research Got It Wrong. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 63, pp. 1–43). Academic Press. https://doi.org/10.1016/bs.plm.2015.03.003

Gustavsson, J. P., Weinryb, R. M., Göransson, S., Pedersen, N. L., & Åsberg, M. (1997). Stability and predictive ability of personality traits across 9 years. *Personality and Individual Differences*, *22*(6), 783–791. https://doi.org/10.1016/S0191-8869(96)00268-1

Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public Views on Policies Involving Nudges. *Review of Philosophy and Psychology*, *6*(3), 439–453. https://doi.org/10.1007/s13164-015-0263-2

Halpern, D., & Sanders, M. (2016). Nudging by government: Progress, impact, & lessons learned. *Behavioral Science & Policy*, *2*(2), 52–65. https://doi.org/10.1353/bsp.2016.0015

Halpern, S. D., Loewenstein, G., Volpp, K. G., Cooney, E., Vranas, K., Quill, C. M., Mckenzie, M. S., Harhay, M. O., Gabler, N. B., Silva, T., Arnold, R., Angus, D. C., & Bryce, C. (2013). Default Options In Advance Directives Influence How Patients Set Goals For End-Of-Life Care. *Health Affairs (Project Hope)*, *32*(2), 408–417. https://doi.org/10.1377/hlthaff.2012.0895

Harvey, M. (2001). Deliberation and Natural Slavery. *Social Theory and Practice*, *27*(1), 41–64. https://doi.org/10.5840/soctheorpract200127120

Hauser, D. J., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J., & Litman, L. (2022). Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01999-x

Hausman, D. M., & Welch, B. (2010). Debate: To Nudge or Not to Nudge*. *Journal of Political Philosophy*, *18*(1), 123–136. https://doi.org/10.1111/j.1467-9760.2009.00351.x

Hershfield, H. E., Goldstein, D. G., Sharpe, W. F., Fox, J., Yeykelis, L., Carstensen, L. L., & Bailenson, J. N. (2011). Increasing Saving Behavior Through Age-Progressed Renderings of the Future Self. *Journal of Marketing Research*, *48*(SPL), S23–S37. https://doi.org/10.1509/jmkr.48.SPL.S23

Hertwig, R. (2017). When to consider boosting: Some rules for policy-makers. *Behavioural Public Policy*, *1*(2), 143–161. https://doi.org/10.1017/bpp.2016.14

Hilbig, B. E. (2009). Sad, thus true: Negativity bias in judgments of truth. *Journal of Experimental Social Psychology*, *45*(4), 983–986. https://doi.org/10.1016/j.jesp.2009.04.012

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied*, *5*, 205–221. https://doi.org/10.1037/1076-898X.5.2.205

Hoang, U., Feltz, S., Offer-Westort, T., & Feltz, A. (2023). Willingness to Consume Fewer Animal Products: A Latent Profile Analysis. *ANTHROZOÖS*, Advance online publication. https://doi.org/10.1080/08927936.2023.2204640

Hogarth, R. M., & Kunreuther, H. (1989). Risk, ambiguity, and insurance. *Journal of Risk and Uncertainty*, *2*(1), 5–35. https://doi.org/10.1007/BF00055709

Holt, J. (2023). An Examination of Common Ethical Considerations Across Domains. *[Manuscript in Preparation]*.

Hurlimann, A., & Dolnicar, S. (2010). When public opposition defeats alternative water projects – The case of Toowoomba Australia. *Water Research*, *44*(1), 287–297. https://doi.org/10.1016/j.watres.2009.09.020

Hursthouse, R., & Pettigrove, G. (2022). Virtue Ethics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2022/entries/ethics-virtue/

Isaac, M. S., & Brough, A. R. (2014). Judging a Part by the Size of Its Whole: The Category Size Bias in Probability Judgments. *Journal of Consumer Research*, *41*(2), 310–325. https://doi.org/10.1086/676126

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, *3*(02), 159–186. https://doi.org/10.1017/bpp.2018.43

Janiszewski, C. (1993). Preattentive Mere Exposure Effects. *Journal of Consumer Research*, *20*(3), 376–392. https://doi.org/10.1086/209356

Jenkins, M. M., & Youngstrom, E. A. (2016). A randomized controlled trial of cognitive debiasing improves assessment and treatment selection for pediatric bipolar disorder. *Journal of Consulting and Clinical Psychology*, *84*, 323–333. https://doi.org/10.1037/ccp0000070

Johnson, E. J., & Goldstein, D. G. (2013). Decisions by default. In E. [Ed Shafir (Ed.), *The behavioral foundations of public policy* (pp. 417–427, Chapter xvii, 511 Pages). Princeton University Press (Princeton, NJ, US).

Johnson, R., & Cureton, A. (2022). Kant's Moral Philosophy. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2022/entries/kant-moral/

Joseph, H. (1933). Some Problems in Ethics. *Revue de Métaphysique et de Morale*, *40*(1), 11–12.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52–54. https://doi.org/10.5395/rde.2013.38.1.52

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling, Fourth Edition*. Guilford Publications.

Krijnen, J. M. T., Tannenbaum, D., & Fox, C. R. (2017). Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy*, *3*(2), i–18. https://doi.org/10.1353/bsp.2017.0010

Kriss, P. H., Loewenstein, G., Wang, X., & Weber, R. A. (2011). Behind the veil of ignorance: Self-serving bias in climate change negotiations. *Judgment and Decision Making*, *6*(7), 602–615. https://doi.org/10.1017/S1930297500002643

Krusemark, E. A., Keith Campbell, W., & Clementz, B. A. (2008). Attributions, deception, and event related potentials: An investigation of the self-serving bias. *Psychophysiology*, *45*(4), 511–515. https://doi.org/10.1111/j.1469-8986.2008.00659.x

Kymlicka, W. (2002). *Contemporary Political Philosophy: An Introduction*. Oxford University Press.

Lambe, K. A., O'Reilly, G., Kelly, B. D., & Curristan, S. (2016). Dual-process cognitive interventions to enhance diagnostic reasoning: A systematic review. *BMJ Quality & Safety*, *25*(10), 808–820. https://doi.org/10.1136/bmjqs-2015-004417

Lee, J. A., Soutar, G. N., Daly, T. M., & Louviere, J. J. (2011). Schwartz Values Clusters in the United States and China. *Journal of Cross-Cultural Psychology*, *42*(2), 234–252. https://doi.org/10.1177/0022022110396867

Lee, S., & Feeley, T. H. (2016). The identifiable victim effect: A meta-analytic review. *Social Influence*, *11*(3), 199–215. https://doi.org/10.1080/15534510.2016.1216891

Leong, C., & Lebel, L. (2020). Can conformity overcome the yuck factor? Explaining the choice for recycled drinking water. *Journal of Cleaner Production*, *242*, 118196. https://doi.org/10.1016/j.jclepro.2019.118196

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madowand, & H. B. Mann (Eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

Lieberman, J. D. (1999). Terror Management, Illusory Correlation, and Perceptions of Minority Groups. *Basic and Applied Social Psychology*, *21*(1), 13–23. https://doi.org/10.1207/s15324834basp2101_2

Lim, T.-S., & Loh, W.-Y. (1996). A comparison of tests of equality of variances. *Computational Statistics & Data Analysis*, *22*(3), 287–301. https://doi.org/10.1016/0167-9473(95)00054-2

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767–778. https://doi.org/10.1093/biomet/88.3.767

Madrian, B. C., & Shea, D. F. (2001). The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. *The Quarterly Journal of Economics*, *116*(4), 1149–1187. https://doi.org/10.1162/003355301753265543

Mahmoud-Elhaj, D., Tanner, B., Sabatini, D., & Feltz, A. (2020). Measuring Objective Knowledge of Potable Recycled Water. *Journal of Community Psychology*, *48*(6), 2033–2052. https://doi.org/10.1002/jcop.22402

Mansbridge, J. (2015). A Minimalist Definition of Deliberation. In P. Heller & V. Rao, *Deliberation and Development: Rethinking the Role of Voice and Collective Action in Unequal Societies* (pp. 27–50). World Bank Publications.

Maoz, I., Ward, A., Katz, M., & Ross, L. (2002). Reactive Devaluation of an “Israeli” vs. “Palestinian” Peace Proposal. *Journal of Conflict Resolution*, *46*(4), 515–546. https://doi.org/10.1177/0022002702046004003

Mather, M., Mazar, N., Gorlick, M. A., Lighthall, N. R., Burgeno, J., Schoeke, A., & Ariely, D. (2012). Risk preferences and aging: The "certainty effect" in older adults' decision making. *Psychology and Aging*, *27*, 801–816. https://doi.org/10.1037/a0030174

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994). Illusory correlation in the perception of groups: An extension of the distinctiveness-based account. *Journal of Personality and Social Psychology*, *67*, 414–429. https://doi.org/10.1037/0022-3514.67.3.414

McCracken, J., Martin, W., & Shaw, B. (1998). Virtue Ethics and The Parable of the Sadhu. *Journal of Business Ethics*, *17*(1), 25–38. https://doi.org/10.1023/A:1017912300819

McNaughton, D. (1996). An Unconnected Heap of Duties? *The Philosophical Quarterly (1950-)*, *46*(185), 433–447. https://doi.org/10.2307/2956354

Mele, A. R. (2001). Compatibilist Autonomy and Autonomous Action. In A. R. Mele, *Autonomous Agents: From Self-control to Autonomy*. Oxford University Press.

Melnyk, V., Carrillat, F. A., & Melnyk, V. (2022). The Influence of Social Norms on Consumer Behavior: A Meta-Analysis. *Journal of Marketing*, *86*(3), 98–120. https://doi.org/10.1177/00222429211029199

Midway, S., Robertson, M., Flinn, S., & Kaller, M. (2020). Comparing multiple comparisons: Practical guidance for choosing the best multiple comparisons test. *PeerJ*, *8*, e10387. https://doi.org/10.7717/peerj.10387

Mill, J. S. (1863). *Utilitarianism*. PARKER, SON, AND BOURN, WEST STRAND.

Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial Bias in Mock Juror Decision-Making: A Meta-Analytic Review of Defendant Treatment. *Law and Human Behavior*, *29*, 621–637. https://doi.org/10.1007/s10979-005-8122-9

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing Decisions: Improved Decision Making With a Single Training Intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129–140. https://doi.org/10.1177/2372732215600886

Neumann, N., & Böckenholt, U. (2014). A Meta-analysis of Loss Aversion in Product Choice. *Journal of Retailing*, *90*(2), 182–197. https://doi.org/10.1016/j.jretai.2014.02.002

Nisbet, M. C. (2005). The Competition for Worldviews: Values, Information, and Public Support for Stem Cell Research. *International Journal of Public Opinion Research*, *17*(1), 90–112. https://doi.org/10.1093/ijpor/edh058

Nisbet, M. C., & Goidel, R. K. (2007). Understanding citizen perceptions of science controversy: Bridging the ethnographic—survey research divide. *Public Understanding of Science*, *16*(4), 421–440. https://doi.org/10.1177/0963662506065558

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, *22*(3), 453–460. https://doi.org/10.1016/j.jcps.2011.08.002

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440–461. https://doi.org/10.1037/tps0000176

OECD. (2017). *Behavioral Insights and Public Policy: Lessons from Around the World*. OECD Publishing.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Perfecto, H., Nelson, L. D., & Moore, D. A. (2018). The category size bias: A mere misunderstanding. *Judgment and Decision Making*, *13*(2), 170–184. https://doi.org/10.1017/S1930297500007087

Peters, E. (2012). Beyond Comprehension: The Role of Numeracy in Judgments and Decisions—Ellen Peters, 2012. *Current Directions in Psychological Science*, *21*(1), 31–35. https://doi.org/10.1177/0963721411429960

Po, M., Kaercher, J. D., & Nancarrow, B. E. (2003). *Literature Review of Factors Influencing Public Perceptions of Water Reuse* (Technical Report No. 54/03). CSIRO Land and Water.

Railton, P. (1984). Alienation, Consequentialism, and the Demands of Morality. *Philosophy & Public Affairs*, *13*(2), 134–171.

Ramasubramanian, M. (2020). Individual Differences and Risk Perception: Numeracy Predicts Differences in General and Specific Risk Perceptions. *The University of Oklahoma - Theses*. https://shareok.org/handle/11244/324406

Ramasubramanian, M. (2022). Toward an Integrated Assessment of Risk Perceptions: Development and Testing of The Berlin Risk Perception Inventory. *The University of Oklahoma - Dissertations*. https://hdl.handle.net/11244/336286

Raphael, D. D. (1994). *Moral Philosophy*. Oxford University Press.

Rettberg, K. (2020). Socioeconomic Declinism and Right-Wing Populist Support. *Res Publica - Journal of Undergraduate Research*, *25*(1). https://digitalcommons.iwu.edu/respublica/vol25/iss1/17

Rosenberg, S. W. (2022). Citizen competence and the psychology of deliberation. In *Citizen competence and the psychology of deliberation* (pp. 98–117). Edinburgh University Press. https://doi.org/10.1515/9780748643509-008

Roth, S., Robbert, T., & Straus, L. (2015). On the sunk-cost effect in economic decision-making: A meta-analytic review. *Business Research*, *8*(1), 99–138. https://doi.org/10.1007/s40685-014-0014-8

Rozin, P., Haddad, B., Nemeroff, C., & Slovic, P. (2015). *Psychological aspects of the rejection of recycled water: Contamination, purification and disgust*. https://scholarsbank.uoregon.edu/xmlui/handle/1794/19440

Sarstedt, M., Neubert, D., & Barth, K. (2017). The IKEA Effect. A Conceptual Replication. *Journal of Marketing Behavior*, *2*(4), 307–312.

Scheffler, S. (1988). *Consequentialism and Its Critics*. Oxford University Press.

Schramme, T. (2021). Capable deliberators: Towards inclusion of minority minds in discourse practices. *Critical Review of International Social and Political Philosophy*, *0*(0), 1–24. https://doi.org/10.1080/13698230.2021.2020550

Schultz, T., & Fielding, K. (2014). The common in-group identity model enhances communication about recycled water. *Journal of Environmental Psychology*, *40*, 296–305. https://doi.org/10.1016/j.jenvp.2014.07.006

Schulz, E., Cokely, E. T., & Feltz, A. (2011). Persistent bias in expert judgments about free will and moral responsibility: A test of the expertise defense. *Consciousness and Cognition*, *20*(4), 1722–1731. https://doi.org/10.1016/j.concog.2011.04.007

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The Role of Numeracy in Understanding the Benefit of Screening Mammography. *Annals of Internal Medicine*, *127*(11), 966–972. https://doi.org/10.7326/0003-4819-127-11-199712010-00003

Schwartz, S. H. (2012). An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, *2*(1). https://doi.org/10.9707/2307-0919.1116

Schwartz, S. H., & Bardi, A. (2001). Value Hierarchies Across Cultures: Taking a Similarities Perspective. *Journal of Cross-Cultural Psychology*, *32*(3), 268–290. https://doi.org/10.1177/0022022101032003002

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*, 663–688. https://doi.org/10.1037/a0029393

Selinger, E., & Whyte, K. (2011). Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture: Practice and Ethics of Choice Architecture. *Sociology Compass*, *5*(10), 923–935. https://doi.org/10.1111/j.1751-9020.2011.00413.x

Sellier, A.-L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing Training Improves Decision Making in the Field. *Psychological Science*, *30*(9), 1371–1379. https://doi.org/10.1177/0956797619861429

Shanahan, K. J., & Hyman, M. R. (2003). The Development of a Virtue Ethics Scale. *Journal of Business Ethics*, *42*(2), 197–208. https://doi.org/10.1023/A:1021914218659

Sidgwick, H. (1874). *The Methods of Ethics*. Macmillan.

Sinnott-Armstrong, W. (2022). Consequentialism. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Winter 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2022/entries/consequentialism/

Slote, M. (1995). Agent-Based Virtue Ethics. In R. Shafer-Landau (Ed.), *Ethical Theory: An Anthology* (2nd ed.). Wiley-Blackwell.

Soster, R. L., Gershoff, A. D., & Bearden, W. O. (2013). The Bottom Dollar Effect: How Resource Availability Influences Perceived Value and Satisfaction. In S. Botti & A. Labroo (Eds.), *NA - Advances in Consumer Research* (Vol. 41). The Association for Consumer Research.

Soster, R. L., Gershoff, A. D., & Bearden, W. O. (2014). The Bottom Dollar Effect: The Influence of Spending to Zero on Pain of Payment and Satisfaction. *Journal of Consumer Research*, *41*(3), 656–677. https://doi.org/10.1086/677223

Stark, C. A. (1997). Decision Procedures, Standards of Rightness and Impartiality. *Noûs*, *31*(4), 478–495. https://www.jstor.org/stable/2216169

Steenvoorden, E., & Harteveld, E. (2018). The appeal of nostalgia: The influence of societal pessimism on support for populist radical right parties. *West European Politics*, *41*(1), 28–52. https://doi.org/10.1080/01402382.2017.1334138

Steininger, D. M., Lorch, M., & Veit, D. J. (2014). The Bandwagon Effect in Digital Environments: An Experimental Study on Kickstarter.com. In D. Kundisch (Ed.), *MKWI 2014—Multikonferenz Wirtschaftsinformatik* (pp. 546–556). Univ. https://www.researchgate.net/publication/260088116_The_Bandwagon_Effect_in_Digital_Environments_An_Experimental_Study_on_Kickstartercom

Stich, S., & Tobia, K. (2015). *Experimental Philosophy's Challenge to the "Great Tradition"* (SSRN Scholarly Paper ID 2543546). Social Science Research Network. https://papers.ssrn.com/abstract=2543546

Sturgis, P., & Allum, N. (2004). Science in Society: Re-Evaluating the Deficit Model of Public Attitudes. *Public Understanding of Science*, *13*(1), 55–74. https://doi.org/10.1177/0963662504042690

Sunstein, C. R. (2015a). Nudges Do Not Undermine Human Agency: A Note. *Journal of Consumer Policy*, *38*, 207–210. https://doi.org/10.2139/ssrn.2594758

Sunstein, C. R. (2015b). The Ethics of Nudging. *Yale Journal on Regulation*, *32*, 413–450. https://doi.org/10.2139/ssrn.2526341

Tanner, B. (2021). Cascading Effects of Nudging and Education on Common Goods: An Experimental Comparison of Potable Water Recycling Interventions. *The University of Oklahoma - Theses*. https://shareok.org/handle/11244/330166

Tanner, B., Cokely, E. T., & Feltz, A. (2023). Strategies for Increasing First and Second-Order Recycled Water Acceptance: Implications for Ethical Consumer Decision Support. *[Manuscript in Preparation]*.

Tanner, B., & Feltz, A. (2022). Comparing effects of default nudges and informing on recycled water decisions. *Journal of Experimental Psychology: Applied*, *28*(2), 399–411. https://doi.org/10.1037/xap0000406

Tanner, B., McDonald, P., Mahmoud-Elhaj, D., Sabatini, D., Chamberlain, J. F., Nijhawan, A., Rainbolt-Forbes, E., & Feltz, A. (2023). Not All Information is Informative: An Exploration of Educational Content on Potable Recycled Water Knowledge and Acceptance. *Manuscript in Preparation*.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Trout, J. D. (2005). Paternalism and Cognitive Bias. *Law and Philosophy*, *24*(4), 393–434.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Uhlmann, V., & Head, B. (2011). *Water Recycling: Recent History of Local Government Initiatives in South East Queensland* (Technical No. 45). Urban Water Security Research Alliance.

Upadhyay, U. D., Dworkin, S. L., Weitz, T. A., & Foster, D. G. (2014). Development and Validation of a Reproductive Autonomy Scale. *Studies in Family Planning*, *45*(1), 19–41. https://doi.org/10.1111/j.1728-4465.2014.00374.x

Utts, J. M. (2014). *Seeing Through Statistics* (4th ed.). Cengage Learning.

Vedam, S., Stoll, K., Martin, K., Rubashkin, N., Partridge, S., Thordarson, D., Jolicoeur, G., & Council, the C. C. in B. S. (2017). The Mother's Autonomy in Decision Making (MADM) scale: Patient-led development and psychometric testing of a new instrument to evaluate experience of maternity care. *PLOS ONE*, *12*(2), e0171804. https://doi.org/10.1371/journal.pone.0171804

Venema, T. A. G., Kroese, F. M., & Ridder, D. T. D. D. (2018). I'm still standing: A longitudinal study on the effect of a default nudge. *Psychology & Health*, *33*(5), 669–681. https://doi.org/10.1080/08870446.2017.1385786

Vuchinich, R. E., & Simpson, C. A. (1998). Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and Clinical Psychopharmacology*, *6*, 292–305. https://doi.org/10.1037/1064-1297.6.3.292

Weinstein, N., Przybylski, A. K., & Ryan, R. M. (2012). The index of autonomous functioning: Development of a scale of human autonomy. *Journal of Research in Personality*, *46*(4), 397–413. https://doi.org/10.1016/j.jrp.2012.03.007

Weißmüller, K. S., Bouwman, R., & Vogel, R. (2021). Satisficing or maximizing in public–private partnerships? A laboratory experiment on strategic bargaining. *Public Management Review*, *0*(0), 1–27. https://doi.org/10.1080/14719037.2021.2013072

Wester, J., Timpano, K. R., Çek, D., & Broad, K. (2016). The psychology of recycled water: Factors predicting disgust and willingness to use. *Water Resources Research*, *52*(4), 3212–3226. https://doi.org/10.1002/2015WR018340

Wichmann, S. S. (2011). Self-Determination Theory: The Importance of Autonomy to Well-Being Across Cultures. *Journal of Humanistic Counseling*, *50*(1), 16–26. https://doi.org/10.1002/j.2161-1939.2011.tb00103.x

Williams, O. F., & Murphy, P. E. (1990). The Ethics of Virtue: A Moral Theory for Marketing. *Journal of Macromarketing*, *10*(1), 19–29. https://doi.org/10.1177/027614679001000103

Yan, H., & Yates, J. F. (2019). Improving acceptability of nudges: Learning from attitudes towards opt-in and opt-out policies. *Judgment and Decision Making*, *14*(1), 26–39. https://www.sas.upenn.edu/~baron/journal/18/181018/jdm181018.pdf

Yeung, S. K., Yay, T., & Feldman, G. (2022). Action and Inaction in Moral Judgments and Decisions: Meta-Analysis of Omission Bias Omission-Commission Asymmetries. *Personality and Social Psychology Bulletin*, *48*(10), 1499–1515. https://doi.org/10.1177/01461672211042315

Zelezny, L. C. (1999). Educational Interventions That Improve Environmental Behaviors: A Meta-Analysis. *The Journal of Environmental Education*, *31*(1), 5–14. https://doi.org/10.1080/00958969909598627

Figure 1

Framework for Skilled Decision-Making Proposed by Cokely et al. (2018)

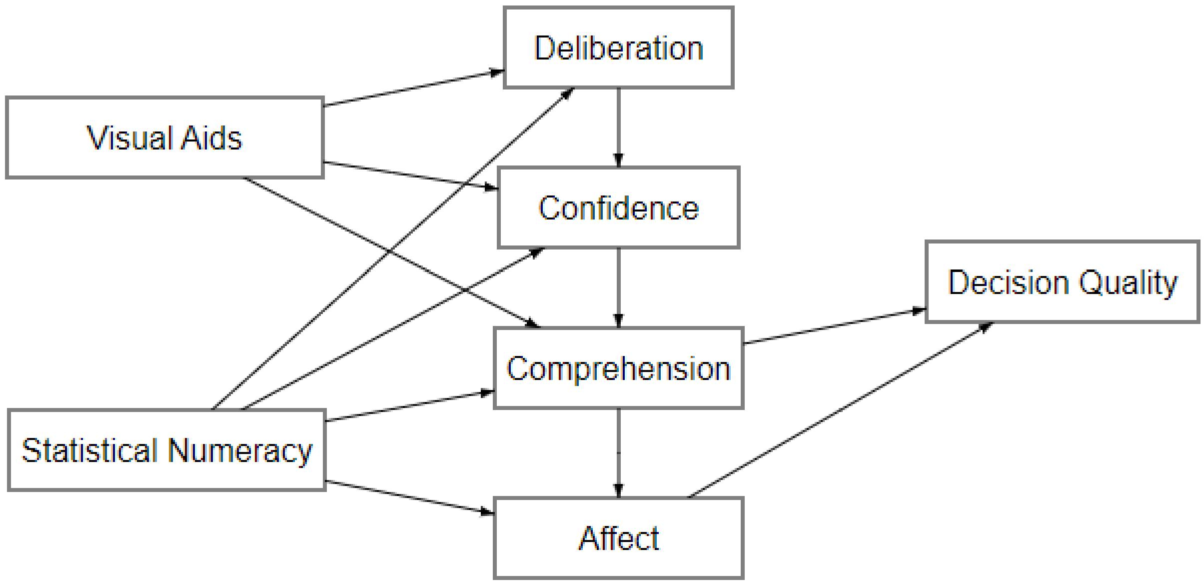


Figure 2

Flowchart of Experimental Procedure for Experiment 1b

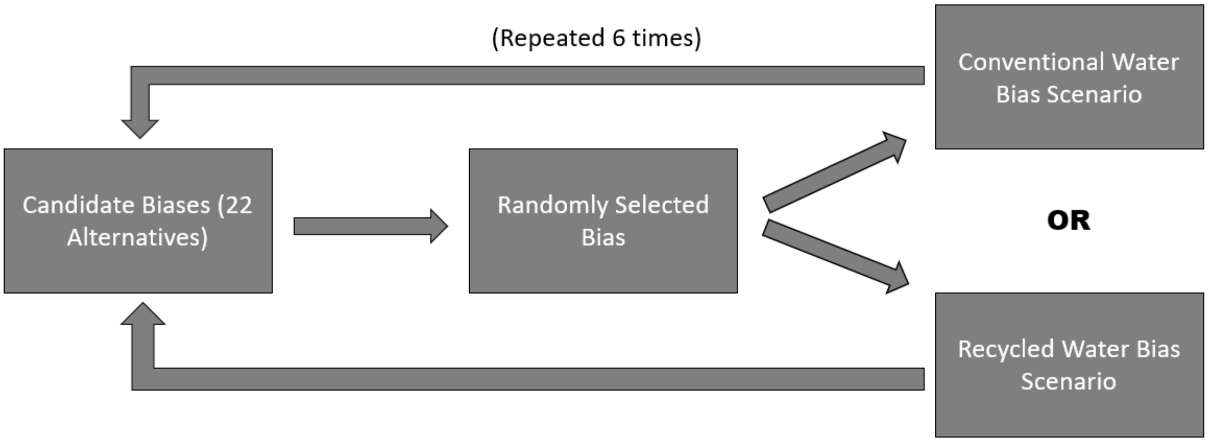
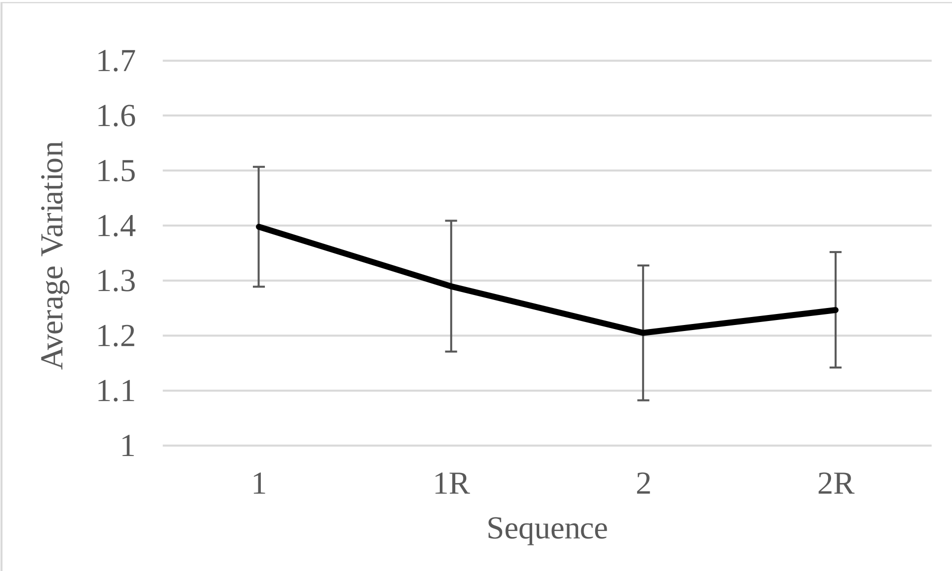


Figure 3

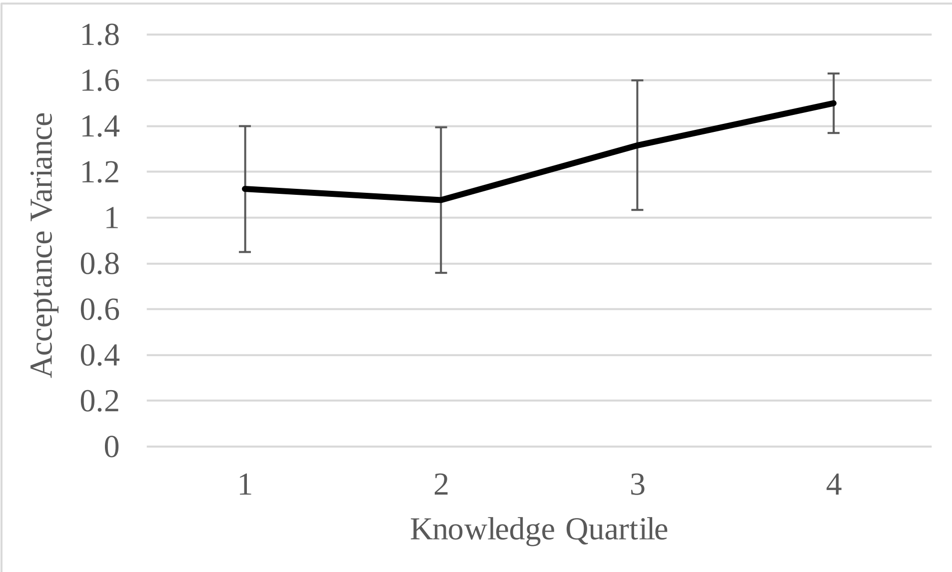
Experiment 1b Variance Split by Sequence



Note. Error bars represent 95% confidence intervals.

Figure 4

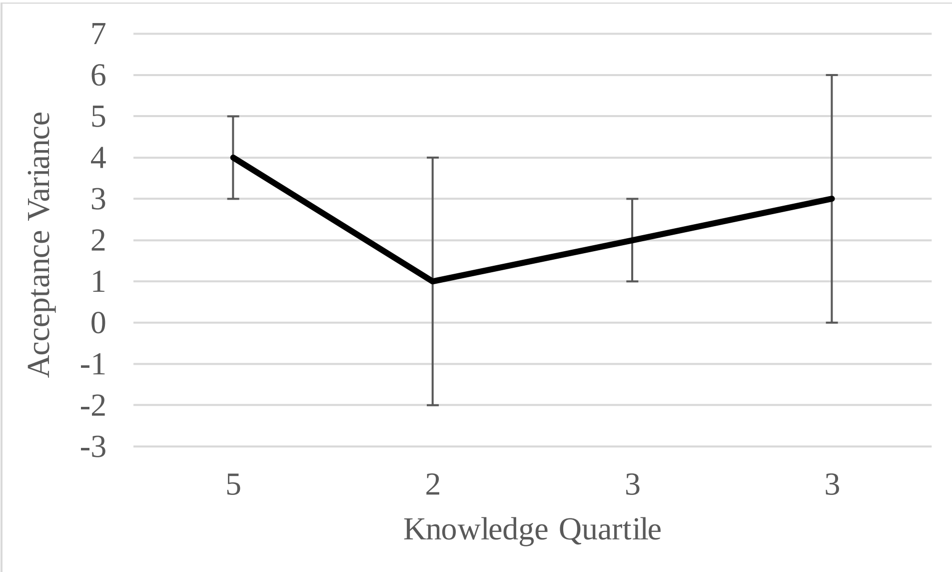
Experiment 1b Variance Split by Knowledge Quartile for Sequence 1R



Note. Error bars represent 95% confidence intervals.

Figure 5

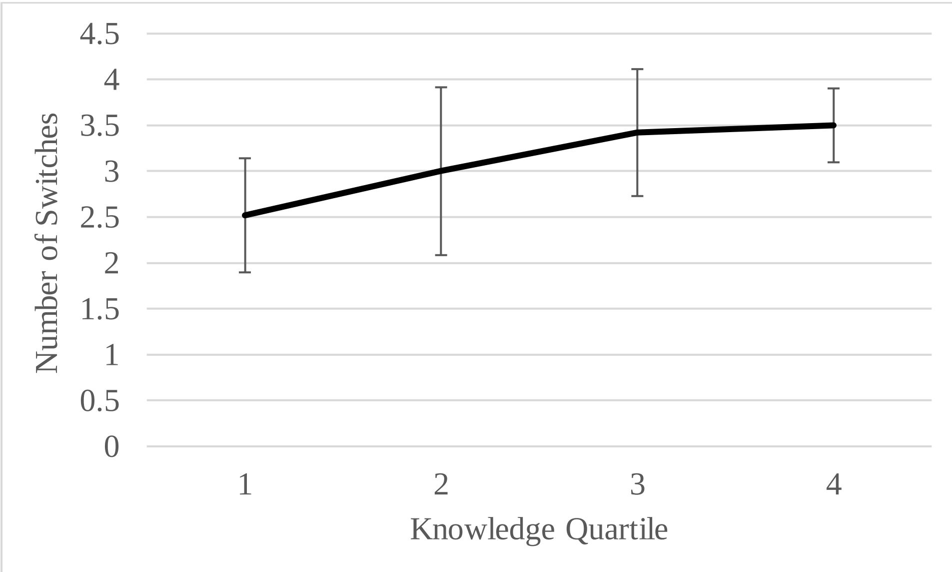
Experiment 1b Variance Split by Knowledge Quartile for Sequence 2R



Note. Error bars represent 95% confidence intervals.

Figure 6

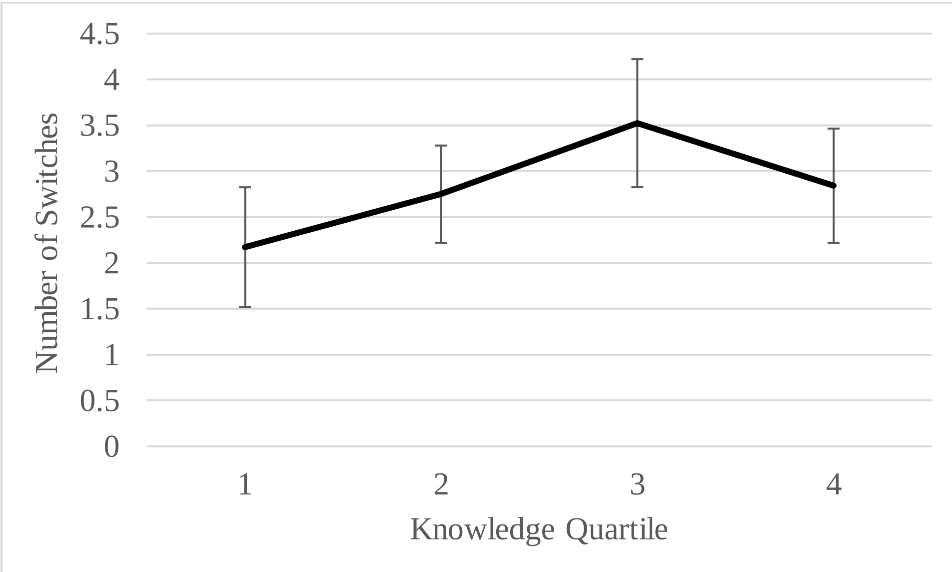
Experiment 1b Switches Split by Knowledge Quartile for Sequence 1R



Note. Error bars represent 95% confidence intervals.

Figure 7

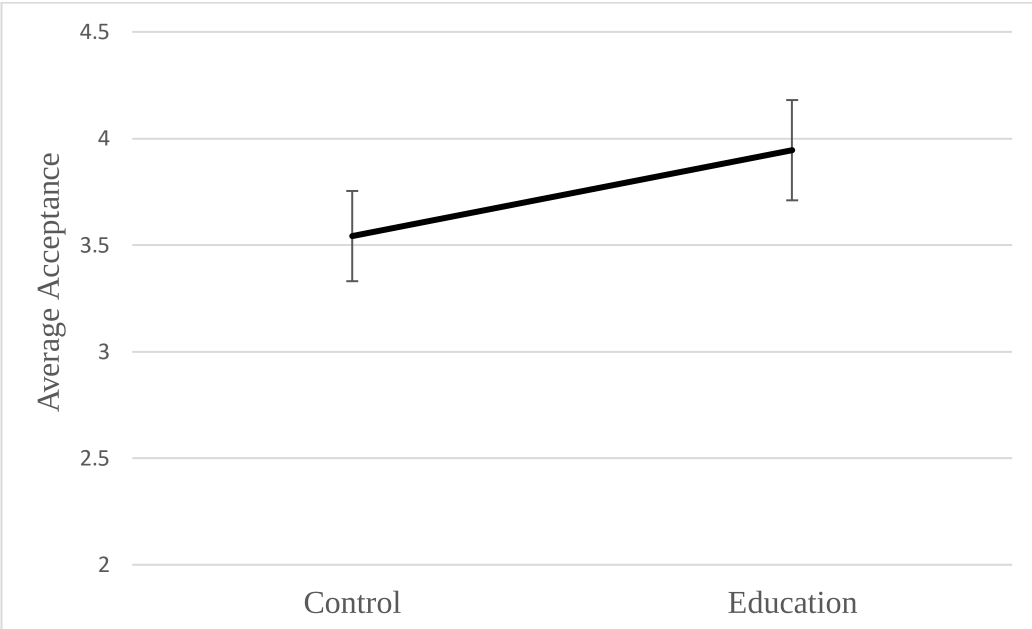
Experiment 1b Switches Split by Knowledge Quartile for Sequence 2



Note. Error bars represent 95% confidence intervals.

Figure 8

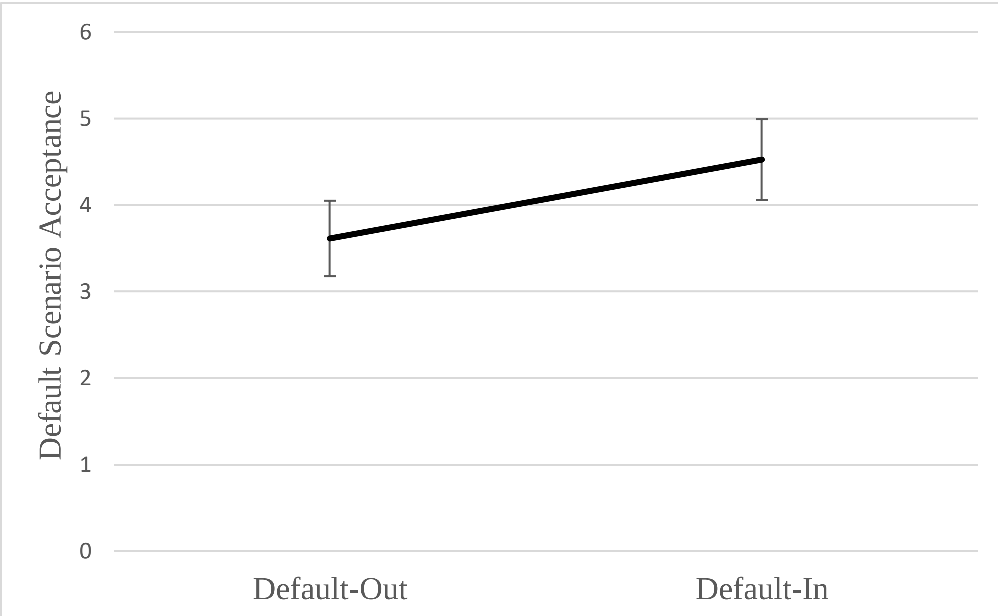
Overall Acceptance by Education Condition Across All Scenarios in Experiment 2



Note. Error bars represent 95% confidence intervals.

Figure 9

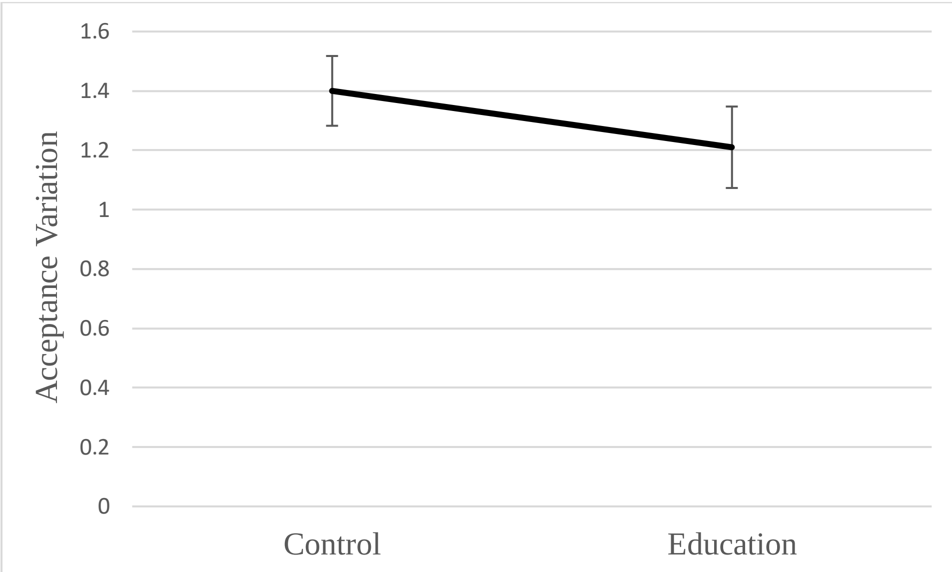
Acceptance of Recycled Water by Default Condition in Experiment 2



Note. Results are from the default scenario alone. Error bars represent 95% confidence intervals.

Figure 10

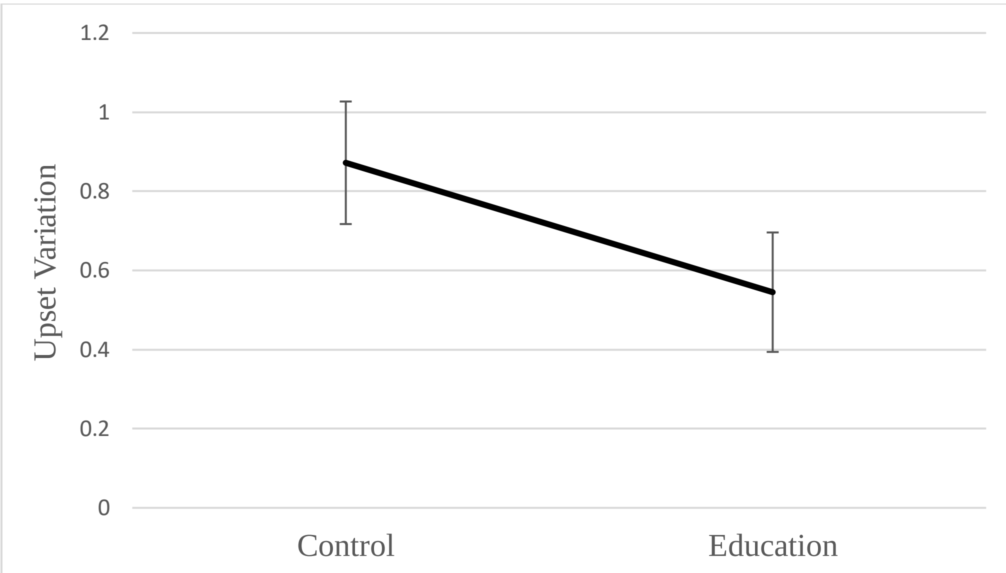
Average Variation in Acceptance of Recycled Water by Education Condition in Experiment 2



Note. Error bars represent 95% confidence intervals.

Figure 11

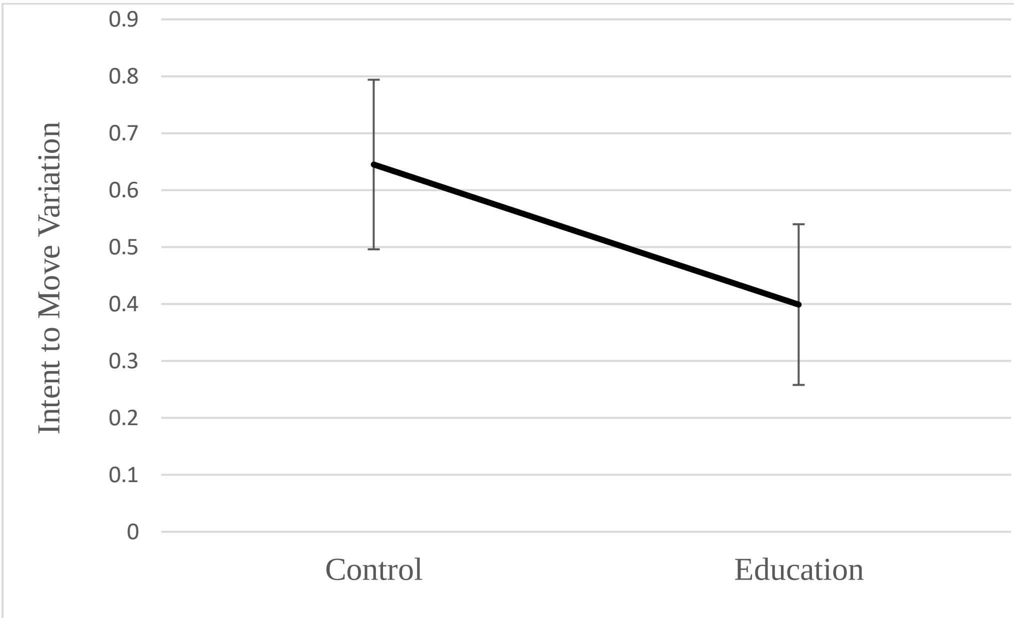
Average Variation in Being Upset About Recycled Water by Education Condition in Experiment 2



Note. Error bars represent 95% confidence intervals.

Figure 12

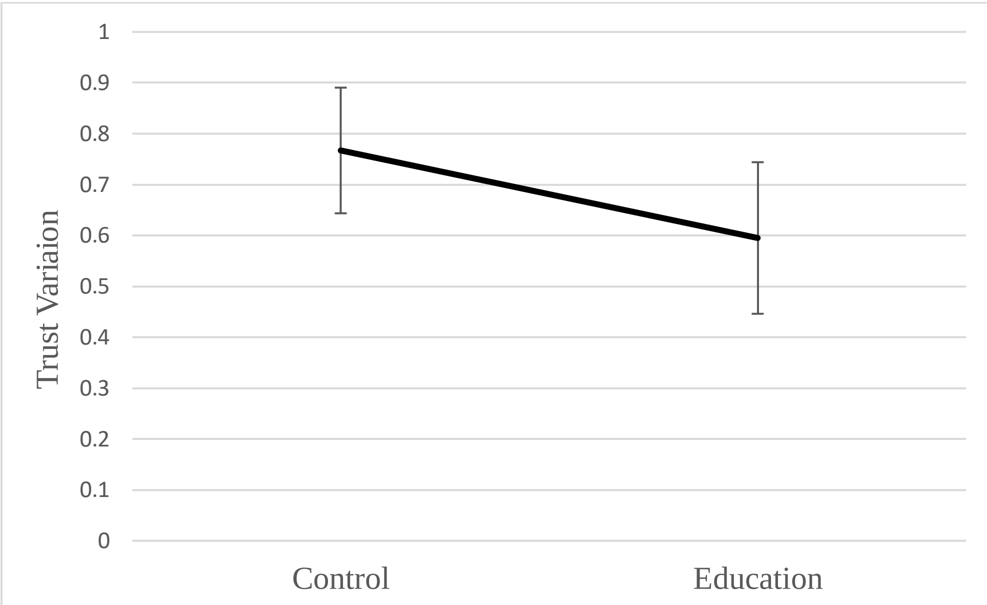
Average Variation in Intentions to Move by Education Condition in Experiment 2



Note. Error bars represent 95% confidence intervals.

Figure 13

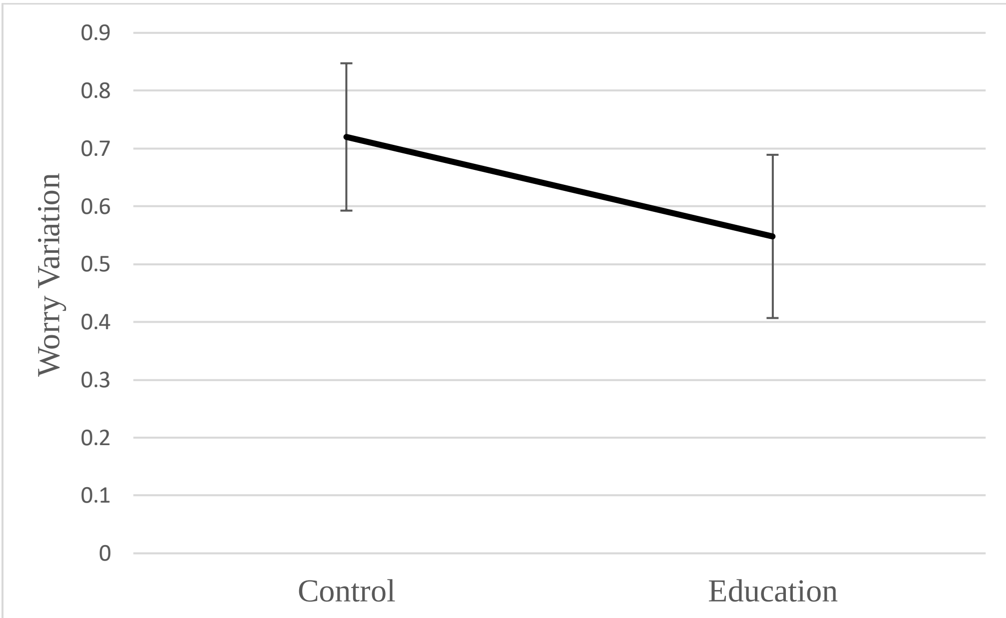
Average Variation in Trust About Recycled Water by Education Condition in Experiment 2



Note. Error bars represent 95% confidence intervals.

Figure 14

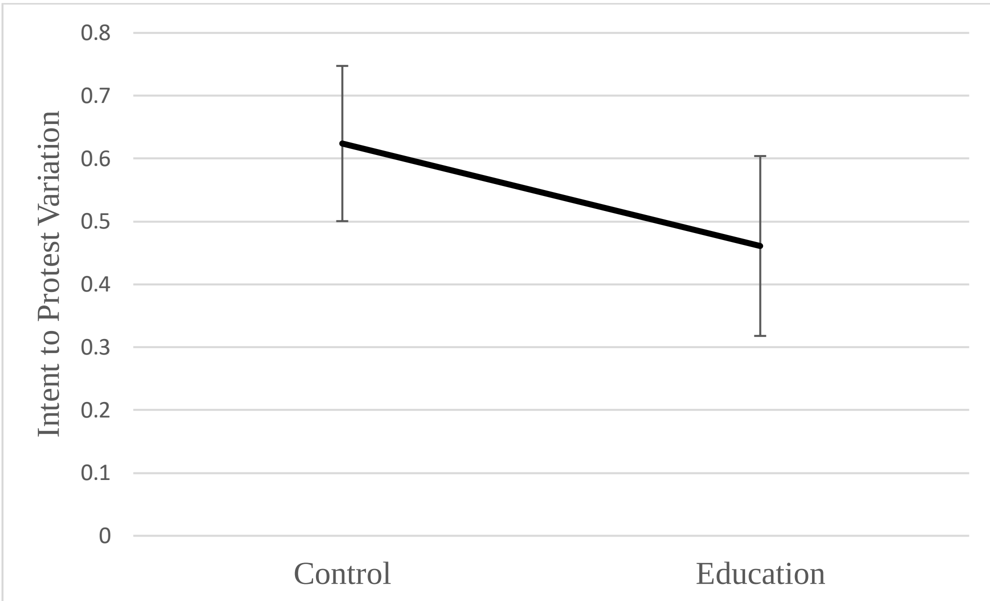
Average Variation in Worry About Recycled Water by Education Condition in Experiment 2



Note. Error bars represent 95% confidence intervals.

Figure 15

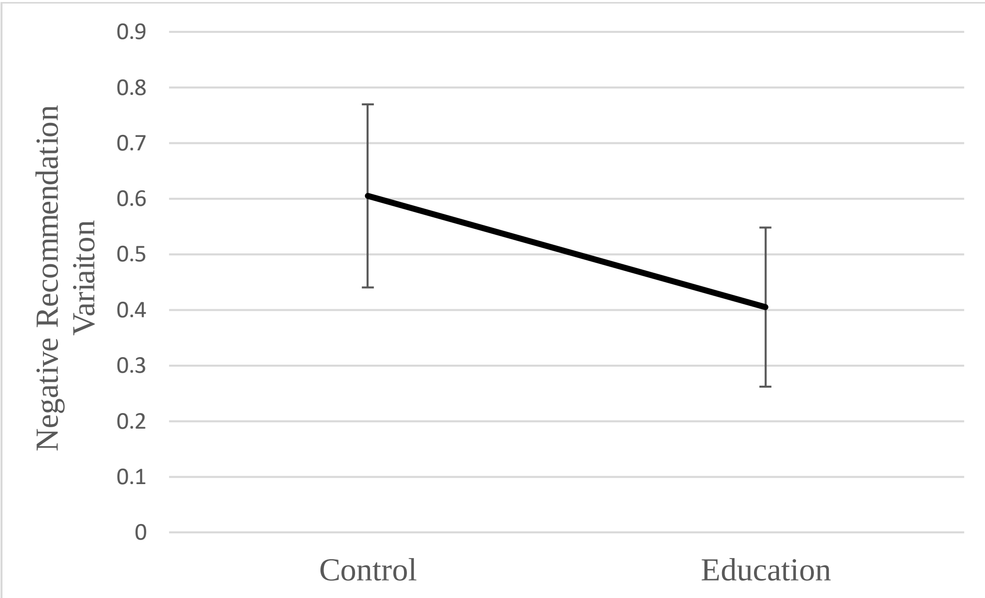
Average Variation in Intentions to Protest by Education Condition in Experiment 2



Note. Error bars represent 95% confidence intervals.

Figure 16

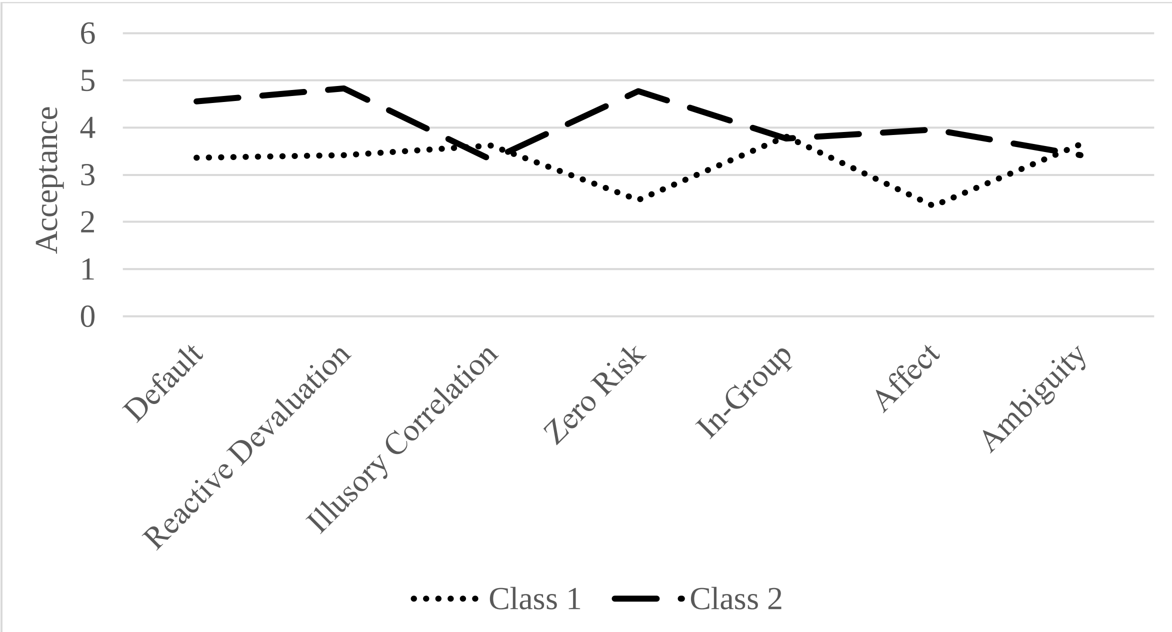
Average Variation in Negative Recommendations by Education Condition in Experiment 2



Note. Error bars represent 95% confidence intervals.

Figure 17

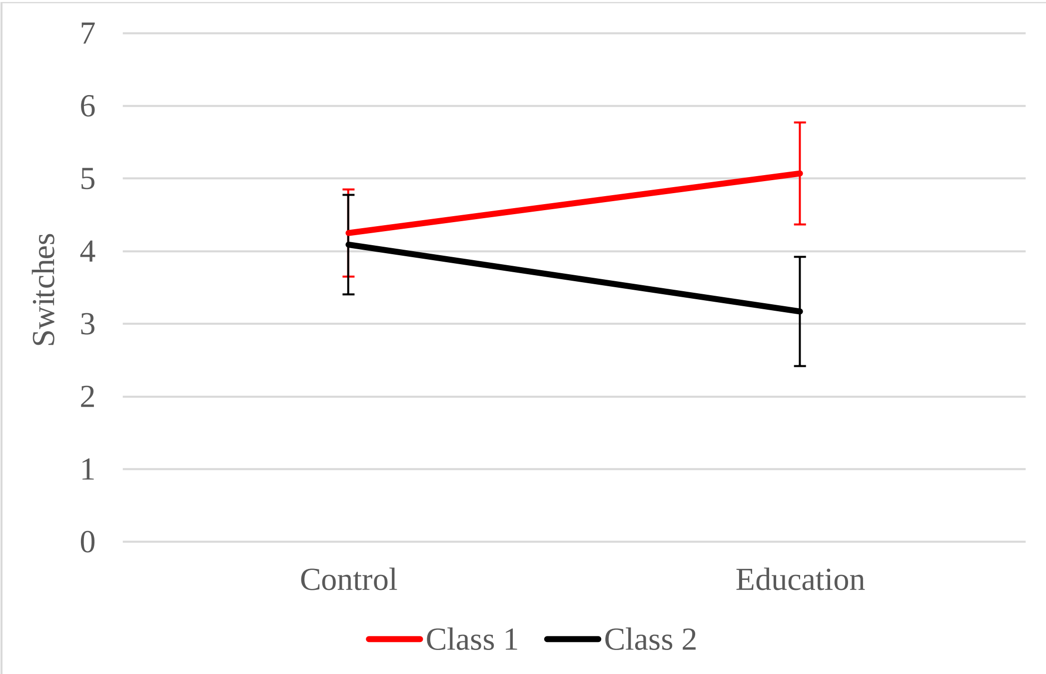
Classes Identified by Latent Profile Analysis in Experiment 2



Note. N = 35 for Class 1; N = 47 for Class 2.

Figure 18

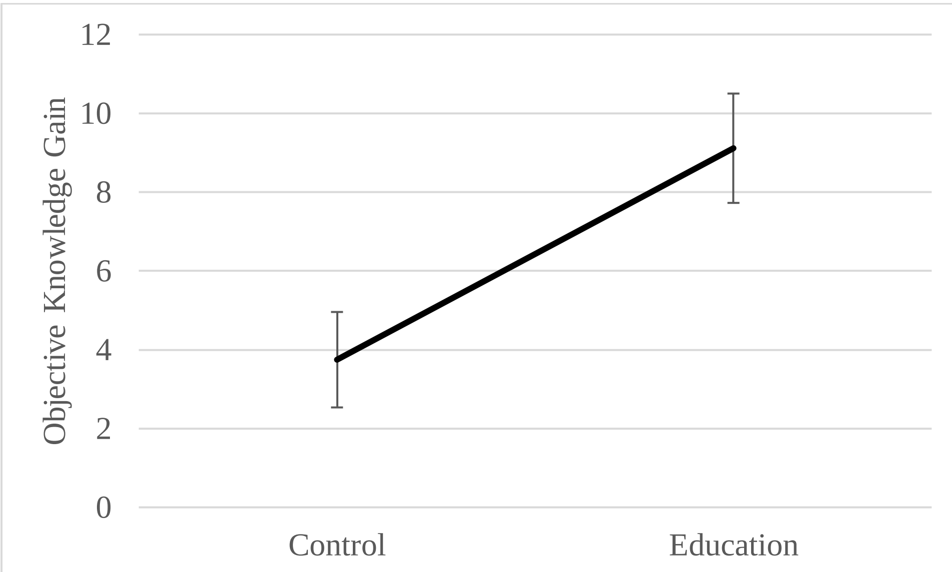
Number of Switches by Education Condition and Latent Class in Experiment 2



Note. Error bars represent 95% confidence intervals.

Figure 19

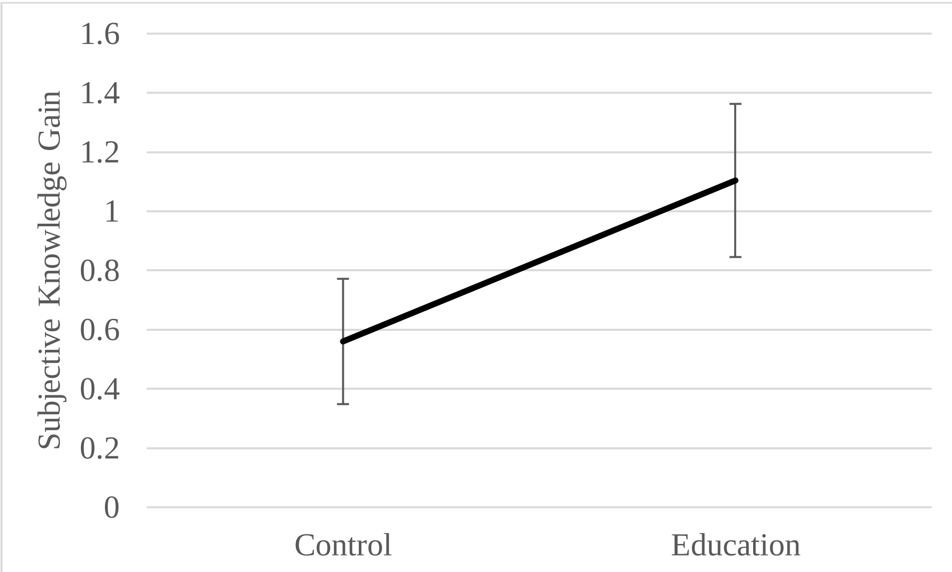
Objective Knowledge Gained by Condition in Experiment 3



Note. Error bars represent 95% confidence intervals.

Figure 20

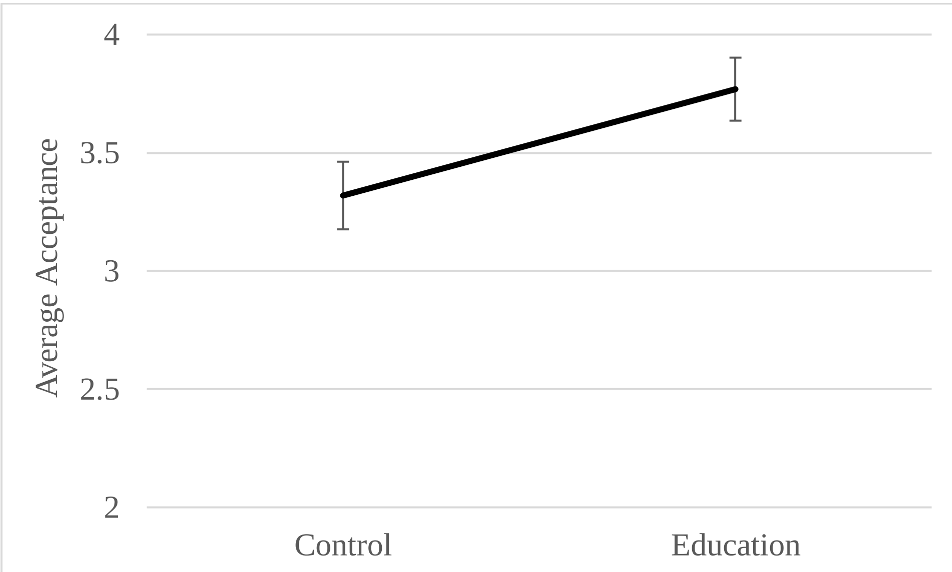
Subjective Knowledge Gained by Condition in Experiment 3



Note. Error bars represent 95% confidence intervals.

Figure 21

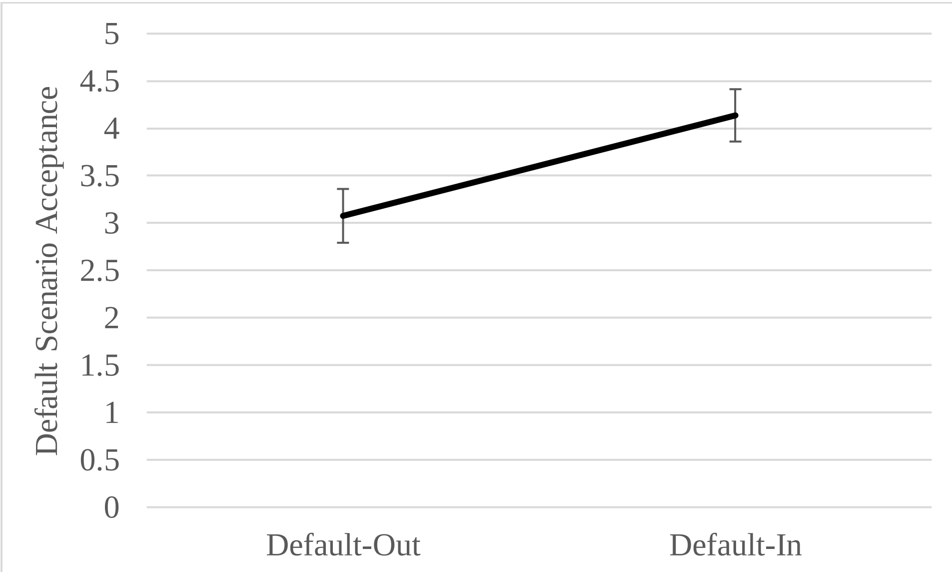
Overall Acceptance by Education Condition Across All Scenarios in Experiment 3



Note. Error bars represent 95% confidence intervals.

Figure 22

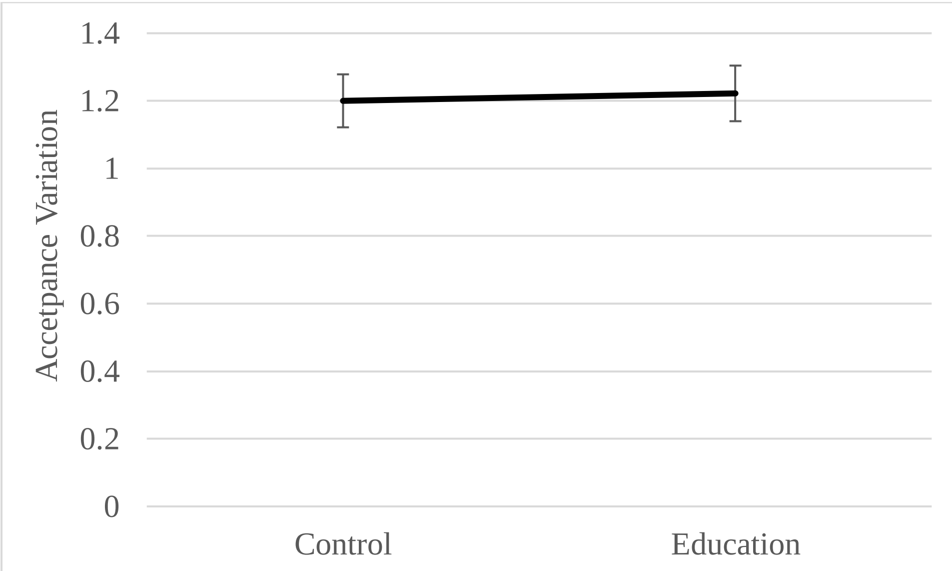
Acceptance of Recycled Water by Default Condition in Experiment 3



Note. Results are from the default scenario alone. Error bars represent 95% confidence intervals.

Figure 23

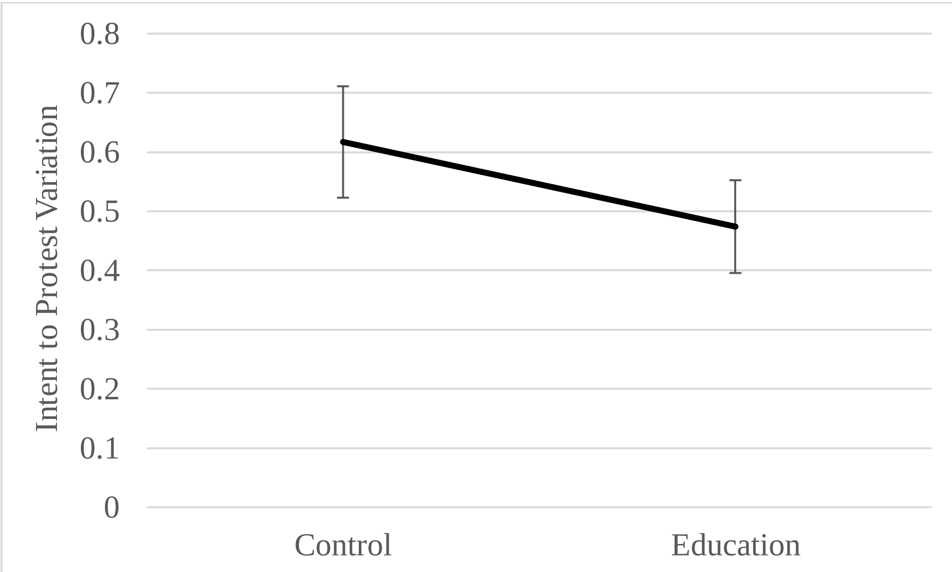
Average Variation in Acceptance of Recycled Water by Education Condition in Experiment 3



Note. Error bars represent 95% confidence intervals.

Figure 24

Average Variation in Intentions to Protest by Education Condition in Experiment 3



Note. Error bars represent 95% confidence intervals.

Figure 25

Classes Identified by Latent Profile Analysis in Experiment 3 (N = 72 for Class 1; N = 115 for Class 2)

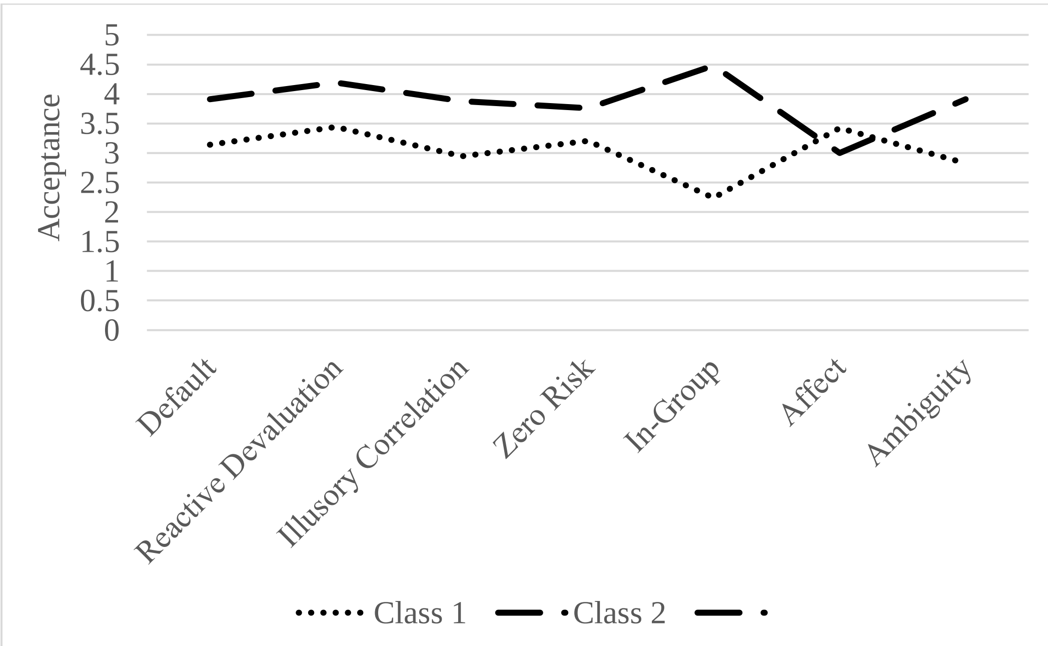
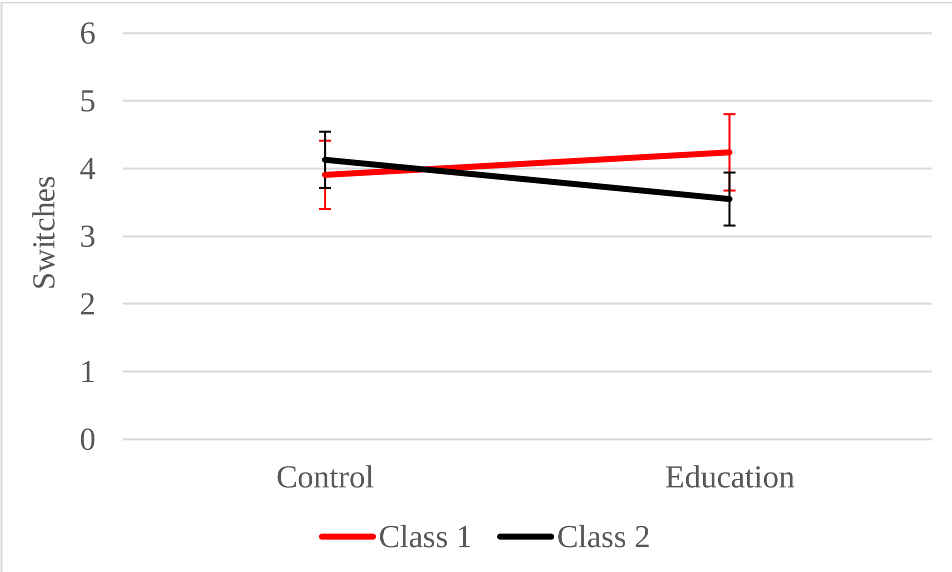


Figure 26

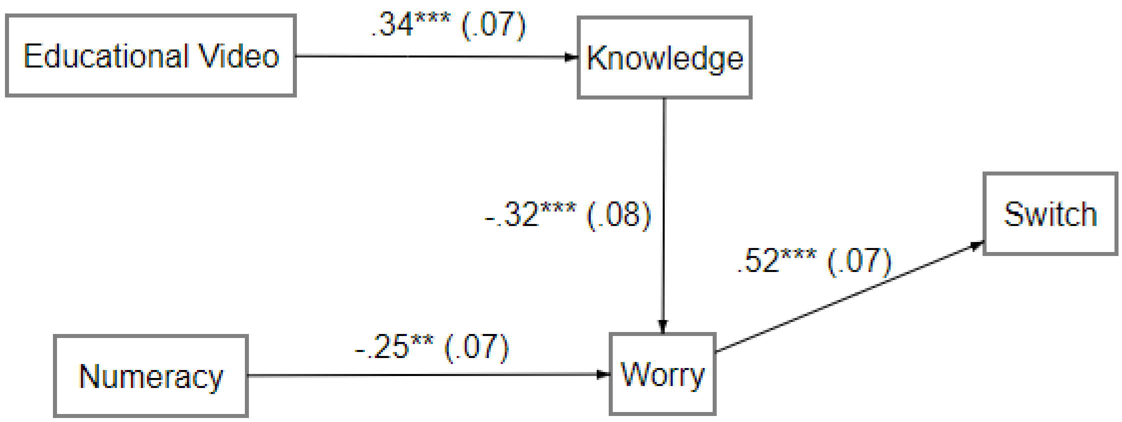
Number of Switches by Education Condition and Latent Class in Experiment 3



Note. Error bars represent 95% confidence intervals.

Figure 27

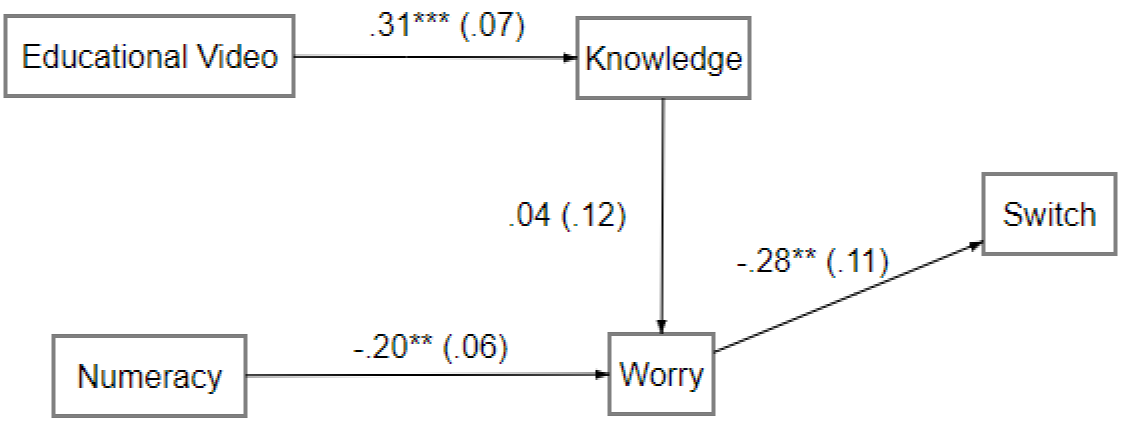
Skilled Decision Path Model from Experiment 3, Class 2 (Higher Acceptance)



Note. Path values are standardized, with standard errors in parentheses; * $p < .05$, ** $p < .01$, *** $p < .001$).

Figure 28

Skilled Decision Path Model from Experiment 3, Class 1 (Lower Acceptance)



Note. Path values are standardized, with standard errors in parentheses; * $p < .05$, ** $p < .01$, *** $p < .001$).

Table 1

List of all Biases Evaluated and Results from Experiment 1a

Bias	Description	Potential Application for Increasing Recycled Water Acceptance	Experiment 1a Results	Retained?
Noble Edge Effect	Tendency to evaluate products based on the producer's social responsibility, rather than the merit of product itself (Chernev & Blair, 2015)	Describe recycled water providers as being engaged in charitable causes and humanitarian work to a greater extent than other providers	t (212) = 6.99, p < .001, d = .96	Yes
Ambiguity Effect	Tendency to view things that are more ambiguous or about which one has less information as less favorable (Hogarth & Kunreuther, 1989)	Describe recycled water providers as being able to provide understandable information and answer questions more easily than other providers	t (208) = 6.55, p < .001, d = .91	Yes
Action Bias	Tendency to prefer taking action over inaction, even when there is no evidence that action will lead to better outcomes (Bar-Eli et al., 2007)	Frame recycled water as a method of taking action to solve various issues (e.g., water shortage, water quality problems)	t (206) = 5.76, p < .001, d = .80	Yes
Affect Heuristic	Tendency to allow strong emotional reactions (e.g., disgust) to dictate decisions, regardless of other factors (Rozin et al., 2015; Wester et al., 2016)	Reduce disgust to recycled water (e.g., by doing tours of water reclamation facilities) or increase disgust of conventional water (e.g., by showing unprocessed water)	t (215) = 5.37, p < .001, d = .73	Yes

Framing (Loss Aversion)	Tendency to weigh losses more heavily than gains, and thus to prefer information framed as avoiding losses over information framed as acquiring equivalent gains (Neumann & Böckenholt, 2014)	Describe recycled water as a way to avoid adverse circumstances (e.g., water shortage, water bills changing) as opposed to describing in terms of potential things to be gained	t (207) = 4.67, p < .001, d = .65	Yes
In-group Bias	Tendency to prefer and trust those belonging to the same group as ourselves, even when group membership is effectively meaningless (Mitchell et al., 2005)	Have a friend provide information about recycled water rather than a stranger	t (217) = 4.83, p < .001, d = .65	Yes
Gambler's Fallacy	Tendency to believe that previous instances of a random event occurring influences the probability of the event occurring again in the future, even when occurrences are independent of each other (Croson & Sundali, 2005)	Present information about clusters of people getting sick when drinking conventional water	t (210) = 3.49, p < .001, d = .48	Yes
Zero-risk Bias (Certainty Effect)	Preference for situations in which absolute certainty can be guaranteed, even if a riskier alternative would potentially offer greater benefits (Mather et al., 2012)	Present information about recycled water framed as a guaranteed (e.g., 100% chance that a water bill will rise by X amount, as opposed to a 50/50 chance that it might)	t (207) = 3.44, p < .001, d = .48	Yes

Bounded Rationality (Satisficing)	Tendency to settle for “good enough” options in situations where additional effort (e.g., information gathering, computation, etc.) would be required to reach an optimal solution (Weißmüller et al., 2021)	Present recycled water as a solution that could quickly help solve problems while avoiding more time- and money-consuming solutions	*t* (211) = 2.92, *p* < .01, *d* = .40	Yes
Illusory Correlation	Tendency to view events, ideas, etc. as having a relationship (e.g., believing one event caused another), even when there is no evidence that the two things are associated (Lieberman, 1999; McConnell et al., 1994)	Present information about adverse events that happened around the same time that changes to a conventional water system were made	*t* (216) = 2.74, *p* < .01, *d* = .37	Yes
Outcome Bias	Tendency to judge the quality of a decision and decision-maker based on its outcomes, rather than the evidence or processes that led to the decision (Baron & Hershey, 1988)	Emphasize examples of how successful other recycled water projects have been, rather than detailed reasons why recycled water would be wise to implement	*t* (210) = 2.53, *p* = .01, *d* = .35	Yes
Reactive Devaluation	Tendency to favor proposals made by those viewed as similar to oneself / devalue proposals made by those viewed as opposite, regardless of the	Describe those attempting to implement recycled water as being similar to those who make up the population of interest	*t* (206) = 2.09, *p* = .04, *d* = .29	Yes

	content of the proposal (e.g. Maoz et al., 2002)			
Omission Bias	Tendency to evaluate harm done because action *was not* taken as more favorable than harm done because an action *was* taken (Yeung et al., 2022)	Describe any harms coming from recycled water as the result of unintentional oversight rather than intentional actions	t (212) = 3.01, $p < .01$, $d = .41$	No
Hyperbolic Discounting (Temporal Discounting)	Tendency to value immediate benefits over future benefits, even if the future benefits would be bigger (Hershfield et al., 2011; Vuchinich & Simpson, 1998)	Describe immediate benefits of recycled water, rather than greater benefits that could arrive in the future	t (209) = 3.15, $p < .01$, $d = .43$	No
Bandwagon Effect	Tendency to align one's beliefs and behaviors with those of the larger group (Steininger et al., 2014)	Present information about the number of communities currently using recycled water	t (214) = 2.96, $p < .01$, $d = .40$	No
Sunk Cost	Tendency to continue doing something even after the benefits stop outweighing the costs because of previously invested time and resources (Roth et al., 2015)	Give individuals recycled water to drink, but refrain from informing them it is recycled until they have already consumed a large portion	t (741) = .25, $p = .80$, $d = .02$	No
Social Norm Bias	Tendency to make decisions that align with collective views of what behaviors are socially approved (Melnyk et al., 2022)	Inform individuals about how many their friends would support recycled water	t (207) = 1.03, $p = .30$, $d = .14$	No

Identifiable Victim Effect	Tendency to be more willing to provide help when a specific victim is identified as being harmed, rather than victims who are anonymous (S. Lee & Feeley, 2016)	Provide example(s) of specific, named people who would suffer should water shortages not be addressed	t (207) = 1.33, p = .19, d = .18	No
Category Size	Tendency to judge the likelihood of an event happening as higher when it comes from a large category, as opposed to a smaller category (Isaac & Brough, 2014; Perfecto et al., 2018)	Simply unfavorable statistics (e.g., instead of saying 25 out of 100 individuals felt recycled water tasted weird, say 1 in 4 felt this way)	t (212) = 1.68, p = .10, d = .23	No
Declinism	Tendency to view the past more positively than the present or future (i.e., the past is viewed as the "good old days" while the present and future are viewed as being in a state of decline (Rettberg, 2020; Steenvoorden & Harteveld, 2018)	Frame recycled water in a way that to appeals to "the good old days" (e.g., recycled water as a way to do what is right by standing up for those who do not have the same water-related privileges we have)	t (214) = .10, p = .92, d = .01	No
Negativity Effect	Tendency to weigh negative events more heavily than positive events, even when positive and negative events are of the same magnitude (Fišar et al., 2022; Hilbig, 2009)	Emphasize positive aspects of recycled water; emphasize any potential negative effects of conventional water	t (203) = .14, p = .89, d = .02	No

Self-serving Bias	Tendency to attribute personal negative outcomes factors outside one's own control, while attributing positive personal outcomes to one's own efforts (Kriss et al., 2011; Krusemark et al., 2008)	Ask individuals to put themselves in the shoes of policymakers and determine what they would do regarding water issues given the information they have	t (213) = 1.93, p = .06, d = .26	No
Bottom Dollar Effect	Tendency to view a product more negatively if it causes one to spend to zero (i.e., to completely exhaust one's budget; Soster et al., 2013, 2014)	Set up a paradigm where participants have to spend the last of their money on conventional water	NA (Eliminated prior to testing)	NA
IKEA Effect	Tendency to value self-made products higher than products made by others (Norton et al., 2012; Sarstedt et al., 2017)	Have individuals assist in the process of purifying recycled water	NA (Eliminated prior to testing)	NA
Mere Exposure Effect	Tendency to view a product as more favorable when one has been exposed to it before, even when the exposure is incidental (Janiszewski, 1993)	Have individuals try recycled water at some point before ratings or decisions are to be made	NA (Eliminated prior to testing)	NA

Table 2

Acceptance Results for Individual Nudge Scenarios by Bias Condition in Experiment 2

Nudge	$M_{Conventional}$ (*SD*)	$M_{Recycled}$ (*SD*)	T-test Results
Default	3.61 (1.48)	4.53 (1.47)	t (80) = 2.80, p < .01, d = .62
Reactive Devaluation	4.40 (1.24)	3.93 (1.23)	t (80) = -1.66, p = .10, d = -.38
Illusory Correlation	3.21 (1.20)	4.14 (1.28)	t (80) = 3.02, p < .01, d = .76
Zero Risk	3.58 (1.50)	4.05 (1.36)	t (80) = 1.48, p = .14, d = .33
In-Group	3.77 (1.56)	3.87 (1.28)	t (80) = .32, p = .75, d = .07
Affect	2.86 (1.51)	3.91 (1.57)	t (80) = 3.04, p < .01, d = .69
Ambiguity	2.84 (1.48)	4.07 (1.39)	t (80) = 3.87, p < .001, d = .86

Note. These effects were derived from the single acceptance questions asked after each scenario was presented (e.g., effects for the default come from the preference for recycled vs. conventional water question asked after the default nudge, effects for reactive devaluation come from the preference question asked after the reactive devaluation scenario, and so on).

Table 3

Average Acceptance Results Across All Scenarios by Nudge and Bias Condition in Experiment 2

Nudge	$M_{Conventional}$ (*SD*)	$M_{Recycled}$ (*SD*)	T-test Results
Default	3.65 (.67)	3.83 (.84)	t (80) = 1.10, p = .28, d = .24
Reactive Devaluation	4.01 (.67)	3.25 (.63)	t (80) = -5.02, p < .001, d = -1.15
Illusory Correlation	3.94 (.70)	3.12 (.53)	t (80) = -4.91, p < .001, d = -1.24
Zero Risk	3.92 (.82)	3.57 (.65)	t (80) = -2.19, p = .03, d = -.49
In-Group	4.04 (.69)	3.23 (.55)	t (80) = -5.61, p < .001, d = -1.28
Affect	3.92 (.81)	3.45 (.56)	t (80) = -2.93, p < .01, d = -.66
Ambiguity	4.03 (.81)	3.48 (.59)	t (80) = -3.56, p < .001, d = -.79

Note. These effects were derived by averaging across responses to the singular recycled vs. conventional water preference question asked after each of the seven scenarios.

Table 4

Correlation Matrix from Experiments 2 (Top) and 3 (Bottom)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Education Condition	1															
2. Default Condition	.09	1														
	.09															
3. Average Acceptance	.27*	.12	1													
	.32**	.07														
4. Acceptance Variance	-.23*	-.06	-.15	1												
	.03	.01	-.13													
5. Upset Variance	-.31**	-.06	-.01	.58**	1											
	-.08	.02	-.24**	.33**												
6. Trust Variance	-.19	-.05	-.04	.43**	.64**	1										
	.02	.01	-.02	.28**	.50**											
7. Worry Variance	-.20	-.02	.00	.42**	.71**	.65**	1									
	-.06	.01	-.09	.35**	.56**	.48**										
8. Protest Variance	-.19	-.04	-.15	.42**	.56**	.52**	.50**	1								
	-.17*	-.14	-.25**	.34**	.58**	.37**	.55**									
9. Move Variance	-.25*	-.17	-.12	.48**	.71**	.66**	.64**	.74**	1							

	-.10	-.02	-.22**	.27**	.59**	.26**	.47**	.67**								
10. Negative	-.19	-.16	-.14	.42**	.67**	.57**	.53**	.74**	.90**	1						
Recommendation Variance	-.11	.01	-.22**	.27**	.63**	.40**	.59**	.62**	.78**							
11. Switches	-.07	-.02	-.37**	.46**	.30**	.23*	.37**	.21	.25*	.21	1					
	-.08	.16*	-.15*	.32**	.18*	.14	.21**	.13	.14	.16*						
12. Post Objective	.25*	-.04	.28*	-.07	-.13	-.13	-.19	-.08	-.14	-.12	-.05	1				
Knowledge	.38**	-.05	.27**	.01	-.07	.12	-.08	-.14	-.07	-.04	-.02					
13. Numeracy	.15	-.05	.19	-.30*	-.21	-.14	-.18	-.08	-.19	-.13	-.14	.29**	1			
	.09	-.06	.20**	-.18*	-.09	-.09	-.04	-.08	-.12	-.05	-.07	.13				
14. Politics	.11	-.11	-.27*	-.07	.04	.12	.15	-.13	.04	.06	.17	-.10	-.13	1		
	-.09	-.01	-.34**	.16*	.10	-.03	.11	.14	.11	.08	.11	-.13	-.06			
15. Age	.12	-.03	.02	.03	.03	.01	.06	.03	.10	.02	.01	.11	.08	-.01	1	
	.08	-.05	.05	.05	.10	.12	.03	.12	.05	.04	-.02	.06	.02	.05		
16. Gender	.09	.04	.04	.01	.04	.17	.13	.09	.11	.09	-.11	.23*	.25*	-.08	.08	1
	.00	.09	.11	-.07	-.07	-.09	-.05	-.10	-.09	-.07	-.09	.10	.31**	.11	.20**	

Note. Correlations come the full samples (*Ns* = 82, 187); variables were coded as follows: education condition (1 = control, 2 = education); default condition (1 = default out, 2 = default in); politics (1 = strongly liberal, 7 = strongly conservative); gender (1 = male, 2 = female); For Experiment 2 gender correlations, participants who were non-binary or declined to specify gender ($n = 2$) were excluded; * $p < .05$, ** $p < .01$

Table 5

Acceptance Results for Individual Nudge Scenarios by Bias Condition in Experiment 3

Nudge	$M_{Conventional}$ (*SD*)	$M_{Recycled}$ (*SD*)	**T-test Results**
Default	3.08 (1.39)	4.14 (1.37)	t (185) = 5.25, p < .001, d = .77
Reactive Devaluation	3.94 (1.11)	3.84 (.98)	t (185) = -.65, p = .51, d = -.10
Illusory Correlation	3.42 (1.30)	3.73 (1.08)	t (185) = 1.58, p = .12, d = .25
In-Group	3.76 (1.28)	3.43 (1.46)	t (185) = -1.67, p = .10, d = -.25
Zero Risk	3.41 (1.37)	3.69 (1.26)	t (185) = 1.47, p = .14, d = .22
Affect	2.89 (1.32)	3.54 (1.73)	t (185) = 2.91, p < .01, d = .43
Ambiguity	3.01 (1.43)	3.79 (1.48)	t (185) = 3.55, p < .001, d = .53

Note. These effects were derived from the single acceptance questions asked after each scenario was presented (e.g., effects for the default come from the preference for recycled vs. conventional water question asked after the default nudge, effects for reactive devaluation come from the preference question asked after the reactive devaluation scenario, and so on).

Table 6

Average Acceptance Results Across All Scenarios by Nudge and Bias Condition Experiment 3

Nudge	$M_{Conventional}$ (*SD*)	$M_{Recycled}$ (*SD*)	T-test Results
Default	3.50 (.73)	3.60 (.70)	t (185) = .91, p = .36, d = .13
Reactive Devaluation	3.86 (.61)	3.16 (.64)	t (185) = -7.62, p < .001, d = -1.12
Illusory Correlation	3.74 (.67)	3.15 (.65)	t (185) = -5.69, p < .001, d = -.90
Zero Risk	3.90 (.64)	3.21 (.62)	t (185) = -7.33, p < .001, d = -1.07
In-Group	3.82 (.62)	3.21 (.68)	t (185) = -6.36, p < .001, d = -.94
Affect	3.85 (.62)	3.13 (.62)	t (185) = -7.84, p < .001, d = -1.16
Ambiguity	3.77 (.62)	3.41 (.73)	t 185) = -3.46, p < .001, d = -.52

Note. These effects were derived by averaging across responses to the singular recycled vs. conventional water preference question asked after each of the seven scenarios.

Table 7

2x2 ANOVA Results for Additional Effectiveness Indicators Education and Default Condition from Experiment 3

Indicator	**Education Main Effect**	**Default Main Effect**	**Interaction**
Upset	F (1, 183) = 9.97, $p < .01$, $\eta^2 = .05$	F (1, 183) = .04, $p = .85$, $\eta^2 < .01$	F (1, 183) = .82, $p = .37$, $\eta^2 < .01$
Trust	F (1, 183) = 8.09, $p < .01$, $\eta^2 = .04$	F (1, 183) = 1.54, $p = .22$, $\eta^2 < .01$	F (1, 183) = .05, $p = .83$, $\eta^2 < .01$
Worry	F (1, 183) = 14.29, $p < .001$, $\eta^2 = .07$	F (1, 183) = 2.08, $p = .15$, $\eta^2 = .01$	F (1, 183) = .44, $p = .51$, $\eta^2 < .01$
Protest	F (1, 183) = 7.12, $p = .01$, $\eta^2 = .04$	F (1, 183) = .63, $p = .43$, $\eta^2 < .01$	F (1, 183) = .03, $p = .88$, $\eta^2 < .01$
Move	F (1, 183) = 3.63, $p = .06$, $\eta^2 = .02$	F (1, 183) < .01, $p = .99$, $\eta^2 < .01$	F (1, 183) = .02, $p = .88$, $\eta^2 < .01$
Negative Recommendation	F (1, 183) = 5.95, $p = .02$, $\eta^2 = .03$	F (1, 183) = .04, $p = .85$, $\eta^2 < .01$	F (1, 183) = .05, $p = .83$, $\eta^2 < .01$

Note. To replicate procedure from Tanner (2021), these analyses come solely from the questions asked after the default nudge; significant effects were all in the hypothesized direction (i.e., the educational intervention increased levels of trust and decreased levels of upset, worry, protest, move, and negative recommendations relative to the control condition).

Table 8

Tests Comparing Variance for Additional Acceptance Indicators from Experiment 3

	Control Group		Education Group		
Indicator	M_{SD}	SD_{SD}	M_{SD}	SD_{SD}	***T-test***
Upset	.75	.44	.68	.45	t (185) = 1.02, p = .31, d = .15
Trust	.79	.39	.80	.40	t (185) = -.24, p = .81, d = -.04
Worry	.70	.40	.65	.41	t (185) = .83, p = .41, d = .12
Protest	.62	.46	.47	.39	t (185) = 2.31, p = .02, d = .34
Move	.62	.46	.53	.47	t (185) = 1.32, p = .19, d = .19
Negative Recommendation	.62	.44	.52	.43	t (185) = 1.49, p = .14, d = .22

Appendix

Wording and Order of Biases Retained After Experiments 1a and 1b

Italics indicate the differences between the two different versions of each bias.

1) Reactive Devaluation:
   a. Recycled water bias: "Suppose you are attending a city council meeting. The mayor–who *is your close personal friend*, *shares many of the same political views* that you do, and who *is known to be well-respected* among much of the community–announces that after much research and deliberation, they have decided to propose a program that will bring recycled water to residential taps rather than conventional water. Rate how likely you would be to be in favor of the mayor's recycled water proposal."
   b. Conventional water bias: "Suppose you are attending a city council meeting. The mayor–who *you have never met*, *holds many of the opposite political views* that you do, and who *is not known to be well-respected* among much of the community–announces that after much research and deliberation, they have decided to propose a program that will bring recycled water to residential taps rather than conventional water. Rate how likely you would be to be in favor of the mayor's recycled water proposal.
2) Illusory Correlation:
   a. Recycled water bias: "Imagine that researchers are studying indicators of health in your community. Among other things, the researchers reported that there was a significant increase in mental health-related emergencies reported right about the time that the city *switched from their old recycled water system to using*

conventional water. Please rate the extent to which you agree with the following statement: Your community should *switch back to the old recycled water system*."

   b. Conventional water bias: "Imagine that researchers are studying indicators of health in your community. Among other things, the researchers reported that there was a significant increase in mental health-related emergencies reported right about the time that the city *switched from their old conventional water system to using recycled water*. Please rate the extent to which you agree with the following statement: Your community should *switch back to the old conventional water system*."

3) Zero Risk Bias (Certainty Effect)

   a. Recycled water bias: "Imagine you live in an area that is currently being affected by a drought. In response to the drought, officials have proposed two competing solutions, both of which unfortunately might increase your water bill. The two programs are (1) A *recycled water program, in which it is 100% certain* that your water bill will increase by $10 per month, and (2) A *new conventional water program, in which there is a 50% chance* your water bill will not increase, but also a 50% chance that your water bill will increase by $20 per month. Which of the programs would you be more likely to support?"

   b. Conventional water bias: "Imagine you live in an area that is currently being affected by a drought. In response to the drought, officials have proposed two competing solutions, both of which unfortunately might increase your water bill. The two programs are (1) A *new conventional water program, in which it is 100% certain* that your water bill will increase by $10 per month, and (2) A *recycled* *water program, in which there is a 50% chance* your water bill will not increase, but also a 50% chance that your water bill will increase by $20 per month. Which of the programs would you be more likely to support?"

4) In-Group Bias

    a. Recycled water bias: "Suppose the city council of your town has decided to consider switching from supplying conventionally treated water to supplying treated recycled water to residential taps. As the mayor is trying to determine how they will vote on the issue, they consult the scientific literature, where they find that the consensus of the experts is that both sources of water would be safe, that it has also provided some advantages such as resilient water supplies during times of drought, but that *cities in your state that have used recycled water have reported the water looking, tasting, and costing different than conventional water.* The mayor was about to vote to switch to recycled water when they remembered that they had a friend who lives in a city using recycled water. The mayor called up *the friend, who reported the water tasted the same after they switched, and that their water bill had remained the same.* Given that the mayor's vote will likely be the deciding vote on the issue, which way do you think they should vote?"

    b. Conventional water bias: "Suppose the city council of your town has decided to consider switching from supplying conventionally treated water to supplying treated recycled water to residential taps. As the mayor is trying to determine how they will vote on the issue, they consult the scientific literature, where they find that the consensus of the experts is that both sources of water would be safe, that it has also provided some advantages such as resilient water supplies during times of drought, and that *cities in your state that have used recycled water have reported the water looking, tasting, and costing the same as conventional water.* The mayor was about to vote to switch to recycled water when they remembered that they had a friend who lives in a city using recycled water. The mayor called up *the friend, who reported the water did not taste the same after they switched, and that their water bill had gone up.* Given that the mayor's vote will likely be the deciding vote on the issue, which way do you think they should vote?"

5) Affect Heuristic

   a. Recycled water bias: "Imagine you take a tour *of your conventional water treatment plant*, and find that at one point during the treatment process, the water looks like this:

   Given this information, would you be more likely to prefer conventional water, or would you prefer recycled water?"

   b. Conventional water bias: "Imagine you take a tour *of your recycled water treatment plant*, and find that at one point during the treatment process, the water looks like this:

Given this information, would you be more likely to prefer conventional water, or would you prefer recycled water?”

6) Ambiguity Effect

   a. Recycled water bias: “Suppose you’re taking a tour of both a local recycled and conventional water plant. *At the recycled water plant, the person guiding you through the tour appears very knowledgeable* about the process and seems to be able to answer every concern and question you have about the treatment process with ease. Meanwhile, *at the conventional water plant, the person guiding you through the tour seems to know only the very basics* of the process and is very ambiguous in his answers to questions and concerns about the treatment process. At the end of both tours, you are given a bottle of water taken straight from the plant. Which would you be more likely to drink?”

   b. Conventional water bias: “Suppose you’re taking a tour of both a local recycled and conventional water plant. *At the conventional water plant, the person guiding you through the tour appears very knowledgeable* about the process and seems to be able to answer every concern and question you have about the treatment process with ease. Meanwhile, *at the recycled water plant, the person guiding you through the tour seems to know only the very basics* of the process and is very ambiguous in his answers to questions and concerns about the treatment process.

At the end of both tours, you are given a bottle of water taken straight from the plant. Which would you be more likely to drink?”