

The Unintended Effects of Ethical Decision Aids in Organizations

Finance Working Paper N° 1041/2025

August 2026

Malte Baader

Norwegian School of Economics

Maxim Egorov

Munich Business School

Baiba Renerte

University of Zurich

Carmen Tanner

University of Zurich

Alexander F. Wagner

University of Zurich

© Malte Baader, Maxim Egorov, Baiba Renerte,
Nicole Witt
Carmen Tanner, Alexander F. Wagner and Nicole
Witt 2026. All rights reserved. Short sections of

text, not to exceed two paragraphs, may be quoted
without explicit permission provided that full credit,
including © notice, is given to the source.

This paper can be downloaded without charge from:
<https://www.ecgi.global/publications/working-papers>

ECGI Working Paper Series in Finance

The Unintended Effects of Ethical Decision Aids in Organizations

Working Paper N° 1041/2025
August 2026

Malte Baader
Maxim Egorov
Baiba Renerte
Carmen Tanner
Alexander F. Wagner
Nicole Witt

We thank the Editor (Daniele Nosenzo), an anonymous Associate Editor, and two anonymous Referees for excellent comments.

© Malte Baader, Maxim Egorov, Baiba Renerte, Carmen Tanner, Alexander F. Wagner and Nicole Witt 2026. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Abstract

Unethical behavior, deception, and fraud are major concerns in corporate governance. This paper examines the effectiveness of contemplation questions (CQs) as ethical decision aids intended to activate moral agency and prompt employees to consider their actions from various perspectives (e.g., self, peers, company). Although we find that 44% of S&P 200 and Fortune Global 200 companies use CQs, their behavioral impact is not well understood. Using two pre-registered, incentivized vignette experiments, we examine the effects of CQs under two implementation formats. In Study 1 (N = 1,986), participants were presented with CQs without being required to respond. In Study 2 (N = 1,450), participants were required to answer the CQs before deciding. Across both studies, CQs have no average effect on ethical behavior. However, in Study 2, the effects are strongly moderated by moral identity. CQs marginally increase ethical decisions among individuals with high moral identity, yet significantly reduce such decisions among those with low moral identity. These findings align with a conceptual framework of motivated moral reasoning and suggest that while CQs can benefit some individuals, they may backfire precisely among those predisposed to unethical behavior.

Keywords:

JEL Classification: D91, C91, M14

Malte Baader

Norwegian School of Economics

Maxim Egorov

Munich Business School

Baiba Renerte

University of Zurich

Carmen Tanner

University of Zurich

Alexander F. Wagner

University of Zurich

Nicole Witt

Zeppelin University

The Unintended Effects of Ethical Decision Aids in Organizations

Malte Baader

Norwegian School of
Economics, NO
malte.baader@nhh.no

Maxim Egorov

Munich Business School, DE
maxim.egorov@munich-business-school.de

Baiba Renerte

University of Zurich, CH
baiba.renerte@df.uzh.ch

Carmen Tanner

University of Zurich, CH
Zeppelin University, DE
carmen.tanner@df.uzh.ch

Alexander F. Wagner

University of Zurich,
CEPR, ECGI, and SFI, CH
alexander.wagner@df.uzh.ch

Nicole Witt

Zeppelin University, DE
nicole.witt@zu.de

August 12, 2026

Abstract

Unethical behavior, deception, and fraud are major concerns in corporate governance. This paper examines the effectiveness of contemplation questions (CQs) as ethical decision aids intended to activate moral agency and prompt employees to consider their actions from various perspectives (e.g., self, peers, company). Although we find that 44% of S&P 200 and Fortune Global 200 companies use CQs, their behavioral impact is not well understood. Using two pre-registered, incentivized vignette experiments, we examine the effects of CQs under two implementation formats. In Study 1 ($N = 1,986$), participants were presented with CQs without being required to respond. In Study 2 ($N = 1,450$), participants were required to answer the CQs before deciding. Across both studies, CQs have no average effect on ethical behavior. However, in Study 2, the effects are strongly moderated by moral identity. CQs marginally increase ethical decisions among individuals with high moral identity, yet significantly reduce such decisions among those with low moral identity. These findings align with a conceptual framework of motivated moral reasoning and suggest that while CQs can benefit some individuals, they may backfire precisely among those predisposed to unethical behavior.

JEL Codes: D91, C91, M14

Keywords: Organizational behavior, ethical decision-making

*This research has been approved by the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology at the University of Zurich (OEC IRB # 2021-020). We thank the Editor (Daniele Nosenzo), an anonymous Associate Editor, and two anonymous Referees for excellent comments. We are grateful for comments and feedback from participants at seminars at the University of Nottingham, University of Konstanz, University of Vienna, NYU Abu Dhabi, NHH Bergen, Thurgau Institute of Economics, Zhejiang University, and University of Zurich. We further appreciate input from Michael Beier, Alexander Cappelen, Georg Gösswein, Christian Hauser, Anna Hochleitner, Nikos Nikiforakis, Ernesto Reuben, Urs Fischbacher, Bertil Tungodden, Roberto Weber and Richard Zeckhauser on this project. We are also thankful to attendants of the Integrity Europe Conference 2023, AOW-Fachgruppentagung 2023, DICO Webcast 2023, TEAP 2024, Maastricht BEPS 2024, Helsinki ESA Europe 2024, THEEM 2025, Beijing ESA World Meeting 2025, and Beijing BEAT Workshop 2025. We also thank Jana Borchert, Maximilian Paul, Etienne Rechsteiner and Francesca Schirinzi for excellent research assistance. This work was supported by the Swiss National Science Foundation [10001CL_201117/1] and by the German Research Foundation [460293325].

1 Introduction

Ethical business scandals, such as those at Volkswagen, Wirecard, Enron, or Cambridge Analytica, continue to rock the corporate world. In response, policymakers across jurisdictions have intensified rulemaking and enforcement to strengthen corporate ethics and compliance (KPMG, 2024). Some of these efforts protect customers, such as the General Data Protection Regulation governing how personal data is handled, whereas others protect employees and workers, such as the EU Whistleblower Directive, which shields those who report wrongdoing, and the German Supply Chain Due Diligence Act. Our study focuses on the individual employee as the ethical decision-maker, whose choices can in turn affect customers and other stakeholders. Reflecting these shifts, 55% of companies reported an expected increase in compliance budgets in 2023 (Thomson Reuters Institute, 2023), alongside efforts to strengthen ethical decision-making beyond traditional codes of conduct (Weaver, 2014). Among these efforts, organizations have increasingly turned to behavioral tools to influence everyday decision-making before misconduct occurs.

One popular tool for guiding ethical behavior in organizations involves brief, reflective prompts, which we refer to as *contemplation questions* (CQs). These are simple, structured questions that aim to help individuals pause and reflect when confronted with ethically ambiguous situations. Often called “checking” or “test” questions in organizational contexts, such prompts are intended not to provide clear-cut answers, but to prompt self-reflection and help individuals think through the potential consequences of their actions. We use the term contemplation questions to emphasize their reflective function. Examples of CQs include the “mirror test” (“Would I like the person I see in the mirror after I have made the decision?”) and the “publicity test” (“Would the public approve of my decision if they read about it in the newspaper?”).

CQs are intended to increase the salience of relevant reference groups and their associated norms and standards. For example, the mirror test cues personal standards, while the publicity test invokes societal norms. By prompting individuals to compare their intended actions against salient standards of others, CQs are thought to activate self-regulatory processes (Bandura, 1991; Gunia et al., 2012), thereby fostering moral agency and overcoming bounded ethical awareness (Bazerman and Sezer, 2016; Bazerman and Tenbrunsel, 2011). Ethical blind spots, that is, failures to recognize the moral dimensions of decisions, may be mitigated when individuals are encouraged to slow down and reflect. In their structure, CQs resemble the simple yes/no prompts shown in the heuristics literature to guide effective professional decisions (e.g., Gigerenzer and Todd, 1999; Dhimi, 2003), yet unlike “fast and frugal” heuristics, they are

intended to slow the decision process and to encourage deliberate ethical reflection. In this way, CQs function as what [Luo et al. \(2023\)](#) call “sludge for good,” namely, helpful frictions that interrupt automaticity and stimulate ethical consideration. The hope, therefore, is that CQs can serve as subtle nudges toward ethical vigilance.

Many organizations have adopted such tools to support ethical decision-making in practice. [Renerte et al. \(2025\)](#) show that 44% of S&P 200 and Fortune Global 200 companies incorporate CQs into their ethics and compliance programs. They are also frequently recommended in practitioner-oriented publications (e.g., [Kouchaki and Smith, 2020](#)). Scholars have highlighted their potential to raise awareness of problematic decisions and encourage principled reflection (e.g., [Bazerman and Sezer, 2016](#); [Martin et al., 2014](#)).

Yet, despite their intuitive appeal and widespread adoption, causal evidence on the effectiveness of CQs is surprisingly sparse. One exception is a study by [Zhang et al. \(2015\)](#) who showed that a single CQ (“Do you see any problem with any of these funds?”) improved participants’ ability to identify fraudulent investment products. However, the broader empirical question of whether, and under what conditions, CQs shape ethical decision-making remains largely unanswered.

This paper addresses this gap by testing the effectiveness of several specific CQs. Do they, in fact, promote more ethical decision-making? Drawing on two pre-registered, incentivized vignette experiments, we examine the effects of CQs on ethical behavior. In Study 1, participants were presented with a CQ without being required to respond; in Study 2, they were required to answer the CQ before making their decision. In both studies, participants encountered CQs referencing different standards: personal values, peer norms, company policies, public expectations, and universal moral principles.

In Study 1 ($N = 1,986$), participants were presented with CQs passively, without being required to respond. We find no effect on ethical decision-making – neither in the aggregate nor across specific reference groups. In Study 2 ($N = 1,450$), we ask participants to answer the questions explicitly. While we again observe no overall effect, we find striking heterogeneity based on individual differences in moral identity.¹ Among participants high in moral identity, CQs marginally increase the likelihood of ethical decisions. Among those low in moral identity, by contrast, CQs significantly reduce ethical choices. We show that this pattern of heterogeneity is robust when moderating instead with other elicited traits relevant for ethi-

¹Moral identity is an established construct in psychology ([Aquino and Reed II, 2002](#); [Reed II and Aquino, 2003](#)), often associated with moral decisions (see e.g., [Hardy et al., 2013](#); [Hertz and Krettenauer, 2016](#); [Winterich et al., 2013](#)). It has also been shown to moderate self-serving tendencies in moral conflict situations ([Tian et al., 2025](#)).

cal decision-making, including honesty-humility and social value orientation. Taken together, these results provide new causal evidence that CQs can have unintended negative effects, particularly for individuals with a weaker internal moral compass.

To explain these findings, we develop a conceptual framework in which CQs increase the salience of standards associated with particular reference groups. Individuals derive utility from aligning their actions with what they believe these reference groups endorse (Akerlof and Kranton, 2000; Bernheim, 1994). We propose two competing mechanisms through which CQs may influence ethical behavior, depending on what individuals believe which decisions the reference group approves of. First, in line with the theory of moral agency (Bandura, 1991), if individuals perceive the reference group to subscribe to moral standards of *right* and *wrong*, CQs may encourage ethical action by activating a sense of moral obligation. Second, drawing on theories of motivated reasoning (Bénabou and Tirole, 2011, 2016), individuals may project their own preferences onto the reference group, thereby using CQs to justify rather than challenge unethical behavior. Our findings suggest that this second mechanism predominates among individuals low in moral identity. This has direct and significant implications for practice: exactly for the employees whom companies most hope to reach, CQs can unintentionally backfire by reinforcing unethical inclinations rather than mitigating them.

Our paper contributes to various strands of literature. First, it adds to research on value- or integrity-oriented interventions that affect ethical decision-making or dishonesty (see e.g., Dimant and Shalvi, 2022; Paine, 1994; Weaver and Treviño, 1999; Zhang et al., 2015). Studies in psychology, economics, and other behavioral sciences have shown that dishonesty is malleable and can be influenced through interventions (see Gerlach et al., 2019; Irlenbusch and Villeval, 2015; Rosenbaum et al., 2014; Shalvi et al., 2025, for reviews). Although in general there appears to be a preference for truth telling (Abeler et al., 2019; Bellemare et al., 2019), with some people remaining honest even when faced with large financial incentives (Cohn et al., 2019; Gibson et al., 2013; Hilbig et al., 2012), dishonest behavior can be even further reduced through interventions such as physical mirrors (Batson et al., 1999; Diener and Wallbom, 1976), being exposed to a pair of eyes on a poster (Bateson et al., 2006), by exposure to moral symbols (Desai and Kouchaki, 2017), by providing moral reminders (Bursztyn et al., 2019), or by the order of presented information (Gneezy et al., 2020). Some of these interventions fall under the umbrella term of moral primes for which our aggregate null effects support their documented instability (Huynh et al., 2024; Kristal et al., 2020; Vaidis et al., 2024; Verschuere et al., 2018).

Second, a large body of literature has documented backfiring of ethical interventions (see e.g., Beshears et al., 2015; Damgaard and Gravert, 2018; Irlenbusch et al., 2020; Pierce et al.,

2025; Sachdeva et al., 2009; Schultz et al., 2007). For example, Morvinski et al. (2023) find that within a workplace environment, individuals stating their beliefs about their appropriate wage *ex ante* can cause more self-serving behavior. Similarly, Köbis et al. (2019) present evidence that promoting intuitive decisions can systematically lead to more dishonest behavior. More recent work by Köbis et al. (2025) and Leib et al. (2024) demonstrates comparable backfiring effects, showing that AI-delegation and -advice can increase dishonesty whilst failing to promote honesty. Our paper advances this literature by focusing on an ethical intervention that we document to be widely used in global corporations. Moreover, a central finding of our study is the heterogeneity in effects. We identify positive effects for some individuals when incorporating moral identity as a moderator, but unintended adverse effects for others. In particular, backfiring occurs for those most prone to unethical behavior.

Lastly, our interpretation that CQs may prompt individuals to justify their priors contributes to the literature on motivated reasoning (Bénabou and Tirole, 2016). More precisely, the observed behavior closely links to motivated *moral* reasoning (Ditto et al., 2009; Haidt, 2001; Uhlmann et al., 2009) or self-serving justification (Shalvi et al., 2015) in psychology, both considered a specific application of motivated reasoning. In this strand of literature, individuals construct a defensible justification that supports the preferred action rather than objectively re-evaluate the actions. A key argument behind motivated moral reasoning is to preserve one’s self-esteem and small opportunities for “moral wiggle room” have frequently been shown to be exploited (Cohn et al., 2022; Dana et al., 2007; Gneezy et al., 2019, 2020; Konow, 2000). Similarly, Saccardo and Serra-Garcia (2023) find that whilst 30–45 percent of individuals limit their capacity for motivated beliefs, 40–55 percent of individuals actively seek such opportunities even when it is costly. Subsequent research supports the notion of motivated moral reasoning by highlighting various motivational forces (for example emotions) when deciding to act morally or immorally (Greene et al., 2001; Monin et al., 2007). Substantial evidence shows that individual’s reasoning leads them to morally disengage (Bandura, 2011; Welsh et al., 2020), misperceive the objectivity of moral situations (Haidt, 2001; Pizarro et al., 2003; Provencher and Fincham, 2000), and results in selective memory of moral events (Dimant et al., 2020; Pizarro et al., 2006).

This paper is structured as follows. Section 2 presents evidence on the prominence and occurrence of CQs in global corporations. Section 3 outlines our conceptual framework to illustrate how CQs can affect ethical decision-making. Sections 4 and 5 describe the design, procedures, and results of Study 1 and 2, respectively. We provide a discussion and concluding remarks in Section 6.

2 The Occurrence of Contemplation Questions

How prevalent is the use of CQs in large companies around the world? To answer this, we reviewed companies that are part of the 2022 Fortune Global 200 and 2023 S&P 200. Here, we summarize the key findings that motivate this paper. For more details, we refer to the companion study [Renerte et al. \(2025\)](#). As of mid-2022, the Fortune Global 200 list comprised the world’s largest companies by annual revenue across 20 countries and 11 sectors, whereas the S&P 200 list included the largest companies listed in the US by market capitalization spanning 11 sectors. From both lists, we identified 355 unique firms that compose our target sample. To explore the patterns of CQ usage among these companies, several research assistants independently identified CQs from publicly available company websites.² Since all identified CQs are publicly available, our data should be considered a lower bound of their prominence in companies.

Overall, across both lists, we find that 156 of the 355 companies (44%) utilize CQs. Among the Global Fortune 200 companies, we can further see that the use of CQs is somewhat heterogeneously spread throughout the world, with 65% of companies using them in North America, and only 13% in Asia-Pacific. This heterogeneity is also reflected in all sectors, where 73% of the companies reviewed in healthcare use CQs, while only 31% of the companies in communication services or even 0% in real estate promote CQs.

[Renerte et al. \(2025\)](#) identify 727 CQs in all 156 companies, which means conditional on utilizing some form of CQ, each company uses on average 4.66 CQs.³ Of these CQs, 72% are positively framed, only 15% have a negative frame, and 12% are neutrally framed.⁴ While some companies embed CQs in decision trees, the vast majority employ them as a list of independent questions. [Renerte et al. \(2025\)](#) further relate the prevalence of CQs to company characteristics and perform a clustering exercise to categorize CQs. This results in five distinct categories, differing in the reference group and standards they address: *Self*, *Peers*, *Company Policies & Culture*, *Law & Reputation*, and *Universal Principles*. The categorization is data-driven, using a *k*-means clustering of employee-survey ratings of the reference groups each CQ invokes,

²They first examined the company websites proactively for CQs or similar questions. Subsequently, they used on-line search engines to look for the company in conjunction with the keywords: “code of ethics”, “code of conduct”, and “questions”. After one assistant completed the search, the other assistant repeated the process for the companies where the questions were not initially found. We provide two examples in Appendix D. The data is provided as supplementary material in [Renerte et al. \(2025\)](#).

³A complete list of CQs and some of their properties is provided as supplementary material in [Renerte et al. \(2025\)](#).

⁴An example of a positively framed CQ is “Is this behavior legal?”, whereas a negatively framed one would be “Is this behavior illegal?”.

verified with human coding in which two independent coders assign each of the 727 CQs to one of the five categories. This ensures that the five categories cover the identified CQs exhaustively.

Given the prominence of CQs in large corporations across the world, the question of their effectiveness looms large. Although it seems surprising that firms promote the use of such decision aids among employees without first testing their effectiveness, this is not unusual. Organizations frequently implement interventions or changes to organizational structures without rigorous examination (Barends et al., 2014; Pfeffer and Sutton, 2006; Staw and Epstein, 2000). For CQs, we contribute to closing this gap. Before describing our empirical evaluation, we first provide a conceptual framework through which the mechanisms of CQs impact ethical decision-making.

3 Conceptual Framework

To conceptualize the decision conflict faced by an individual, we draw on established theoretical work from social and moral psychology (Bandura, 1986, 1991; Tajfel, 1979) and economics (Akerlof and Kranton, 2000; Bénabou and Tirole, 2011; Bernheim, 1994). This framework was not pre-registered; it was developed after data collection and is intended to clarify the mechanisms underlying CQs.

3.1 Individual Preferences

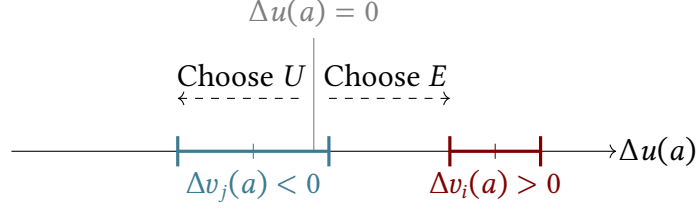
Suppose an individual i faces an ethical conflict in which they need to choose an action a . For simplicity, we assume that a is binary, $a \in \{E, U\}$, where E is the ethical and U the unethical action. The utility of choosing a depends both on the individual valuation of a , $v_i(a)$ as well as on the expectation of i about how a will be perceived by another reference group, $r_i(a)$. This reference group can differ between individuals and decisions and could be, for example, family, society, the employer, or even oneself. We can capture an individual's utility by

$$u_i(a) = v_i(a) + \lambda_i r_i(a). \quad (1)$$

$\lambda_i \in [0, 1]$ describes how much i cares about the evaluation of the reference group.⁵ For

⁵In principle, λ_i could be negative if an individual deliberately opposed the reference group's evaluation. We restrict $\lambda_i \in [0, 1]$ for simplicity, as CQs are intended to increase the weight individuals place on salient reference standards, rather than to induce opposition to them.

Figure 1: Net valuation without reference group salience



Note. The figure depicts two examples for the intuitive net utility of two individuals i (red), and j (blue). Individual i has $\Delta u_i(a) > 0$ (red) and j has $\Delta u_j(a) < 0$ (blue).

simplicity, we assume $r_i(a)$ to be binary with $r_i(a) \in \{0, 1\}$. The social evaluation of action a is positive if the reference group approves of it. This means that the individual receives a utility increase of λ_i if the reference group approves of the action a . An individual would choose a if the utility obtained from this action is greater than the utility from the alternative, that is, if $u_i(a) > u_i(a')$. Defining $u_i(E) - u_i(U)$ as $\Delta u_i(a)$, and analogously $\Delta v_i(a) \equiv v_i(E) - v_i(U)$ and $\Delta r_i(a) \equiv r_i(E) - r_i(U)$, i chooses E if the net utility $\Delta u_i(a) > 0$, that is, if $\Delta v_i(a) + \lambda_i \Delta r_i(a) > 0$.

Let us first consider a case where $\lambda_i = 0$, i.e., a situation where the utility of an individual depends only on $v_i(a)$. $\lambda_i = 0$ can occur if individuals do not care at all about their social image or simply do not consider a given reference group when making their decision. For example, what i 's family would think of a moral choice at work might simply not be a salient question for i . To represent the heterogeneity of preferences within the population, we draw $\Delta v_i(a)$ from a distribution F with $F \sim N(\mu, \delta^2)$. We do not explicitly model how an individual i arrives at their relative evaluation of available actions, but it will involve some weighing of the material rewards against preferences to act ethically or honestly (similar to [Bénabou and Tirole, 2011](#)). $\Delta v_i(a)$ thus defines an ad hoc preference based on an individual's *moral compass*. Due to the instinctive nature of this process, we model $\Delta v_i(a)$ to be accompanied by some noise $\epsilon_i \sim N(0, \delta^2)$ ([Recalde et al., 2018](#)). An individual then chooses E if $\Delta v_i(a) + \epsilon_i > 0$, is indifferent if $\Delta v_i(a) + \epsilon_i = 0$ and chooses U otherwise. Note that due to the noise, some individuals with $v_i(U) > v_i(E)$ still choose E and vice versa.

To illustrate this, Figure 1 represents the net utility of two individuals, $\Delta u_i(a)$ on the horizontal line with the point of indifference in gray. The figure captures the utilities of two individuals, i (red) and j (blue). Both net utilities are displayed as an interval that illustrates the decision noise around $v_i(a)$. The figure shows that while individual i always chooses the ethical action E , as $\Delta u_i > 0$ for all values of ϵ_i , there are some instances where individual j chooses E , despite $\Delta v_j(a)$ being negative.

3.2 Introducing Contemplation Questions

Presumably, the purpose of using CQs as decision aids is to encourage individuals to self-reflect on their actions in relation to other reference groups, including other groups or even oneself. They increase the importance of different reference standards that could influence an individual’s evaluation of the decision conflict. Within our framework, we capture this element of increased salience by a larger λ_i .

However, the effect of this increased salience depends strongly on the perceived approval of the two actions ($\Delta r_i(a)$). In the following, we outline two possible mechanisms of how individuals may form these perceptions and how they can, in turn, influence the choice between E and U . First, in the case of *moral agency*, perceived approval is normatively fixed and individuals align their action with this perceived approval. In contrast, under *motivated moral reasoning*, we keep the individuals’ preferred action constant, but allow them to endogenously distort their beliefs regarding the perceived approval of their action.

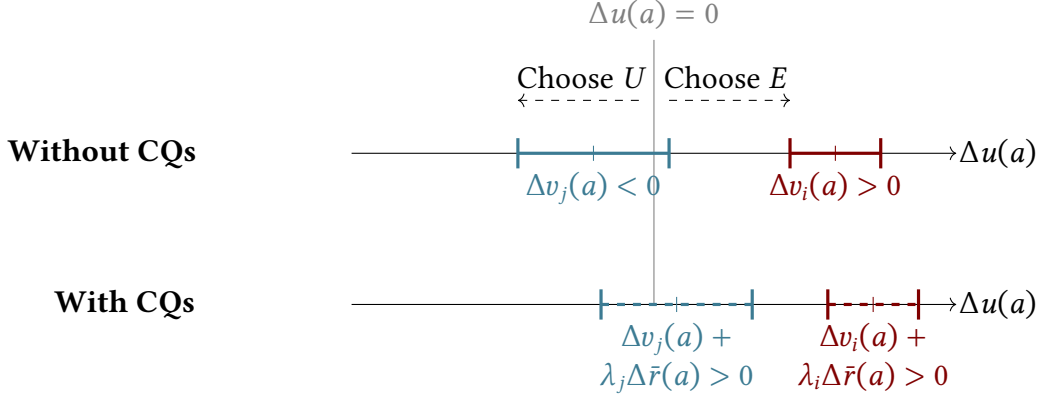
3.2.1 Moral Agency

Let us first consider an agentic perspective on moral decision-making. According to Bandura’s moral agency approach (Bandura, 1986, 1991), moral decisions are guided and evaluated by adopted standards of *right or wrong*. Drawing on this self-regulatory approach, CQs are suggested to enhance the salience of different reference standards or norms and, consequently, influence decisions in favor of what is perceived as *right* (Bandura, 2006). Assuming there exists an objectively *right and wrong*, in accordance with societally beneficial outcomes, we model this such that the respective reference group always prefers the ethical action E , implying that $r_i(E) = 1$, $r_i(U) = 0$, and, since this perceived approval is common across individuals, $\Delta \bar{r}(a) = 1$. This means that choosing the ethical action yields an additional utility gain, shifting Δu_i to the right (see Figure 2). Depending on the size of λ_i , as well as the noise ϵ_i , this means that some individuals change their preferred action from U to E (blue).

From the visualization, we can see that this scenario is most likely for individuals with a small negative net utility $\Delta v_i(a)$ or large λ_i , meaning high concerns for maintaining a positive moral self-image. This implication allows us to formulate a first testable hypothesis.

Hypothesis 1. Moral agency predicts, on average, an increase in the likelihood of ethical decisions in the presence of CQs.

Figure 2: The effect of CQs through *Moral Agency*



Note. According to moral agency, both individuals shift their net utility towards the ethical action.

3.2.2 Motivated Moral Reasoning

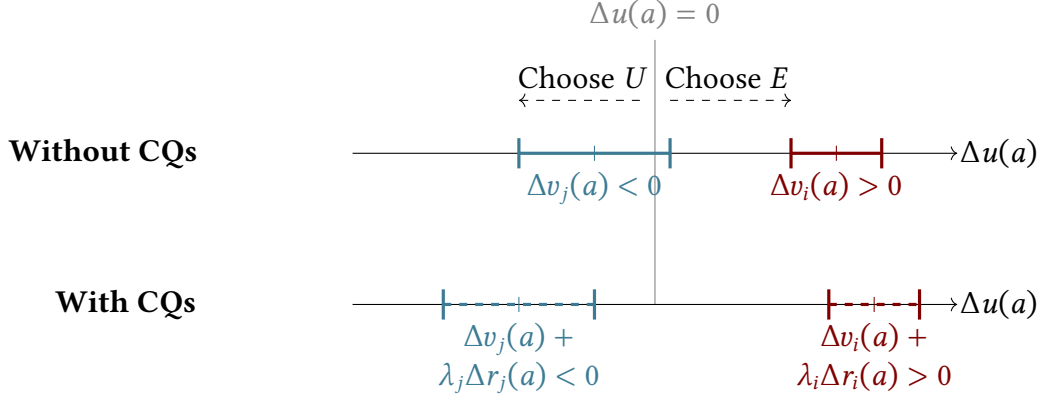
Another possibility of modeling how individual actions can be influenced through $r_i(a)$ is through the lens of motivated moral reasoning. In economics (Bénabou and Tirole, 2011, 2016) and moral psychology (Ditto et al., 2009; Haidt, 2001; Uhlmann et al., 2009), there is ample evidence that individuals engage in motivated moral reasoning, be it through selective memory (Golman et al., 2017; Zimmermann, 2020), confirmation bias (Charness and Dave, 2017; Nickerson, 1998) or the justification of held intuitions (Gigerenzer, 2008; Weaver et al., 2014). This angle suggests that individuals do not objectively evaluate their actions through self-reflection, but rather employ biased reasoning to justify an already preferred action.

Within our conceptual framework, we capture this mechanism by making $\Delta r_i(a)$ endogenous to the decision-maker. Thus, instead of a normatively fixed $\Delta \bar{r}(a)$, individuals now justify their individual valuation ($\Delta v_i(a)$) by forming corresponding perceptions about its approval $\Delta r_i(a)$. Consider two individuals, i and j , with different individual valuations, where i has a higher valuation of E ($\Delta v_i(a) > 0$) and j of U ($\Delta v_j(a) < 0$). Both now engage in motivated moral reasoning by self-servingly aligning $\Delta r_i(a)$ to their individual valuation. This means that i believes that $\Delta r_i(a)$ is positive, while j believes it to be negative. Thus, in contrast to moral agency, where $\Delta \bar{r}(a) = 1$, under motivated moral reasoning $\Delta r_i(a)$ equals 1 when an individual justifies the ethical action and -1 when they justify the unethical one, i.e. $\Delta r_i(a) \in \{-1, 1\}$.⁶

Let us now consider the implications of making $\Delta r_i(a)$ endogenous. Figure 3 again shows the net utility of two individuals with different individual evaluations with and without CQs.

⁶We assume that the reference group is not indifferent between the two actions ($r_i(E) \neq r_i(U)$), which rules out $\Delta r_i(a) = 0$. In addition, as our results do not depend on the existence of costs for motivated moral reasoning, we abstract from them here to simplify the model.

Figure 3: The effect of CQs through *Motivated Moral Reasoning*



Note. According to motivated moral reasoning individuals use CQs to rationalize their individual valuations and are thus more likely to choose in line with them.

For i , $\Delta r_i(a)$ is positive, implying that the introduction of CQs shifts the net utility of i to the right. However, since the individual already chose E before the introduction of CQs, there is no change in action. For j , on the contrary, $\Delta v_j(a)$ and therefore, due to motivated reasoning, $\Delta r_j(a)$ is negative, shifting Δu_j to the left. This can lead to a change from E to U .

Figure 3 thus shows that if individuals engage in motivated moral reasoning, more individuals act in line with their individual valuation. This results in the following testable hypothesis that contrasts Hypothesis 1.

Hypothesis 2. Motivated moral reasoning predicts that individuals reinforce their initial net utilities and are more likely to act in line with their individual valuations when being confronted with CQs.

4 Study 1

To test the effectiveness of CQs and our hypotheses, in Study 1, we place individuals in honesty-related workplace conflicts and study the effects of a mere exposure to CQs on ethical decision-making.

4.1 Design

4.1.1 Ethical Workplace Conflicts

To construct ethical conflicts for participants, we develop workplace vignettes encompassing ethical dilemmas. We thereby follow other literature in economics (Ambuehl and Ockenfels, 2017; Krawczyk et al., 2015; Gibson et al., 2013) and business ethics (reviewed in Mudrack and

Mason, 2013) using vignette-based designs to study ethical dilemmas. All vignettes are developed such that study participants take on the role of an employee facing an ethical conflict. The participant must then decide between two actions, one of which we designed to be ethical and one unethical. In our case, we consider the ethicality of an action in line with the definition of honesty such that lying, deception, or corruption is unethical. To incentivize the participant’s decision and encompass the fact that such unethical actions in organizations are often associated with monetary benefits (for example, through bonuses, promotions, or reputation), we follow Gibson et al. (2013) by including a financial incentive of €5 if a participant chooses the unethical action, while the ethical action only yields €1.⁷ As an example, in one of our vignettes, participants take on the role of a member of the sales team at a car manufacturer and must decide whether to sell premium accessories to a customer. Although the customer explicitly declared that they would like the cheapest configuration possible, more premium accessories are still within their budget. As the salesperson, the participant now faces the choice to either sell the cheapest configuration possible (ethical) or increase their own bonus by selling the more premium package (unethical). Screenshots of all vignettes are provided in Appendix C.

To capture a range of ethical tensions in workplaces, we previously conducted ten expert interviews and collected survey evidence ($N = 1,493$) on prevalent ethical challenges in the workplace, allowing us to develop a taxonomy of common conflict types (Witt et al., 2025). We distinguish between four types of conflict. First, situations where an unethical action directly translates into a financial benefit (*self*). Second, situations where an unethical action has already occurred by mistake and one can now revoke the action (*mistake*). Third, situations where another individual (e.g., a manager or colleague) encourages one to act unethically (*loyalty*) and lastly, a combination of the latter two, where another entity asks one not to revoke an unethical action that happened unintentionally (*mistake + loyalty*).⁸ See Table 1 for a summary of the four different types of conflict.

While keeping the incentives identical, we systematically vary our vignettes along these four dimensions to test the effectiveness of CQs in various types of workplace conflicts. In total, we developed four workplace scenarios (related to car accessories, overtime, faulty head-

⁷Although financial stakes are modest, the relative payoff structure aims to simulate real-world temptation asymmetries. Our vignette design follows the literature on dishonesty, where an unethical action (being dishonest) is financially beneficial to the participant without consequences to other participants (Benistant et al., 2022; Irlebusch et al., 2020; Gneezy et al., 2018; Fehrler et al., 2020; Fischbacher and Föllmi-Heusi, 2013; Gneezy et al., 2013; Friesen and Gangadharan, 2012; Shalvi et al., 2011; Sutter, 2009).

⁸The latter two encompass a dimension of hierarchy that has previously been shown to be relevant for honesty in groups (Rilke et al., 2018).

Table 1: Overview of conflict types

Self	Situations where unethical actions offer personal or organizational financial gain, creating a tension between integrity and personal gain.
Mistake	Situations where an unethical action occurred by mistake and individuals face the dilemma of preserving integrity by admitting it or protecting themselves by concealing it to avoid negative consequences.
Loyalty	Situations where individuals are encouraged or pressured by managers or colleagues to act unethically, creating a conflict between maintaining personal integrity and preserving loyalty or obedience within workplace relationships.
Loyalty + Mistake	Situations where an unintentional unethical action has occurred, and individuals face pressure from managers or colleagues not to revoke it, creating a conflict between maintaining integrity and preserving loyalty, obedience, or self-protection.

lights, environmental certificates; see Appendix C), each in four versions, varying in distinct conflict type as outlined above. This results in a total of 16 possible vignettes, combining workplace scenarios and conflict types. Within the study, each participant randomly faces four distinct combinations of conflict types and workplace scenarios (see Table 3). To increase involvement such that participants can immerse themselves in the vignettes, we display vignettes as interactive situations, consisting of short descriptions in story form. We also ask participants to select an avatar to represent themselves across all vignettes.⁹

4.1.2 Contemplation Questions

As outlined in Section 1, we identified a total of 727 CQs that appear in organizations.¹⁰ From this pool, we select 15 CQs, three for each of the five categories as described in Section 1 and in Renerte et al. (2025). The categories differ in the reference group they address and are labeled: *Self*, *Peers*, *Company Policy*, *Law & Reputation*, and *Universal Principles*. As these categories represent all identified CQs, these 15 CQs allow us to examine possible heterogeneity by categories and individual CQs (see Table 2).

Using a between-participant design, we assign each participant either to one of the 15 CQs or the control condition, where no CQ is present. Because 78% of companies present CQs in a list format that employees can read but are not required to answer (unlike, for example,

⁹Similar to Tergiman and Villeval (2023), in a pilot, we also varied the risk of detection within our study. We found no interactions between detectability and the effectiveness of contemplation questions. This is why we do not include this dimension in our design. Appendix F reports details.

¹⁰Many CQs such as “Is the behavior legal?” occur multiple times. The list of 727 thus includes duplicates.

flowcharts), we begin by studying this common implementation. In line with the practice observed in firms that most often simply list CQs as moral primes in codes of conduct, on posters, or in handbooks (Renerte et al., 2025), we present the CQ before the final decision situation and simply ask participants “to read through the question before making their final decision.” Since we want to study each CQ individually, we avoid contamination by always assigning a unique (or no) CQ per participant in all four vignette decisions. Table 3 provides an example of how we combine vignettes and CQs in the study. To determine the final bonus payments, we randomly select one of the four vignettes at the end of the study and pay the bonus according to the taken decision (€1 for ethical, €5 for unethical).

4.2 Procedures

We pre-registered the study (see <https://doi.org/10.17605/OSF.IO/GK2ER>) and received ethical approval from the University of Zurich. The interface was programmed using Qualtrics and

Table 2: Overview of all studied contemplation questions

Self (1 - 3)
1. “Would I have a bad conscience about the decision?”
2. “Would I like the person I see in the mirror after I have made the decision?”
3. “Am I sure that I will not regret the decision later?”
Peers (4 - 6)
4. “Do my peers accept my decision?”
5. “Would I be proud if my family and friends found out about my decision?”
6. “Would my closest friends advise it?”
Company Policy (7 - 9)
7. “Is my decision in line with our company policy?”
8. “Is my decision in line with our company values?”
9. “Does my behavior reflect the culture of my company?”
Law & Reputation (10 - 12)
10. “Would the behavior be legal?”
11. “Will the reputation of our company be damaged if my decision is made public?”
12. “Would the public approve of my decision if they read about it in the newspaper?”
Universal Principles (13 - 15)
13. “Do I treat others the way I would like to be treated?”
14. “Do I act honestly and sincerely?”
15. “Could I take responsibility for my decision?”

Table 3: Example of vignettes and CQ combination

Conflict type	Workplace scenarios				CQ
	Car accessories	Overtime	Faulty headlights	Environmental certificates	
Self	X				“Would I have a bad conscience about the decision?”
Mistake			X		
Loyalty		X			
Mistake + Loyalty				X	

Note. The table depicts an example for a single participant. In the study, the combination of conflict types and workplace scenarios, as well as the CQ was randomized for each participant.

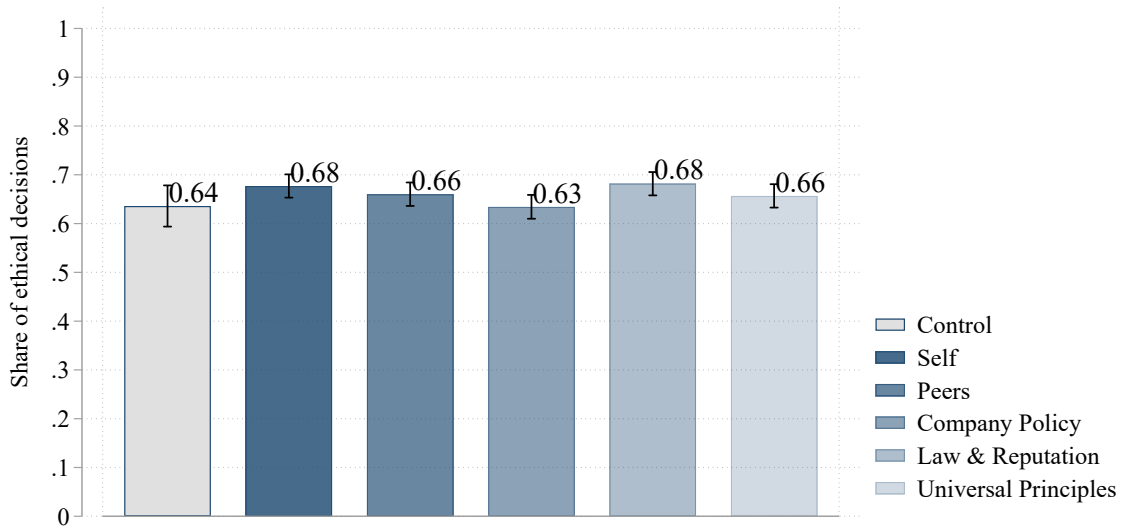
we ran the study in June 2023 using the participant panel provider Bilendi.¹¹ Of the 2,669 participants who completed the first-wave personality survey (more detail below), 83% ($N = 2,208$) completed the main study. We excluded the pre-registered 10% (222) of participants due to slow or fast responses or failure of the attention check. In total, this left us with a sample of 1,986 participants for our main analysis. However, all results also hold when considering the full sample.¹² As we are interested in workplace conflicts, we only recruited German employees aged 18 to 67 who have been employed for at least six months during the last two years, excluding self-employed participants. Our final sample consists of 43% women and 57% men, and none did not declare their gender. The average age is 49.1 years (median: 51.0). In addition to our main variable of interest, the share of ethical actions, we also elicited measures of stress, decision difficulty (both [Hanselmann and Tanner, 2008](#)) and decision avoidance ([Janis and Mann, 1977](#)) for each decision. In a separate data collection two weeks prior to our study, we collected additional demographic information, as well as data on moral identity ([Black and Reynolds, 2016](#)), protected values ([Gibson et al., 2013](#)), six personality traits (HEXACO, [De Vries, 2013](#)), need for cognition ([Beißert et al., 2014](#)), self-efficacy ([Hannah and Avolio, 2010](#)), risk- and social preferences ([Falk et al., 2023](#)).¹³

¹¹See Appendix Tables A1 and A2, for a breakdown of participants across treatments, conflict types and workplace situations.

¹²We excluded participants in the slowest and fastest 5th percentile, as well as participants indicating that they did not pay attention during the study.

¹³We slightly modified the stress, decision certainty, decision avoidance and self-efficacy scales to better match our context. All utilized scales can be found in Appendix B.

Figure 4: The effect of CQ categories on ethical decision-making



Note. The control includes all participants that did not see a CQ. Each of the other bar comprises three CQs corresponding to the respective category. Bars indicate 95% confidence intervals.

4.3 Results

Overall, we find that 66% of participants choose the ethical action. 80% of participants act ethically in vignettes that only give themselves an additional payoff (*self*) whereas only 47% of participants choose the ethical action in vignettes where an unethical action had already been taken and can now be revoked (*mistake*). In *loyalty* and *mistake + loyalty* conflicts, the share of ethical actions lies in between (*loyalty*: 71%; *mistake + loyalty*: 63%). Thus, despite financial incentives to act unethically without consequences for others, people often choose the honest option, broadly in line with the literature on preferences for truthfulness (Abeler et al., 2019; Gibson et al., 2013). We also find heterogeneity across vignettes, as 78% of the participants do not take the same action in all vignettes.¹⁴ This highlights the importance of context in our vignettes, since they all have the exact same financial incentives.

Turning to the effectiveness of CQs, on aggregate, 64% of the participants in the control condition choose the ethical action. Among all participants that were confronted with a CQ, this share is 66%, thus slightly higher but not statistically different ($z = 1.34$, $p = 0.182$). These results are based on a random-effect probit regression with the ethical choice as the dependent, binary variable and a CQ present indicator as the independent variable (see Appendix Table

¹⁴Whilst we do observe this heterogeneity, we cannot find any evidence of wealth effects or moral licensing. There are no statistical trends in the data between the first and last decisions of individuals.

A3 for details). Since each participant makes multiple decisions, we estimate random-effects probit models with an individual-specific intercept. Randomization of treatment assignment at the individual level makes the random-effects exogeneity assumption plausible. As a robustness check, for all subsequent regression analyses we also re-estimate our specifications using our pre-registered pooled probit with standard errors clustered at the individual level. The results are virtually identical and available in Appendix A.1. Figure 4 presents this result for each of the five different categories of CQs, showing no statistically significant differences. In addition, Table 4 shows the effects by individual CQ instead of categories. The coefficients present marginal effects from random-effect probit regressions of ethical decisions on CQ-dummies as treatment indicators.

In each of the four regression models (1)–(4), 13–14 of the 15 CQs (87–93%) show a positive coefficient, but only one CQ regarding *family and friends* is marginally significant at the 5% level ($z < 2.04$, $p > 0.041$). Moreover, when examining other outcome variables such as stress, decision difficulty, or decision avoidance, we also find no effects of CQs ($t < 1.33$, $p > 0.185$, based on linear regression models including random effects; see Appendix Table A4; identical using our pre-registered Tobit specification, see Appendix Table A13). The effects on ethical decisions, as well as other outcome variables, are robust across conflict types, workplace scenarios, and to the inclusion of individual controls.¹⁵ Our results suggest that CQs do not affect ethical decision-making when implemented as moral primes, which is the most commonly observed practice in companies.

When examining demographic characteristics, we find that older participants ($z = 6.18$, $p < 0.001$) are more likely to choose the ethical action. Moreover, individual moral identity predicts ethical actions ($z = 4.83$, $p < 0.001$).¹⁶ Taking moral identity as a proxy for individual valuations, $\Delta v_i(a)$, our findings are in line with the conceptual framework. Both possible mechanisms – moral agency as well as motivated moral reasoning – predict that a *moral compass* (proxied by moral identity) predicts ethical decision-making in general.

Implication 1. The mere exposure to CQs does not affect ethical decision-making.

Overall, Study 1 examines an implementation in which participants are presented with a CQ but are not required to answer it. Under this implementation, we find no average effect on

¹⁵We also tested a *pure deliberation* treatment as in Gunia et al. (2012) in a pilot study in which we asked participants to take time and reflect without CQ guidance prior to their decision, but we did not find a causal effect of this treatment relative to the absence of reflection on ethical decisions ($z = 0.89$, $p = 0.376$).

¹⁶Complete regression tables are provided in Appendix Tables A3 and A4.

Table 4: The effect of individual CQs on ethical decision-making

	(1)	(2)	(3)	(4)
Self				
Bad conscience	0.046 (0.031)	0.047 (0.031)	0.047 (0.031)	0.051 (0.029)
Mirror	0.041 (0.031)	0.042 (0.031)	0.042 (0.031)	0.050 (0.029)
No regret	0.049 (0.031)	0.050 (0.031)	0.050 (0.031)	0.045 (0.029)
Peers				
Personal network	0.022 (0.031)	0.023 (0.031)	0.023 (0.031)	0.021 (0.029)
Family and friends	0.060* (0.030)	0.061* (0.030)	0.061* (0.030)	0.058* (0.029)
Closest friends	-0.001 (0.031)	0.000 (0.031)	0.000 (0.031)	0.011 (0.029)
Company policy				
Company guidelines	-0.040 (0.032)	-0.039 (0.031)	-0.039 (0.031)	-0.042 (0.030)
Company values	0.036 (0.031)	0.036 (0.031)	0.036 (0.031)	0.048 (0.029)
Company culture	0.014 (0.031)	0.014 (0.031)	0.014 (0.031)	0.028 (0.029)
Law & Reputation				
Legal	0.046 (0.031)	0.047 (0.031)	0.047 (0.031)	0.043 (0.029)
Reputation	0.056 (0.031)	0.057 (0.031)	0.057 (0.031)	0.055 (0.029)
Newspaper	0.056 (0.031)	0.057 (0.031)	0.057 (0.031)	0.056 (0.029)
Universal Principles				
Others as self	0.032 (0.031)	0.032 (0.031)	0.032 (0.031)	0.029 (0.029)
Honesty and integrity	0.016 (0.031)	0.017 (0.031)	0.017 (0.031)	0.017 (0.029)
Accountable	0.019 (0.031)	0.020 (0.031)	0.020 (0.031)	0.026 (0.029)
Workplace scenario FEs		X	X	X
Dilemma type FEs			X	X
Individual controls				X
Observations	8832	8832	8832	8832
Individuals	2,208	2,208	2,208	2,208

Note. All models present marginal effects of random-effect probit regressions. The dependent variable is the binary decision to choose the ethical action. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits and risk preferences. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 5: Overview of contemplation questions tested in Study 2

Self	“Would I have a bad conscience about the decision?”
Peers	“Would I be proud if my family and friends found out about my decision?”
Company Policy	“Is my decision in line with our company values?”
Law & Reputation	“Would the public approve of my decision if they read about it in the newspaper?”
Universal Principles	“Do I treat others the way I would like to be treated?”

ethical decision making. While moral identity remains predictive of ethical decision-making in Study 1, its interaction with CQ presence is positive but does not reach conventional significance ($z = 1.33$, $p = 0.184$, on the interaction). We next examine a second implementation in which participants answer the CQ before deciding.

5 Study 2

5.1 Design

In Study 2, participants are required to answer the CQ before making their final decision, rather than being presented with it without having to respond. [Renerte et al. \(2025\)](#) find that 22% of companies provide CQs as flow charts or checklists, which encourage employees to actively answer a question before moving on to the next question. Our second study captures this use of CQs. More specifically, we ask participants to name which action (ethical or unethical) best fits a specific CQ, e.g., “*for which action would you have a bad conscience after you made the decision?*” The final decision itself remains a binary choice between the ethical and the unethical action, as in Study 1, and is made on a separate screen directly after the CQ; screenshots of both the CQ-answer and the decision screens are provided in [Appendix C](#).

In addition, we now focus on a single CQ per category, as we did not detect significant differences of individual CQs within categories in Study 1 ($\chi^2 < 1.42$, $p > 0.234$ for all positive coefficients in [Table 4](#)). We select the CQ with the largest effect size in each category based on the marginal coefficients depicted in [Table 4](#), column 4, and list the resulting five CQs in [Table 5](#). In addition to testing these five CQs individually, we also include one treatment arm in which all five CQs are shown simultaneously.

Further, in Study 2, we move from a between-participant design to a within-participant set-up, administered across two waves. This way we can study individual decisions in control

as well as treatment conditions, allowing us to examine switches and gain more insight into the role of personality traits. Because moral agency and motivated moral reasoning are distinguished by how an individual's moral reference standard responds to the introduction of CQs, observing the same participants with and without CQs holds individual characteristics fixed and is better suited to examining these mechanisms than a purely between-participant comparison. In Wave 1, all participants are part of the control condition and make four decisions in the absence of CQs. Two weeks later, in Wave 2, they then make another four decisions, seeing either one of the five single CQs or all five CQs simultaneously. In addition, a subset of participants is assigned to a control group that faces no CQ in Wave 2. This within-study control allows us to gauge any other influences on ethical behavior between the two waves, such as learning or repeated exposure to the vignettes. Across both waves, we use eight vignettes from Study 1 and split these into two sets of four parallel vignettes. Each set includes four different workplace scenarios with four different conflict types. The participants were then randomly presented with the one set in Wave 1 and the other set in Wave 2. In the analysis, we then match vignettes based on conflict types across waves.¹⁷

5.2 Procedures

We pre-registered this study (<https://doi.org/10.17605/OSF.IO/G3EWA>) and received ethical approval from the University of Zurich.¹⁸ We used Qualtrics for programming and followed the same inclusion criteria as Study 1, and ran the study in December 2023 using the Bilendi panel with around two weeks between Waves 1 and 2. Of the 2,212 participants who completed Wave 1, 81% ($N = 1,784$) completed Wave 2. Of those completing both experimental waves, 1,646 received a contemplation question in Wave 2, while 138 were assigned to the no-CQ control group. After excluding the pre-registered 12% (196) of the CQ participants due to fast or slow aggregate completion time across both waves or failed attention checks, our main analysis sample comprises $N = 1,450$ participants (alongside $N = 125$ in the control group). Again, our results hold when including the entire sample. The final sample composition is similar to Study 1 with slightly more women (53%) than men (47%). Again, everyone declared their gender. The average age is 45.1 (median: 45.0). Furthermore, since we saw a strong inter-item correlation (Cronbach's $\alpha = 0.93$) between our measures of stress, decision difficulty, and

¹⁷The vignettes assigned to the two sets are provided in Appendix C. We also provide a description of why we selected this subset of vignettes and scenarios for Study 2.

¹⁸As specified in the pre-registration, we also tested positive and negative framing of CQs. However, since our analysis did not reveal differential effects of the frames, we pool our data across both frames. Appendix Table A10 reports the effect of CQs separately for each frame.

decision avoidance in Study 1, we now collapsed them into one scale with five items, which we refer to as “decision confidence” and administered at the end of each wave.¹⁹ In addition, since we were interested in whether CQs affect the level of moral disengagement, we administered a state-moral disengagement scale after each wave (Barsky, 2011). As pre-registered, we asked participants at the end of Wave 2 to recall the CQ they had seen. Again, we collected data on moral identity (Black and Reynolds, 2016), six personality traits (HEXACO, De Vries, 2013), risk- (Falk et al., 2023) and social preferences (Murphy et al., 2011), as well as emotional state (PANAS, Krohne et al., 1996) in a separate third wave between three to six weeks after Wave 2. Demographic information is available for all participants. Within our main analysis sample, these dispositional measures are available for 88% ($N = 1,274$) treated participants and for 93% ($N = 116$) pure control participants, and analyses that use them are restricted accordingly.

5.3 Results

Even though participants now actively answer the CQs, we replicate the results of Study 1 and again find, no average effect of CQs on the share of ethical decisions ($z = -1.81$, $p = 0.070$, based on a probit regression with random effects). Within our framework, one possible interpretation is that actively answering the CQs increases their salience, corresponding to a larger λ . This result goes against our Hypothesis 1 regarding moral agency and suggests that, on average, individuals do not shift their net utilities in favor of the ethical action, E .

However, depending on the distribution of net utilities in our sample, this overall null effect can still be a consequence of motivated moral reasoning, as outlined in our Hypothesis 2.²⁰ To test this more thoroughly, we rely on our evidence from Study 1 and utilize moral identity as a proxy for an individual’s *moral compass* ($\Delta v_i(a)$).²¹ Moral identity is a widely established concept in the psychological literature (Aquino and Reed II, 2002; Reed II and Aquino, 2003) that, as in Study 1, has been frequently shown to predict individual moral behavior (see e.g., Hardy et al., 2013; Hertz and Krettenauer, 2016; Winterich et al., 2013). This is also the case in Study 2, where moral identity is a strong and statistically significant predictor of ethical decision-making ($z = 4.12$, $p < 0.001$, see Appendix Table A5).

More importantly, we find that moral identity serves as a moderator in the effectiveness

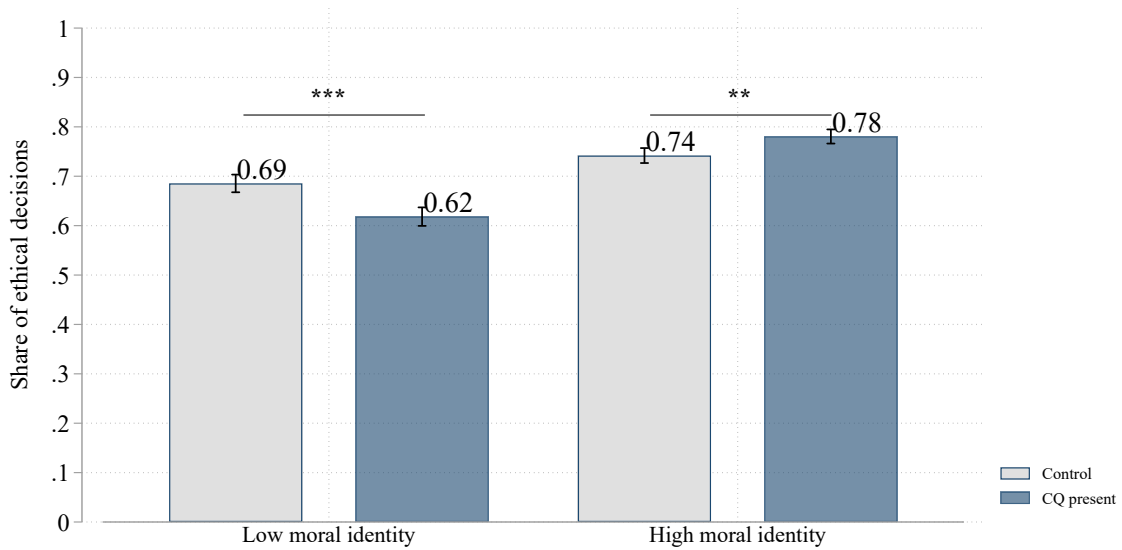
¹⁹The new scale with five items is provided in Appendix B.

²⁰If there are no individuals where the decision noise passes the indifferent point, motivated moral reasoning leads to stronger net utilities in either direction, but does not affect the share of individuals choosing E or U .

²¹Our pre-registration specified splitting the sample by participants’ pro-ethical versus pro-unethical responses to the CQs. Because these responses are endogenous to the decision-making process, we instead moderate by moral identity, which we had pre-registered as an exploratory moderator and which is significantly correlated with these responses ($r = 0.19$, $p < 0.001$). The pre-registered response-based split is reported in Appendix E.3.

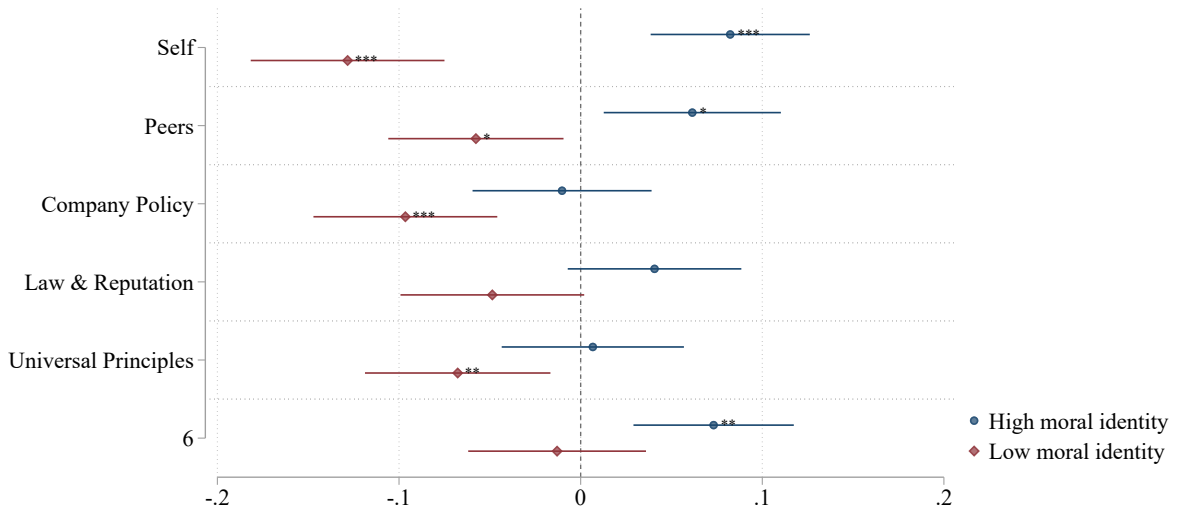
of CQs ($z = 7.89$, $p < 0.001$ on the interaction, see Appendix Table A5). Figure 5 presents the share of ethical actions with and without CQs split for low and high moral identity (based on a median split). First, according to the concept of a *moral compass* in the absence of CQs, participants with a high moral identity choose the ethical action 6pp more often than those with a low moral identity ($z = 3.70$, $p < 0.001$). Second, when we split the sample by moral identity (median split), we observe heterogeneous effects of CQs, where for individuals with low moral identity (left), CQs significantly reduce the probability of choosing ethical action (Appendix Table A5, $z = -5.85$, $p < 0.001$), whereas for individuals with high moral identity (right) there is a positive and significant effect (Appendix Table A5, $z = 3.64$, $p < 0.001$) of CQs on ethical decisions. We further find that this heterogeneity is robust when, instead of moral identity, we examine other elicited dispositional traits such as honesty-humility (Appendix Table A6, $z = 5.60$, $p < 0.001$) or social value orientation (Appendix Table A6, $z < -3.45$, $p < 0.001$). This suggests that any proxy of a *moral compass* can unveil these heterogeneous effects of CQs. These category-level effects are robust to multiple-hypothesis testing. Appendix Table A8 reports Benjamini-Hochberg false-discovery-rate-adjusted p-values for all category effects across the three proxies, and the moral-identity heterogeneity remains significant after

Figure 5: Effects of CQs across moral identity



Note. The control consists of decisions in Wave 1 (gray bar), where participants did not see a CQ. In the CQ present bar (blue), we aggregate all participants that were in one of the six CQ treatment conditions. On the left, we compare participants with low moral identity (Black and Reynolds, 2016); on the right, individuals with high moral identity based on a median split. Significance levels in the figure are based on Wilcoxon rank sum test and regression results are presented in Appendix Table A5. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Bars indicate 95% confidence intervals.

Figure 6: Moral identity moderates the effectiveness of CQs



Note. We present marginal effects of individual CQs from one random-effects probit regression. The dependent binary variable is whether participants made an ethical choice. We present interactions with a high moral identity dummy in blue and with a low moral identity dummy in red. Coefficients capture the effect relative to the control condition (Wave 1) in the absence of CQs. We include situation and vignette fixed effects, as well as age, gender, social value orientation, HEXACO as additional individual controls. We also control whether an individual has a manager, is in a managerial role and how long they have been employed. Regression results are presented in Appendix Table A7. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Bars indicate 95% confidence intervals.

correction. Due to the high predictive power of moral identity for ethical decision-making in the aggregate results, we will, in the following, focus on this dimension as a moderator. These results are in line with motivated moral reasoning as they indicate a “boomerang effect” of CQs for individuals with low moral identity and amplify the difference between people with low and high moral identity.

Including moral identity as a moderator, we next examine different categories of CQs (see regression results in Appendix Table A7). In Figure 6, we see that the differential effects described above are in fact driven mainly by CQs referring to the *Self* and *Peers*, which could be considered more relevant for an individual’s identity (Aquino and Reed II, 2002; Tajfel, 1979).

In our study, the unintended consequences of CQs for individuals with low moral identity are stronger than the beneficial effects for individuals with high moral identity. For example, our results suggest that on average, for individuals with low moral identity, the *Self* CQ reduces the share of ethical actions by 13pp, whilst it increases this share by 8pp for individuals with high moral identity. Even more notably, CQs addressing *Company policy* and *Universal Principles* only show significant negative effects on the probability of ethical actions for individuals with low moral identity, without any significant positive effects for those with high

moral identity. This suggests that using CQs in practice might backfire and promote unethical behavior. In Appendix Figure A1 we replicate this finding using honesty-humility and social value orientation as moderators, further demonstrating the robustness of our results across different proxies of a moral compass.

We interpret these findings as consistent with motivated moral reasoning (Hypothesis 2), under which individual moral valuations ($v_i(a)$) are strengthened through interaction with CQs, rather than as direct evidence of the mechanism. This interpretation is consistent with the observed separation between individuals with low and high moral identity, which constitutes our second main result and implication.

Additional, suggestive evidence consistent with this mechanism comes from a separate pilot study for Study 2 in which participants justified their choices in open text. We find that justifications given for the unethical action draw disproportionately on cost-dissolving rationales (e.g., denying harm or citing a superior’s instruction), whereas those for the ethical action rest on more stable standards (see Appendix A.2). As the pilot did not vary CQs within participants, we regard this pattern as suggestive, and therefore cannot establish that CQs causally elicit or strengthen such justifications.

Implication 2. When participants are required to answer them, CQs can lead to unintended “boomerang effects” for individuals with low moral identity.

The moral-identity moderation is precisely estimated in Study 2 but imprecisely estimated in Study 1. However, a formal cross-study test cannot statistically distinguish the moderation estimates (study $\times$ CQ $\times$ moral-identity interaction, $p = 0.35$). Because the studies also differ in design and in the CQs tested, we do not interpret this pattern as evidence that required answering or greater engagement causes the moderation.

Finally, we also test what happens if we present five CQs simultaneously, pre-registered as an exploratory analysis. Why would the effect of five CQs differ from that of a single CQ? We do not model multiple CQs in our conceptual framework above. One possible explanation, which our design does not allow us to test directly, is that there is a psychological cost associated with motivated moral reasoning (Ditto and Lopez, 1992; von Hippel and Trivers, 2011): individuals might be able to justify unethical actions with respect to one reference group, but fail to do so with respect to multiple. This could result in more individuals choosing the ethical action E . We present the results of this treatment in the last row of Figure 6. We now indeed observe a significant effect of multiple CQs on the share of ethical actions among individuals with high moral identity ($z = 2.49$, $p = 0.013$, see Appendix Table A7) and even more im-

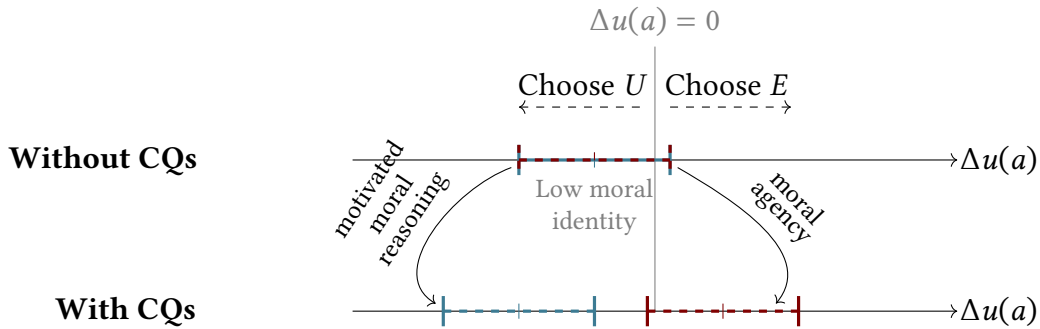
portantly, no longer observe a significant “boomerang effect” for individuals with low moral identity ($z = -0.44$, $p = 0.663$, see Appendix Table A7). Importantly, among individuals with low moral identity, the effect of multiple CQs is significantly less negative than the effect of a single CQ (a difference of 7.2pp, $p = 0.010$). Among individuals with high moral identity, by contrast, the effects of multiple and single CQs do not differ significantly ($p = 0.123$). These results are robust to order effects in the presentation of the five CQs (Appendix Table A9).

Implication 3. Multiple CQs mitigate the negative effects of CQs and may promote ethical decision-making for individuals with high moral identity.

5.4 Decomposition of individual types

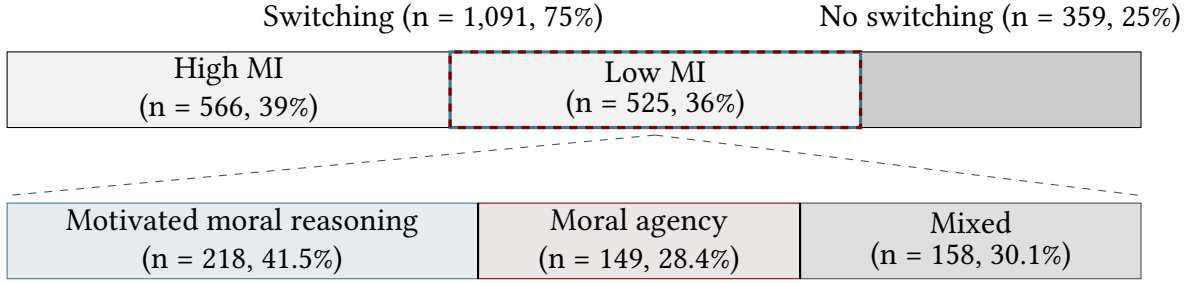
In general, our results are in line with motivated moral reasoning. In practice, however, some participants may engage in motivated moral reasoning while others act according to moral agency. Although our design does not allow us to disentangle mechanisms for every participant, we can decompose individual types by examining how each participant’s choices change between the two waves. We conduct this decomposition at the participant level, tracking for each individual how frequently and in which direction their decisions switch. Note that this constitutes a post-hoc analysis, which is not part of the pre-registration. Of the 1,450 participants, 359 (25%) do not change any of their decisions between waves; since these participants are consistent in their choices, we are unable to say more about the mechanisms of CQs. Among the remaining 1,091 (75%) who switch at least one decision, the 566 (39%) with high moral identity are uninformative for distinguishing mechanisms, because both moral agency and motivated moral reasoning imply a stronger inclination towards the ethical action when $\Delta u_i(a) > 0$. This leaves 525 (36%) participants with low moral identity who switch at least

Figure 7: Conceptual decomposition of individual types



Note. The figure depicts two examples for the intuitive net utility of two individuals i (red), and j (blue). Individual i has $\Delta u_i(a) > 0$ (red) and j has $\Delta u_j(a) < 0$ (blue).

Figure 8: Decomposition results



Note. Switching refers to individuals who change at least one of their ethical decisions between Wave 1 (without CQ) and Wave 2 (with CQ); percentages in the top bar are shares of all participants. We decompose individuals with low moral identity (*MI*) who switch into those switching consistently in line with motivated moral reasoning (blue), consistently in line with moral agency (red), and those exhibiting both patterns (gray); percentages in the bottom bar are shares of low moral identity switchers.

once and whose behavior we can decompose.

Figure 7 illustrates our decomposition exercise with a hypothetical example. Consider two individuals (red and blue) with identical net utilities, both initially favoring the unethical action (low moral identity). Due to differences in their decision error, ϵ_i , individual red chooses the unethical action whilst individual blue chooses the ethical action in absence of CQs in Wave 1. Once confronted with CQs in Wave 2, suppose that blue engages in motivated moral reasoning. This shifts their net utility to the left, and they switch actions from ethical to unethical. In contrast, red acts in line with moral agency and therefore shifts their net utility to the right and switches from the unethical to the ethical action. Thus, for participants that switch behavior and have low moral identity, we can differentiate between the two underlying mechanisms as follows: If an individual changes from *E* to *U*, this behavior is in line with motivated moral reasoning. If, in contrast, an individual changes from *U* to *E*, it is in line with moral agency.

Using this definition, we classify each of the 525 low moral identity participants who switch by the direction of their switches. Those whose switches are exclusively from *E* to *U* are consistent with motivated moral reasoning, those whose switches are exclusively from *U* to *E* are consistent with moral agency, and those who switch in both directions across their decisions exhibit a mixed pattern. We find that 218 (41.5%) switch consistently in line with motivated moral reasoning, 149 (28.4%) consistently in line with moral agency, and 158 (30.1%) exhibit both patterns. Figure 8 visualizes this participant-level decomposition across the whole sample. Thus, roughly 70% of switchers switch in only one direction, and among these motivated moral reasoning is the modal pattern (59% versus 41%). At the same time, the 30% who switch

in both directions are consistent with the two mechanisms operating within the same individual, or with idiosyncratic noise in individual decisions, which we examine next. Overall, while we observe more behavior in line with motivated moral reasoning than moral agency, the decomposition suggests substantial levels of heterogeneity between individuals.²²

Some switches may reflect idiosyncratic noise rather than a genuine change in an individual's reference standard. Control participants, who experience no change in CQ exposure, switch symmetrically across waves (binomial $p = 0.27$), whereas the consistent switches among low moral identity participants are skewed toward the unethical action (59% versus 41%). The participant-level decomposition thus documents seemingly systematic switching patterns, but does not by itself identify the underlying mechanism, since vignette-specific variation and decision noise may also generate one-directional switches.

6 Discussion and Concluding Remarks

Organizations frequently utilize CQs (contemplation questions) in an attempt to promote ethical decision-making and encourage employees to reflect on their actions. Presumably, they particularly target so-called “bad apples”, those (unknown) employees with a higher propensity for unethical behavior (Kish-Gephart et al., 2010). Our results indicate that a *single* contemplation question can inadvertently produce a “boomerang effect” for this group, increasing the likelihood that individuals with lower moral identities act unethically. This pattern is consistent with motivated moral reasoning, where individuals use the CQ to justify an already preferred unethical action. This backfiring is specific to isolated CQs rather than a general property of the tool. When we instead present *multiple* CQs that invoke different reference standards, the negative effect for individuals with low moral identity disappears (Implication 3), plausibly because justifying an unethical choice against several standards at once is considerably harder. For practice, the message is that the effect of CQs depends on how they are used: a single, isolated CQ can backfire for the very employees it is meant to reach, whereas a set of CQs spanning several standards mitigates this risk.

Our study has some limitations, as we use stylized vignettes to capture honesty-related workplace conflicts. Studying the effectiveness of CQs within organizations would be the gold standard, but ethical conflicts and unethical actions are often not observable until they lead to a scandal. Our experimental setup allows us to “peek behind the curtain” and examine

²²When examining response times in ethical decisions, we also find a large variation across individuals, but not systematic patterns that correlate with individuals characterized by motivated moral reasoning.

situations that are impossible or at least difficult to observe within firms ([Gangadharan et al., 2020](#)). In addition, our design does not identify engagement as the mechanism behind the moderation observed in Study 2. A direct test would randomly assign participants within the same study to read versus answer the same CQs and verify that the manipulation changes engagement. We leave this test to future research.

Our findings also have direct implications for practitioners. In particular, they caution against deploying single, isolated CQs as simple moral primes, for example a lone question on a poster or in a code of conduct, since these are the prompts that individuals with a weaker moral compass can use to rationalize unethical choices. If organizations nonetheless wish to use CQs, our results suggest presenting them as a short checklist that spans several reference standards rather than as a single question, consistent with the more benign effects we observe for multiple CQs (Implication 3). More broadly, the prominence of CQs within the world's largest companies is noteworthy given our results. There may be several reasons why firms adopt them so readily. First, anecdotal evidence suggests that large companies, particularly in the area of ethics and compliance, often adopt "copy-and-paste" approaches to new initiatives, driven by a fear of falling behind competitors ([Taylor, 2024](#)). Second, firms may prioritize the signaling value of such initiatives, for example to attract the right talent, over their actual direct impact, which our study does not evaluate. However, even if virtue signaling is a key motive, the popularity of CQs in global corporations suggests that practitioners do not expect negative consequences of such aids. Our study therefore shows the importance of not only carefully designing behavioral interventions, but also rigorously testing them before implementation, as even the best-intentioned efforts can lead to unintended consequences. Future research should also explore how CQs might be redesigned or supplemented to avoid reinforcing biased justifications.

References

- Abeler J, Nosenzo D, Raymond C (2019) Preferences for truth-telling. *Econometrica* 87(4):1115–1153, URL <http://dx.doi.org/10.3982/ECTA14673>.
- Akerlof GA, Kranton RE (2000) Economics and identity. *The Quarterly Journal of Economics* 115(3):715–753, URL <http://dx.doi.org/10.1162/003355300554881>.
- Ambuehl S, Ockenfels A (2017) The ethics of incentivizing the uninformed: A vignette study. *American Economic Review* 107(5):91–95, URL <http://dx.doi.org/10.1257/aer.p20171109>.
- Aquino K, Reed II A (2002) The self-importance of moral identity. *Journal of Personality and Social Psychology* 83(6):1423–1440, URL <http://dx.doi.org/10.1037/0022-3514.83.6.1423>.
- Bandura A (1986) *Social Foundations of Thought and Action: A Social Cognitive Theory* (Englewood Cliffs, NJ: Prentice-Hall), ISBN 9780138156145.
- Bandura A (1991) Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes* 50(2):248–287, URL [http://dx.doi.org/10.1016/0749-5978\(91\)90022-L](http://dx.doi.org/10.1016/0749-5978(91)90022-L).
- Bandura A (2006) Toward a psychology of human agency. *Perspectives on Psychological Science* 1(2):164–180, URL <http://dx.doi.org/10.1111/j.1745-6916.2006.00011.x>.
- Bandura A (2011) Moral disengagement. Christie DJ, ed., *The Encyclopedia of Peace Psychology* (Wiley-Blackwell), URL <https://doi.org/10.1002/9780470672532.wbepp140>.
- Barends E, Rousseau DM, Briner RB (2014) Evidence-based management: The basic principles. URL <https://cebma.org/assets/Uploads/Booklet-The-Basic-Principles.pdf>, accessed on 2020-02-20.
- Barsky A (2011) Investigating the effects of moral disengagement and participation on unethical work behavior. *Journal of Business Ethics* 104(1):59–75, URL <http://dx.doi.org/10.1007/s10551-011-0891-6>.
- Bateson M, Nettle D, Roberts G (2006) Cues of being watched enhance cooperation in a real-world setting. *Biology Letters* 2(3):412–414, URL <http://dx.doi.org/10.1098/rsbl.2006.0509>.
- Batson CD, Thompson ER, Seufferling G, Whitney H, Strongman JA (1999) Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology* 77(3):525–537, URL <http://dx.doi.org/10.1037/0022-3514.77.3.525>.

- Bazerman MH, Sezer O (2016) Bounded awareness: Implications for ethical decision making. *Organizational Behavior and Human Decision Processes* 136:95–105, URL <http://dx.doi.org/10.1016/j.obhdp.2016.05.003>.
- Bazerman MH, Tenbrunsel AE (2011) *Blind Spots: Why We Fail to Do What's Right and What to Do About It* (Princeton University Press), ISBN 9780691147505.
- Beißert H, Köhler M, Rempel M, Beierlein C (2014) Eine deutschsprachige Kurzsкала zur Messung des Konstrukts Need for Cognition: Die Need for Cognition Kurzsкала (NfC-K). *Diagnostica* 60(4):243–256, URL <http://dx.doi.org/10.1026/0012-1924/a000124>.
- Bellemare C, Deversi M, Englmaier F (2019) Complexity and distributive fairness interact in affecting compliance behavior. Technical Report 7899, CESifo, URL https://EconPapers.repec.org/RePEc:ces:ceswps:_7899.
- Benistant J, Galeotti F, Villeval MC (2022) Competition, information, and the erosion of morals. *Journal of Economic Behavior & Organization* 204:148–163, URL <http://dx.doi.org/10.1016/j.jebo.2022.10.008>.
- Bernheim BD (1994) A theory of conformity. *Journal of Political Economy* 102(5):841–877, URL <http://dx.doi.org/10.1086/261958>.
- Beshears J, Choi JJ, Laibson D, Madrian BC, Milkman KL (2015) The effect of providing peer information on retirement savings decisions. *The Journal of Finance* 70(3):1161–1201, ISSN 0022-1082, URL <http://dx.doi.org/10.1111/jofi.12258>.
- Black JE, Reynolds WM (2016) Development, reliability, and validity of the moral identity questionnaire. *Personality and Individual Differences* 97:120–129, URL <http://dx.doi.org/10.1016/j.paid.2016.03.075>.
- Bursztyn L, Fiorin S, Gottlieb D, Kanz M (2019) Moral incentives in credit card debt repayment: Evidence from a field experiment. *Journal of Political Economy* 127(4):1641–1683, ISSN 0022-3808, URL <http://dx.doi.org/10.1086/701605>.
- Bénabou R, Tirole J (2011) Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics* 126(2):805–855, URL <http://dx.doi.org/10.1093/qje/qjr009>.
- Bénabou R, Tirole J (2016) Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives* 30(3):141–164, URL <http://dx.doi.org/10.1257/jep.30.3.141>.

- Charness G, Dave C (2017) Confirmation bias with motivated beliefs. *Games and Economic Behavior* 104:1–23, URL <http://dx.doi.org/10.1016/j.geb.2017.02.015>.
- Cohn A, Gesche T, Maréchal MA (2022) Honesty in the digital age. *Management Science* 68(2):827–845, URL <http://dx.doi.org/10.1287/mnsc.2021.3985>.
- Cohn A, Maréchal MA, Tannenbaum D, Zünd CL (2019) Civic honesty around the globe. *Science* 365(6448):70–73, URL <http://dx.doi.org/10.1126/science.aau8712>.
- Damgaard MT, Gravert C (2018) The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics* 157:15–26, ISSN 0047-2727, URL <http://dx.doi.org/10.1016/j.jpubeco.2017.11.005>.
- Dana J, Weber RA, Kuang JX (2007) Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory* 33(1):67–80, URL <http://dx.doi.org/10.1007/s00199-006-0153-z>.
- De Vries RE (2013) The 24-Item Brief HEXACO Inventory (BHI). *Journal of Research in Personality* 47(6):871–880, URL <http://dx.doi.org/10.1016/j.jrp.2013.09.003>.
- Desai SD, Kouchaki M (2017) Moral symbols: A necklace of garlic against unethical requests. *Academy of Management Journal* 60(1):7–28, URL <http://dx.doi.org/10.5465/amj.2014.0341>.
- Dhami MK (2003) Psychological models of professional decision making. *Psychological Science* 14(2):175–180, URL <http://dx.doi.org/10.1111/1467-9280.01438>.
- Diener E, Wallbom M (1976) Effects of self-awareness on antinormative behavior. *Journal of Research in Personality* 10(1):107–111, URL [http://dx.doi.org/10.1016/0092-6566\(76\)90088-X](http://dx.doi.org/10.1016/0092-6566(76)90088-X).
- Dimant E, Shalvi S (2022) Meta-nudging honesty: Past, present, and future of the research frontier. *Current Opinion in Psychology* 47:Article101426, URL <http://dx.doi.org/10.1016/j.copsyc.2022.101426>.
- Dimant E, van Kleef GA, Shalvi S (2020) Requiem for a nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization* 172:247–266, URL <http://dx.doi.org/10.1016/j.jebo.2020.02.015>.
- Ditto PH, Lopez DF (1992) Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology* 63(4):568–584, URL <http://dx.doi.org/10.1037/0022-3514.63.4.568>.

- Ditto PH, Pizarro DA, Tannenbaum D (2009) Motivated moral reasoning. *Psychology of Learning and Motivation* 50:307–338, URL [http://dx.doi.org/10.1016/S0079-7421\(08\)00410-6](http://dx.doi.org/10.1016/S0079-7421(08)00410-6).
- Falk A, Becker A, Dohmen T, Huffman D, Sunde U (2023) The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science* 69(4):1935–1950, URL <http://dx.doi.org/10.1287/mnsc.2022.01178>.
- Fehrler S, Fischbacher U, Schneider MT (2020) Honesty and self-selection into cheap talk. *The Economic Journal* 130(632):2468–2496.
- Fischbacher U, Föllmi-Heusi F (2013) Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association* 11(3):525–547, URL <http://dx.doi.org/10.1111/jeea.12014>.
- Friesen L, Gangadharan L (2012) Individual level evidence of dishonesty and the gender effect. *Economics Letters* 117(3):624–626, URL <http://dx.doi.org/10.1016/j.econlet.2012.08.005>.
- Gangadharan L, Grossman PJ, Vecchi J (2020) Antisocial behavior in the workplace. Zimmermann KF, ed., *Handbook of Labor, Human Resources and Population Economics* (Springer, Cham), URL http://dx.doi.org/10.1007/978-3-319-57365-6_139-1.
- Gerlach P, Teodorescu K, Hertwig R (2019) The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin* 145(1):1–44, URL <http://dx.doi.org/10.1037/bul0000174>.
- Gibson R, Tanner C, Wagner AF (2013) Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review* 103(1):532–548, URL <http://dx.doi.org/10.1257/aer.103.1.532>.
- Gigerenzer G (2008) Moral intuition = fast and frugal heuristics? Sinnott-Armstrong W, ed., *Moral Psychology, Volume 2: The Cognitive Science of Morality*, 1–26 (MIT Press), ISBN 9780262195683.
- Gigerenzer G, Todd PM (1999) Fast and frugal heuristics: The adaptive toolbox. *Simple Heuristics That Make Us Smart*, 3–34 (Oxford University Press), ISBN 9780195143812.
- Gneezy U, Kajackaite A, Sobel J (2018) Lying aversion and the size of the lie. *American Economic Review* 108(2):419–453, URL <http://dx.doi.org/10.1257/aer.20161553>.
- Gneezy U, Rockenbach B, Serra-Garcia M (2013) Measuring lying aversion. *Journal of Economic Behavior & Organization* 93:293–300, URL <http://dx.doi.org/10.1016/j.jebo.2013.03.025>.

- Gneezy U, Saccardo S, Serra-Garcia M, van Veldhuizen R (2020) Bribing the self. *Games and Economic Behavior* 120:311–324, URL <http://dx.doi.org/10.1016/j.geb.2020.01.003>.
- Gneezy U, Saccardo S, van Veldhuizen R (2019) Bribery: Behavioral drivers of distorted decisions. *Journal of the European Economic Association* 17(3):917–946, URL <http://dx.doi.org/10.1093/jeea/jvy043>.
- Golman R, Hagmann D, Loewenstein G (2017) Information avoidance. *Journal of Economic Literature* 55(1):96–135, URL <http://dx.doi.org/10.1257/jel.20151245>.
- Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001) An fmri investigation of emotional engagement in moral judgment. *Science* 293(5537):2105–2108, URL <http://dx.doi.org/10.1126/science.1062872>.
- Gunia BC, Wang L, Huang L, Wang J, Murnighan JK (2012) Contemplation and conversation: Subtle influences on moral decision making. *Academy of Management Journal* 55(1):13–33, URL <http://dx.doi.org/10.5465/amj.2009.0873>.
- Haidt J (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108(4):814–834, URL <http://dx.doi.org/10.1037/0033-295X.108.4.814>.
- Hannah ST, Avolio BJ (2010) Moral potency: Building the capacity for character-based leadership. *Consulting Psychology Journal: Practice and Research* 62(4):291–310, URL <http://dx.doi.org/10.1037/a0022351>.
- Hanselmann M, Tanner C (2008) Taboos and conflicts in decision making: Sacred values, decision difficulty, and emotions. *Judgment and Decision Making* 3(1):51–63.
- Hardy SA, Francis SW, Zamboanga BL, Kim SY, Anderson SG, Forthun LF (2013) The roles of identity formation and moral identity in college student mental health, health-risk behaviors, and psychological well-being. *Journal of Clinical Psychology* 69(4):364–382, URL <http://dx.doi.org/10.1002/jclp.21913>.
- Hertz SG, Krettenauer T (2016) Does moral identity effectively predict moral behavior? a meta-analysis. *Review of General Psychology* 20(2):129–140, URL <http://dx.doi.org/10.1037/gpr0000062>.
- Hilbig BE, Zettler I, Heydasch T (2012) Personality, punishment and public goods: Strategic shifts towards cooperation as a matter of dispositional honesty–humility. *European Journal of Personality* 26(3):245–254, URL <http://dx.doi.org/10.1002/per.831>.

- Huynh LDT, Stratmann P, Rilke RM (2024) No influence of simple moral awareness cues on cheating behaviour in an online experiment. *Journal of Behavioral and Experimental Economics* 108:102126, ISSN 2214-8043, URL <http://dx.doi.org/10.1016/j.socec.2023.102126>.
- Irlenbusch B, Mussweiler T, Saxler DJ, Shalvi S, Weiss A (2020) Similarity increases collaborative cheating. *Journal of Economic Behavior & Organization* 178:148–173, URL <http://dx.doi.org/10.1016/j.jebo.2020.06.022>.
- Irlenbusch B, Villeval MC (2015) Behavioral ethics: How psychology influenced economics and how economics might inform psychology? *Current Opinion in Psychology* 6:87–92, URL <http://dx.doi.org/10.1016/j.copsyc.2015.04.004>.
- Janis I, Mann L (1977) *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment* (London: Cassel and Collier Macmillan), ISBN 9780029161601.
- Kish-Gephart JJ, Harrison DA, Treviño LK (2010) Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work. *Journal of Applied Psychology* 95(1):1–31, URL <http://dx.doi.org/10.1037/a0017103>.
- Köbis NC, Verschuere B, Bereby-Meyer Y, Rand D, Shalvi S (2019) Intuitive honesty versus dishonesty: Meta-analytic evidence. *Perspectives on Psychological Science* 14(5):778–796, URL <http://dx.doi.org/10.1177/1745691619851778>.
- Konow J (2000) Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review* 90(4):1072–1092, URL <http://dx.doi.org/10.1257/aer.90.4.1072>.
- Kouchaki M, Smith IH (2020) Building an ethical career. *Harvard Business Review* 98(1):135–139.
- KPMG (2024) KPMG Global Chief Ethics and Compliance Officer Survey. URL <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2024/01/stepping-up-to-a-new-level-of-compliance.pdf.coredownload.pdf>.
- Krawczyk M, Tyrowicz J, Kukla-Gryz A, Hardy W (2015) “piracy is not theft!” is it just students who think so? *Journal of Behavioral and Experimental Economics* 54:32–39, URL <http://dx.doi.org/10.1016/j.socec.2014.11.003>.
- Kristal AS, Whillans AV, Bazerman MH, Gino F, Shu LL, Mazar N, Ariely D (2020) Signing at the beginning versus at the end does not decrease dishonesty. *Proceedings of the National Academy of Sciences* 117(13):7103–7107, URL <http://dx.doi.org/10.1073/pnas.1911695117>.

- Krohne HW, Egloff B, Kohlmann CW, Tausch A (1996) Positive and negative affect schedule—german version. *Diagnostica* 42:139–156.
- Köbis N, Rahwan Z, Rilla R, Supriyatno BI, Bersch C, Ajaj T, Rahwan I (2025) Delegation to artificial intelligence can increase dishonest behaviour. *Nature* 646(8083):126–134.
- Leib M, Köbis N, Rilke RM, Hagens M, Irlenbusch B (2024) Corrupted by algorithms? how AI-generated and human-written advice shape (dis)honesty. *The Economic Journal* 134(658):766–784.
- Luo Y, Li A, Soman D, Zhao J (2023) A meta-analytic cognitive framework of nudge and sludge. *Royal Society Open Science* 10(11):230053, URL <http://dx.doi.org/10.1098/rsos.230053>.
- Martin SR, Kish-Gephart JJ, Detert JR (2014) Blind forces: Ethical infrastructures and moral disengagement in organizations. *Organizational Psychology Review* 4(4):295–325.
- Monin B, Pizarro DA, Beer JS (2007) Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology* 11(2):99–111, URL <http://dx.doi.org/10.1037/1089-2680.11.2.99>.
- Morvinski C, Saccardo S, Amir O (2023) Mis-nudging morality. *Management Science* 69(1):464–474, URL <http://dx.doi.org/10.1287/mnsc.2022.4344>.
- Mudrack PE, Mason ES (2013) Dilemmas, conspiracies, and sophie’s choice: Vignette themes and ethical judgments. *Journal of Business Ethics* 118(3):639–653, URL <http://dx.doi.org/10.1007/s10551-012-1611-0>.
- Murphy RO, Ackermann KA, Handgraaf MJJ (2011) Measuring social value orientation. *Judgment and Decision Making* 6(8):771–781.
- Nickerson RS (1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2(2):175–220, URL <http://dx.doi.org/10.1037/1089-2680.2.2.175>.
- Paine LS (1994) Managing for organizational integrity. *Harvard Business Review* 72(2):106–117.
- Pfeffer J, Sutton RI (2006) Evidence-based management. *Harvard Business Review* 84(1):62–74.
- Pierce L, Rees-Jones A, Blank C (2025) The negative consequences of loss-framed performance incentives. *American Economic Journal: Economic Policy* 17(1):506–539, ISSN 1945-7731, URL <http://dx.doi.org/10.1257/pol.20220512>.

- Pizarro DA, Laney C, Morris EK, Loftus EF (2006) Ripple effects in memory: Judgments of moral blame can distort memory for events. *Memory & Cognition* 34(3):550–555, URL <http://dx.doi.org/10.3758/BF03193582>.
- Pizarro DA, Uhlmann E, Salovey P (2003) Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science* 14(3):267–272, URL <http://dx.doi.org/10.1111/1467-9280.03433>.
- Provencher HL, Fincham FD (2000) Attributions of causality, responsibility and blame for positive and negative symptom behaviours in caregivers of persons with schizophrenia. *Psychological Medicine* 30(4):899–910, URL <http://dx.doi.org/10.1017/S0033291799002665>.
- Recalde MP, Riedl A, Vesterlund L (2018) Error-prone inference from response time: The case of intuitive generosity in public-good games. *Journal of Public Economics* 160:132–147, URL <http://dx.doi.org/10.1016/j.jpubeco.2018.02.010>.
- Reed II A, Aquino KF (2003) Moral identity and the expanding circle of moral regard toward out-groups. *Journal of Personality and Social Psychology* 84(6):1270–1286, URL <http://dx.doi.org/10.1037/0022-3514.84.6.1270>.
- Renerte B, Baader M, Tanner C, Wagner AF, Witt N (2025) Organizational ethics in action: The use of contemplation questions as decision aids in large companies. Technical Report 25-103, Swiss Finance Institute, URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5680982, working Paper.
- Rilke RM, Danilov A, Irlenbusch B, Weisel O, Shalvi S (2018) The honest leader effect – how hierarchies affect honesty in groups. *Academy of Management Proceedings* 2018(1):10365, URL <http://dx.doi.org/10.5465/AMBPP.2018.10365abstract>.
- Rosenbaum SM, Billinger S, Stieglitz N (2014) Let’s be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology* 45:181–196, URL <http://dx.doi.org/10.1016/j.joep.2014.10.004>.
- Saccardo S, Serra-Garcia M (2023) Enabling or limiting cognitive flexibility? evidence of demand for moral commitment. *American Economic Review* 113(2):396–429, URL <http://dx.doi.org/10.1257/aer.20201333>.
- Sachdeva S, Iliev R, Medin DL (2009) Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science* 20(4):523–528, ISSN 0956-7976, URL <http://dx.doi.org/10.1111/j.1467-9280.2009.02326.x>.

- Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V (2007) The constructive, destructive, and reconstructive power of social norms. *Psychological Science* 18(5):429–434, ISSN 0956-7976, URL <http://dx.doi.org/10.1111/j.1467-9280.2007.01917.x>.
- Shalvi S, Dana J, Handgraaf MJJ, De Dreu CKW (2011) Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes* 115(2):181–190, URL <http://dx.doi.org/10.1016/j.obhdp.2011.02.001>.
- Shalvi S, Gino F, Barkan R, Ayal S (2015) Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science* 24(2):125–130, URL <http://dx.doi.org/10.1177/0963721414553264>.
- Shalvi S, Levine EE, Thielmann I, Jayawickreme E, van Rooij B, Teodorescu K, Schurr A, Furr RM, Aglioti SM, Zettler I, Cohen TR, Pittarello A, Barkan R, Köbis N, Leib M, Mitkidis P, Schulz J, Dimant E, van Kleef GA, ŁŚcigała KA, Rilke RM, Ayal S, Beersma B, Plonsky O, Hilbig B, Weisel O, Butera F, Feldman Y, Verschuere B, Zanetti C, Hochman G, Kret ME, Peer E, Capraro V, Dorrough AR, Speer S, Ritov I (2025) The science of honesty: A review and research agenda. Gawronski B, ed., *Advances in Experimental Social Psychology*, volume 72, 241–327 (Academic Press / Elsevier), URL <http://dx.doi.org/10.1016/bs.aesp.2025.04.004>.
- Staw BM, Epstein LD (2000) What Bandwagons Bring: Effects of Popular Management Techniques on Corporate Performance, Reputation, and CEO Pay. *Administrative Science Quarterly* 45(3):523–556, URL <http://dx.doi.org/10.2307/2667108>.
- Sutter M (2009) Deception through telling the truth?! experimental evidence from individuals and teams. *The Economic Journal* 119(534):47–60.
- Tajfel H (1979) An integrative theory of intergroup conflict. Austin WG, Worchel S, eds., *The Social Psychology of Intergroup Relations*, 33–47 (Brooks/Cole), ISBN 9780818503217.
- Taylor A (2024) *Higher Ground: How Business Can Do the Right Thing in a Turbulent World* (Harvard Business Press), ISBN 9781647825745.
- Tergiman C, Villeval MC (2023) The way people lie in markets: Detectable vs. deniable lies. *Management Science* 69(6):3340–3357, URL <http://dx.doi.org/10.1287/mnsc.2022.4526>.
- Thomson Reuters Institute (2023) 2023 risk & compliance report. URL <https://www.thomsonreuters.com/en-us/posts/wp-content/uploads/sites/20/2023/10/2023-Risk-Compliance-Report.pdf>.

- Tian H, Wei X, Yu F, Xu L (2025) The preference for unsolicited advice: When we have to make a trade-off between self-interest and morality. *Journal of Business Ethics* 1–19, URL <http://dx.doi.org/10.1007/s10551-025-05977-z>.
- Uhlmann EL, Pizarro DA, Tannenbaum D, Ditto PH (2009) The motivated use of moral principles. *Judgment and Decision Making* 4(6):479–491.
- Vaidis DC, Slegers WW, Van Leeuwen F, DeMarree KG, Sætrevik B, Ross RM, Priolo D, et al. (2024) A multilab replication of the induced-compliance paradigm of cognitive dissonance. *Advances in Methods and Practices in Psychological Science* 7(1):25152459231213375, URL <http://dx.doi.org/10.1177/25152459231213375>.
- Verschuere B, Meijer EH, Jim A, Hoogesteyn K, Orthey R, McCarthy RJ, Skowronski JJ, Acar OA, Aczel B, Bakos BE (2018) Registered replication report on mazar, amir, and ariely (2008). *Advances in Methods and Practices in Psychological Science* 1(3):299–317, URL <http://dx.doi.org/10.1177/2515245918781032>.
- von Hippel W, Trivers R (2011) The evolution and psychology of self-deception. *Behavioral and Brain Sciences* 34(1):1–16; discussion 16–56, URL <http://dx.doi.org/10.1017/S0140525X10001354>.
- Weaver GR (2014) Encouraging ethics in organizations: A review of some key research findings. *American Criminal Law Review* 51:293–314.
- Weaver GR, Reynolds SJ, Brown ME (2014) Moral intuition: Connecting current knowledge to future organizational research and practice. *Journal of Management* 40(1):100–129, URL <http://dx.doi.org/10.1177/0149206313511272>.
- Weaver GR, Treviño LK (1999) Compliance and values oriented ethics programs: Influences on employees' attitudes and behavior. *Business Ethics Quarterly* 9(2):315–335, URL <http://dx.doi.org/10.2307/3857471>.
- Welsh DT, Baer MD, Sessions H, Garud N (2020) Motivated to disengage: The ethical consequences of goal commitment and moral disengagement in goal setting. *Journal of Organizational Behavior* 41(7):663–677, URL <http://dx.doi.org/10.1002/job.2454>.
- Winterich KP, Aquino K, Mittal V, Swartz R (2013) When moral identity symbolization motivates prosocial behavior: The role of recognition and moral identity internalization. *Journal of Applied Psychology* 98(5):759–770, URL <http://dx.doi.org/10.1037/a0033177>.

Witt N, Tanner C, Baader M, Egovor M, Renerte B, Wagner AF (2025) Ethical conflicts in the workplace: Insights from employees and experts, unpublished manuscript.

Zhang T, Fletcher PO, Gino F, Bazerman MH (2015) Reducing bounded ethicality: How to help individuals notice and avoid unethical behavior. *Organizational Dynamics* 44(4):310–317, URL <http://dx.doi.org/10.1016/j.orgdyn.2015.07.004>.

Zimmermann F (2020) The dynamics of motivated beliefs. *American Economic Review* 110(2):337–363, URL <http://dx.doi.org/10.1257/aer.20180080>.

Appendix for

The Unintended Effects of Ethical Decision Aids

in Organizations

Malte Baader

Norwegian School
of Economics, NO
malte.baader@nhh.no

Maxim Egorov

Munich Business School, DE
maxim.egorov@munich-business-school.de

Baiba Renerte

University of Zurich, CH
baiba.renerte@df.uzh.ch

Carmen Tanner

University of Zurich, CH
Zeppelin University, DE
carmen.tanner@df.uzh.ch

Alexander F. Wagner

University of Zurich, CH
CEPR, ECGI, and SFI
alexander.wagner@df.uzh.ch

Nicole Witt

Zeppelin University, DE
nicole.witt@zu.de

Contents

A Additional analyses	p. 2
A.1 Robustness regressions using pooled probit	p. 12
B Overview of Scales and Questionnaires	p. 21
C Screenshots	p. 30
D Examples of Contemplation Questions	p. 47
E Pre-registrations	p. 48
F The Role of Detectability	p. 61

A Additional analyses

Table A1: Study 1 - Overview of observations per vignette and situation

Situation	Vignette			
	Accessories	Work hours	Headlights	Env. Certificates
Self	476	510	465	535
Mistake	509	454	521	502
Loyalty	509	519	526	432
Mistake + Loyalty	492	503	474	517
Total	1,986	1,986	1,986	1,986

Note. The numbers in each cell represent the number of answers per vignette x situation combination.

Table A2: Study 1 - Overview of observations per treatment condition

Control	No CQ 500		
Self	Bad conscience 508	Mirror 480	No regret 496
Peers	Personal network 476	Family and friends 496	Close friends 520
Company Policy	Guidelines 496	Values 504	Culture 496
Law & Reputation	Legal 484	Reputation 504	Newspaper 472
Universal Principles	Others as self 512	Honesty 504	Accountable 496

Note. The numbers in each cell represent the number of answers per CQ treatment condition.

Table A3: Study 1 - Aggregate result of CQs

	(1)	(2)	(3)	(4)
<i>Treatment dummy</i>				
CQ present	0.030 (0.023)	0.031 (0.023)	0.030 (0.023)	0.033 (0.021)
<i>Situations</i>				
Hours worked		0.064*** (0.013)	0.067*** (0.014)	0.064*** (0.013)
Quality of headlights		0.059*** (0.013)	0.055*** (0.014)	0.060*** (0.013)
Environmental certificates		0.040** (0.013)	0.041** (0.014)	0.040** (0.013)
Mistake		-0.333*** (0.013)		-0.333*** (0.013)
Loyalty		-0.084*** (0.012)		-0.085*** (0.012)
Mistake + Loyalty		-0.169*** (0.013)		-0.170*** (0.013)
<i>Controls</i>				
Age				0.003*** (0.000)
Female				0.022 (0.011)
Prosociality				0.003 (0.004)
Moral identity				0.060*** (0.012)
Protected values (direct)				0.016 (0.012)
Protected values (indirect)				0.011 (0.006)
Self-efficacy				0.001 (0.008)
Need for cognition				0.017** (0.006)
Honesty humility				0.020* (0.010)
Emotionality				0.006 (0.009)
Extraversion				-0.007 (0.009)
Agreeableness				0.007 (0.010)
Conscientiousness				-0.002 (0.010)
Openness to experiences				-0.010 (0.008)
Risk				-0.001 (0.002)
Observations	8,832	8,832	8,832	8,832
Individuals	2,208	2,208	2,208	2,208
Vignette FE		X	X	X
Dilemma FE			X	X
Controls				X

Note. All models present marginal effects of probit regression models including random effects. The dependent variable is binary, capturing the ethical choice in each dilemma. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Random effects are included to account for multiple choices per participant. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table A4: Study 1 - Effect of CQs on stress, decision difficulty and decision avoidance

	Stress	Decision Difficulty	Decision Avoidance
CQ present	0.116 (0.103)	0.121 (0.102)	0.023 (0.106)
Hours worked	-0.121*** (0.036)	-0.210*** (0.039)	-0.226*** (0.039)
Quality of headlights	0.233*** (0.036)	0.108** (0.039)	0.358*** (0.039)
Environmental certificates	0.180*** (0.036)	0.103** (0.039)	0.363*** (0.039)
Age	-0.003 (0.002)	-0.002 (0.002)	-0.006* (0.002)
Female	0.239*** (0.055)	0.134* (0.055)	0.230*** (0.057)
Prosociality	0.039* (0.018)	0.034 (0.018)	0.047* (0.019)
Moral identity	-0.228*** (0.061)	-0.287*** (0.061)	-0.273*** (0.064)
Protected values (direct)	0.112 (0.059)	-0.009 (0.059)	0.069 (0.062)
Protected values (indirect)	-0.058* (0.028)	-0.063* (0.028)	-0.049 (0.029)
Self-efficacy	-0.245*** (0.038)	-0.200*** (0.038)	-0.252*** (0.040)
Need for cognition	-0.063* (0.030)	-0.082** (0.030)	-0.104*** (0.031)
Honesty humility	-0.163*** (0.048)	-0.156*** (0.047)	-0.106* (0.049)
Emotionality	0.264*** (0.043)	0.193*** (0.043)	0.222*** (0.045)
Extraversion	-0.183*** (0.045)	-0.095* (0.045)	-0.135** (0.047)
Agreeableness	0.090 (0.047)	0.100* (0.046)	0.098* (0.048)
Conscientiousness	-0.084 (0.047)	-0.071 (0.046)	-0.130** (0.049)
Openness to experiences	0.150*** (0.041)	0.098* (0.041)	0.120** (0.043)
Risk	0.021 (0.012)	0.021 (0.012)	-0.002 (0.012)
Constant	4.566*** (0.375)	5.345*** (0.372)	5.497*** (0.389)
Observations	8832	8832	8832
Clustered SEs	X	X	X
R ²			

Note. All models present marginal effects of linear regression models including random effects. The dependent variables are stress, decision difficulty and decision avoidance, each measured on a 7-point Likert scale. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Random effects are included to account for multiple choices per participant.
 * p < 0.05; ** p < 0.01; *** p < 0.001.

Table A5: Study 2 - Moral identity as a moderator

	(1)	(2)	(3)	(4)
	Moral Identity	Interaction	High Moral Identity	Low Moral Identity
CQ present		-1.423*** (0.176)	0.040*** (0.011)	-0.072*** (0.012)
Moral identity	0.065*** (0.016)	0.061 (0.061)		
Moral identity × CQ present		0.343*** (0.044)		
Mistake	-0.009 (0.014)	-0.034 (0.053)	-0.026 (0.018)	0.008 (0.021)
Loyalty	-0.043** (0.016)	-0.158** (0.059)	-0.062** (0.021)	-0.023 (0.024)
Mistake + Loyalty	-0.085*** (0.014)	-0.301*** (0.050)	-0.090*** (0.018)	-0.077*** (0.021)
Hours worked	0.045** (0.017)	0.153** (0.059)	0.075*** (0.023)	0.015 (0.026)
Quality of headlights	0.085*** (0.014)	0.298*** (0.050)	0.096*** (0.019)	0.074*** (0.021)
Environmental certificates	0.096*** (0.015)	0.339*** (0.052)	0.105*** (0.020)	0.089*** (0.022)
Observations	9,672	9,672	4,864	4,808
Vignette FEs	X	X	X	X
Dilemma FEs	X	X	X	X
Individual controls	X	X	X	X

Note. Models (1), (3) and (4) present marginal effects of probit regression models including random effects. Model (2) represents standard coefficients of a probit regression model including random effects due to the interaction term. The dependent variable in all models is binary, capturing the ethical choice in each dilemma. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Random effects are included to account for multiple choices per participant.

* p < 0.05; ** p < 0.01; *** p < 0.001.

Table A6: Study 2 - CQs are moderated across moral compass

	(1)	(2)	(3)
	Moral Identity	Honesty-Humility	Social Value Orientation
CQ present	-1.423*** (0.176)	-0.913*** (0.156)	0.165*** (0.041)
Moral identity	0.061 (0.061)		
Honesty-humility	0.082* (0.041)	0.039 (0.042)	0.152*** (0.037)
<i>Reference: SVO: Altruist</i>			
SVO: Prosocial	-0.071 (0.054)	-0.099 (0.054)	0.079 (0.063)
SVO: Individualist	-0.414*** (0.079)	-0.443*** (0.079)	-0.071 (0.095)
SVO: Competitive	-0.138 (0.135)	-0.217 (0.135)	0.091 (0.163)
Moral identity × CQ present	0.343*** (0.044)		
Honesty-humility × CQ present		0.225*** (0.040)	
SVO: Prosocial × CQ present			-0.357*** (0.065)
<i>Reference: SVO: Altruist</i>			
SVO: Individualist × CQ present			-0.740*** (0.100)
SVO: Competitive × CQ present			-0.605*** (0.173)
Mistake	-0.034 (0.053)	-0.034 (0.053)	-0.035 (0.053)
Loyalty	-0.158** (0.059)	-0.158** (0.059)	-0.158** (0.059)
Mistake + Loyalty	-0.301*** (0.050)	-0.300*** (0.050)	-0.301*** (0.050)
Hours worked	0.153** (0.059)	0.152** (0.059)	0.154** (0.059)
Quality of headlights	0.298*** (0.050)	0.297*** (0.050)	0.299*** (0.050)
Environmental certificates	0.339*** (0.052)	0.338*** (0.052)	0.343*** (0.052)
Observations	9672	9672	9672
Vignette FEs	X	X	X
Dilemma FEs	X	X	X
Individual controls	X	X	X

Note. All models present marginal effects of probit regression models including random effects. The dependent variable is binary, capturing the ethical choice in each dilemma. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Random effects are included to account for multiple choices per participant. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table A7: Study 2 – Effects of categories by high and low moral identity

	(1)	(2)
	High Moral Identity	Low Moral Identity
Self	0.082*** (0.022)	-0.128*** (0.027)
Peers	0.061* (0.025)	-0.058* (0.025)
Company policy	-0.010 (0.025)	-0.097*** (0.026)
Law & Reputation	0.041 (0.024)	-0.049 (0.026)
Universal Principles	0.007 (0.026)	-0.068** (0.026)
Multiple CQs	0.073** (0.023)	-0.013 (0.025)
Observations	9,672	9,672
Controls	X	X

Note. Both columns present marginal effects of a single random-effects probit regression. The dependent variable is binary, capturing the ethical choice in each dilemma. Participants are classified as high or low moral identity based on a *median split* of the moral identity scale; we interact a high-moral-identity dummy with each CQ category and report marginal effects separately for the two groups. Each row corresponds to a CQ category, presented in the same order as in the Study 1 tables, and “Multiple CQs” denotes the treatment in which all five CQs are shown simultaneously. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Random effects account for multiple choices per participant. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table A8: Study 2 – Multiple-hypothesis-testing correction for the moral-compass heterogeneity

	Low			High		
	ME	<i>p</i>	<i>q</i>	ME	<i>p</i>	<i>q</i>
<i>Panel A: Moral identity</i>						
Self	−0.128	< 0.001	< 0.001	0.082	< 0.001	0.001
Peers	−0.058	0.019	0.033	0.062	0.013	0.027
Company Policy	−0.097	< 0.001	0.001	−0.010	0.684	0.746
Law & Reputation	−0.049	0.060	0.089	0.041	0.095	0.127
Universal Principles	−0.068	0.009	0.022	0.007	0.793	0.793
Multiple CQs	−0.013	0.604	0.724	0.073	0.001	0.003
<i>Panel B: Honesty-humility</i>						
Self	−0.061	0.008	0.030	0.045	0.106	0.141
Peers	−0.041	0.079	0.119	0.052	0.055	0.109
Company Policy	−0.088	< 0.001	0.002	−0.001	0.983	0.983
Law & Reputation	−0.041	0.070	0.119	0.059	0.034	0.082
Universal Principles	−0.069	0.005	0.027	0.022	0.415	0.498
Multiple CQs	0.008	0.724	0.790	0.061	0.018	0.054
<i>Panel C: Social value orientation</i>						
Self	−0.094	< 0.001	0.001	0.051	0.036	0.062
Peers	−0.042	0.077	0.102	0.040	0.125	0.151
Company Policy	−0.090	< 0.001	0.001	−0.015	0.557	0.557
Law & Reputation	−0.090	< 0.001	0.001	0.086	< 0.001	0.001
Universal Principles	−0.129	< 0.001	< 0.001	0.043	0.065	0.097
Multiple CQs	−0.028	0.265	0.289	0.084	< 0.001	0.001

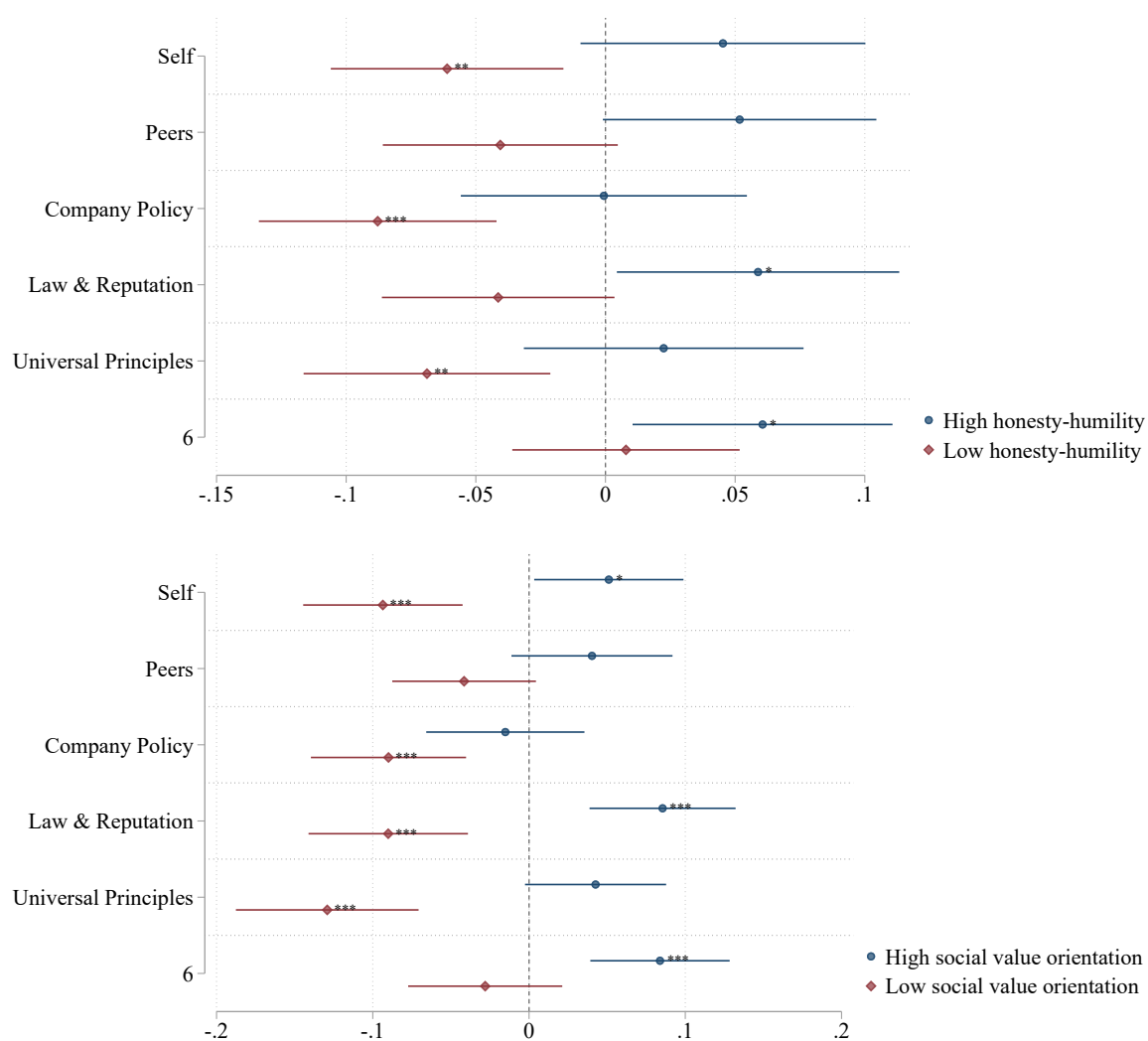
Note. Each cell reports the marginal effect (ME) of a CQ category on the probability of an ethical decision, estimated from the random-effects probit models underlying Table A7 and Appendix Figure A1, separately for participants below (*Low*) and above (*High*) the median of each moral-compass proxy (moral identity, honesty-humility, and social value orientation). *p* is the unadjusted p-value and *q* is the Benjamini-Hochberg false-discovery-rate-adjusted p-value computed within each proxy family (the 12 tests per panel). The moral-identity effects (Panel A), including the boomerang for low-moral-identity participants, remain significant after correction; a small number of secondary effects significant only at the 0.05 level do not. Individual controls and the estimation sample are as in Table A7.

Table A9: Study 2 – Robustness to the display order of the five simultaneous CQs

	(1) RE probit	(2) Pooled probit
<i>Display position of each CQ (0 = shown first)</i>		
Self	−0.043 (0.095)	−0.033 (0.065)
Peers	−0.088 (0.087)	−0.065 (0.060)
Company Policy	−0.048 (0.090)	−0.042 (0.065)
Law & Reputation	−0.152 (0.092)	−0.106 (0.063)
Negative framing	−0.307 (0.198)	−0.253 (0.134)
Joint test of positions (χ^2_4)	3.15	3.32
<i>p</i> -value	0.533	0.506
Observations	808	808
Participants	202	202
Individual controls	X	X
Situation & vignette FEs	X	X

Note. The sample is restricted to the Wave-2 decisions made under the multiple-CQ treatment, in which all five CQs are presented simultaneously in a randomised order ($N = 808$ decisions by 202 participants). The dependent variable is whether the participant made an ethical choice. Each regressor is the display position of the corresponding CQ (coded 0 for shown first through 4 for shown last); Universal Principles is the omitted reference, since the five positions are a permutation and sum to a constant. Column (1) is a random-effects probit with a participant-level random effect; column (2) is a pooled probit with standard errors clustered by participant. Both include the same individual controls and situation and vignette fixed effects as the main Study 2 specification (Table A7). Neither any individual position term nor a joint Wald test of all four positions is statistically significant, indicating that the effect of presenting five CQs is not driven by the order in which they appear. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Figure A1: Study 2 – Effects of CQs by honesty-humility and social value orientation



Note. In each panel above, we present marginal effects of individual CQs from one random-effect probit regression. The dependent binary variable is whether participants made an ethical choice. In the top panel, we present interactions with a high honesty-humility dummy in blue and with a low honesty-humility dummy in red. In the bottom panel, we present the same for social value orientation. Coefficients capture the effect relative to the control condition (Wave 1) in the absence of CQs. We include situation and vignette fixed effects, as well as age, gender, social value orientation, HEXACO as additional individual controls. We also control whether an individual has a manager, is in a managerial role and how long they have been employed. Regression results are presented in Appendix A7. * p < 0.05; ** p < 0.01; *** p < 0.001. Bars indicate 95% confidence intervals.

Table A10: Study 2 - Effect of CQs by framing

	Positive frame	Negative frame
Low moral identity	-0.049** (0.017)	-0.088*** (0.016)
High moral identity	0.048** (0.016)	0.037* (0.017)
Observations	4,832	4,840

Note. Each cell reports the marginal effect of CQs (Wave 2 relative to the Wave 1 control) on the probability of an ethical decision, estimated from random-effects probit regressions run separately for positively and negatively framed CQs and split by a median split on moral identity. The frame does not significantly moderate the effect of CQs (frame $\times$ CQ interaction $z = 1.64$, $p = 0.101$). Individual controls include situation and vignette fixed effects, age, gender, social value orientation, HEXACO traits, regulatory focus, moral motivation, and managerial role and tenure. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

A.1 Robustness regressions using pre-registered pooled probit and Tobit specifications

Table A11: Study 1 - Aggregate result of CQs (pooled probit)

	(1)	(2)	(3)	(4)
<i>Treatment dummy</i>				
CQ present	0.026 (0.024)	0.026 (0.024)	0.027 (0.024)	0.028 (0.024)
<i>Situations</i>				
Hours worked		0.075*** (0.015)	0.069*** (0.014)	0.069*** (0.014)
Quality of headlights		0.063*** (0.015)	0.066*** (0.014)	0.066*** (0.014)
Environmental certificates		0.034* (0.015)	0.033* (0.014)	0.033* (0.014)
Mistake			-0.345*** (0.014)	-0.345*** (0.014)
Loyalty			-0.093*** (0.012)	-0.093*** (0.012)
Mistake + Loyalty			-0.174*** (0.013)	-0.174*** (0.013)
<i>Controls</i>				
Age				0.003*** (0.001)
Female				0.026* (0.012)
Prosociality				0.003 (0.004)
Moral identity				0.058*** (0.013)
Protected values (direct)				0.011 (0.013)
Protected values (indirect)				0.010 (0.006)
Self-efficacy				0.002 (0.008)
Need for cognition				0.016* (0.006)
Honesty humility				0.019 (0.010)
Emotionality				0.007 (0.009)
Extraversion				-0.007 (0.009)
Agreeableness				0.010 (0.010)
Conscientiousness				-0.004 (0.010)
Openness to experiences				-0.013 (0.009)
Risk				0.001 (0.003)
Observations	7,944	7,944	7,944	7,944
Clustered SEs	X	X	X	X
Pseudo R ²	0.00	0.06	0.00	0.08

Note. All models present marginal effects of pooled probit regression models. The dependent variable is binary, capturing the ethical choice in each dilemma. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Standard errors are clustered at the individual level. * p < 0.05; ** p < 0.01; *** p < 0.001.

Table A12: Study 1 - Effect of CQs on stress, decision difficulty and decision avoidance (pooled probit)

	Stress	Decision Difficulty	Decision Avoidance
CQ present	0.113 (0.099)	0.137 (0.103)	-0.002 (0.109)
Hours worked	-0.123** (0.039)	-0.204*** (0.043)	-0.234*** (0.041)
Quality of headlights	0.247*** (0.038)	0.116** (0.041)	0.374*** (0.041)
Environmental certificates	0.195*** (0.037)	0.121** (0.040)	0.392*** (0.039)
Age	-0.004 (0.002)	-0.003 (0.002)	-0.007** (0.003)
Female	0.203*** (0.058)	0.093 (0.057)	0.195** (0.060)
Prosociality	0.031 (0.019)	0.025 (0.019)	0.033 (0.019)
Moral identity	-0.132 (0.072)	-0.215** (0.073)	-0.188* (0.075)
Protected values (direct)	0.100 (0.065)	-0.033 (0.065)	0.043 (0.068)
Protected values (indirect)	-0.063* (0.031)	-0.065* (0.031)	-0.052 (0.032)
Self-efficacy	-0.256*** (0.044)	-0.207*** (0.044)	-0.261*** (0.045)
Need for cognition	-0.064* (0.031)	-0.080** (0.030)	-0.099** (0.032)
Honesty humility	-0.137** (0.052)	-0.128* (0.053)	-0.074 (0.055)
Emotionality	0.267*** (0.044)	0.196*** (0.043)	0.220*** (0.046)
Extraversion	-0.152** (0.049)	-0.073 (0.048)	-0.097 (0.050)
Agreeableness	0.074 (0.049)	0.085 (0.050)	0.090 (0.052)
Conscientiousness	-0.096* (0.049)	-0.081 (0.049)	-0.160** (0.050)
Openness to experiences	0.167*** (0.044)	0.114** (0.044)	0.125** (0.046)
Risk	0.011 (0.012)	0.011 (0.013)	-0.011 (0.013)
Constant	4.270*** (0.390)	5.135*** (0.402)	5.378*** (0.411)
Observations	7,944	7,944	7,944
Clustered SEs	X	X	X
R ²	0.10	0.08	0.11

Note. All models present marginal effects of pooled probit regression models. The dependent variables are stress, decision difficulty and decision avoidance, each measured on a 7-point Likert scale. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Standard errors are clustered at the individual level. * p < 0.05; ** p < 0.01; *** p < 0.001.

Table A13: Study 1 - Effect of CQs on stress, decision difficulty and decision avoidance (Tobit)

	Stress	Decision Difficulty	Decision Avoidance
CQ present	0.204 (0.172)	0.177 (0.167)	0.073 (0.173)
Hours worked	-0.279*** (0.057)	-0.413*** (0.061)	-0.418*** (0.059)
Quality of headlights	0.332*** (0.056)	0.114 (0.060)	0.503*** (0.058)
Environmental certificates	0.261*** (0.056)	0.136* (0.060)	0.520*** (0.058)
Age	-0.002 (0.004)	-0.003 (0.004)	-0.007 (0.004)
Female	0.341*** (0.092)	0.198* (0.089)	0.332*** (0.093)
Prosociality	0.074* (0.030)	0.055 (0.030)	0.077* (0.031)
Moral identity	-0.386*** (0.102)	-0.475*** (0.100)	-0.454*** (0.104)
Protected values (direct)	0.125 (0.099)	-0.048 (0.096)	0.062 (0.100)
Protected values (indirect)	-0.089 (0.048)	-0.100* (0.046)	-0.074 (0.048)
Self-efficacy	-0.493*** (0.064)	-0.400*** (0.062)	-0.456*** (0.064)
Need for cognition	-0.072 (0.050)	-0.120* (0.049)	-0.159** (0.051)
Honesty humility	-0.220** (0.080)	-0.202** (0.078)	-0.143 (0.081)
Emotionality	0.440*** (0.073)	0.308*** (0.071)	0.358*** (0.074)
Extraversion	-0.279*** (0.075)	-0.156* (0.073)	-0.191* (0.076)
Agreeableness	0.013 (0.078)	0.077 (0.076)	0.055 (0.079)
Conscientiousness	-0.120 (0.078)	-0.093 (0.076)	-0.189* (0.079)
Openness to experiences	0.246*** (0.069)	0.184** (0.067)	0.187** (0.070)
Risk	0.047* (0.020)	0.046* (0.019)	0.007 (0.020)
Constant	5.756*** (0.629)	6.884*** (0.613)	6.979*** (0.636)
Observations	8832	8832	8832

Note. All models are random-effects Tobit regressions—the specification pre-registered for the secondary outcomes—with the dependent variable censored at the lower (1) and upper (7) bounds of the 7-point Likert scale. The dependent variables are stress, decision difficulty and decision avoidance. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits and risk preferences. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table A14: Study 2 - Moral identity as a moderator (pooled probit)

	(1)	(2)	(3)	(4)
	Moral Identity	Interaction	High Moral Identity	Low Moral Identity
CQ present		-1.121*** (0.225)	0.037* (0.015)	-0.067*** (0.016)
Moral identity	0.073*** (0.016)	0.097 (0.058)		
Moral identity × CQ present		0.267*** (0.055)		
Mistake	-0.017 (0.013)	-0.058 (0.043)	-0.035* (0.017)	-0.000 (0.019)
Loyalty	-0.050*** (0.015)	-0.160*** (0.048)	-0.064*** (0.019)	-0.035 (0.022)
Mistake + Loyalty	-0.092*** (0.013)	-0.289*** (0.042)	-0.093*** (0.017)	-0.089*** (0.020)
Hours worked	0.046** (0.018)	0.138* (0.053)	0.085*** (0.024)	0.010 (0.026)
Quality of headlights	0.085*** (0.015)	0.262*** (0.045)	0.090*** (0.020)	0.081*** (0.021)
Environmental certificates	0.095*** (0.015)	0.295*** (0.047)	0.109*** (0.021)	0.084*** (0.022)
Observations	8,880	8,880	4,264	4,616
Clustered individual SEs	X	X	X	X
Vignette FEs	X	X	X	X
Dilemma FEs	X	X	X	X
Individual controls	X	X	X	X
Pseudo R ²	0.05	0.06	0.05	0.04

Note. Models (1), (3) and (4) present marginal effects of pooled probit regression models. Model (2) represents standard coefficients of a pooled probit regression model due to the interaction term. The dependent variable in all models is binary, capturing the ethical choice in each dilemma. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Standard errors are clustered at the individual level. * p < 0.05; ** p < 0.01; *** p < 0.001.

Table A15: Study 2 - Moral identity as a moderator across moral compass (pooled probit)

	(1)	(2)	(3)
	Moral Identity	Honesty-Humility	Social Value Orientation
CQ present	-1.222*** (0.206)	-0.782*** (0.176)	0.141** (0.047)
Moral identity	0.053 (0.056)		
Honesty-humility	0.068 (0.037)	0.032 (0.040)	0.129*** (0.034)
<i>Reference: SVO: Altruist</i>			
SVO: Prosocial	-0.061 (0.047)	-0.085 (0.047)	0.071 (0.057)
SVO: Individualist	-0.365*** (0.070)	-0.388*** (0.069)	-0.060 (0.083)
SVO: Competitive	-0.130 (0.133)	-0.197 (0.128)	0.069 (0.159)
Moral identity × CQ present	0.294*** (0.051)		
Honesty-humility × CQ present		0.192*** (0.045)	
SVO: Prosocial × CQ present			-0.313*** (0.073)
<i>Reference: SVO: Altruist</i>			
SVO: Individualist × CQ present			-0.646*** (0.112)
SVO: Competitive × CQ present			-0.519* (0.209)
Mistake	-0.037 (0.041)	-0.036 (0.041)	-0.037 (0.041)
Loyalty	-0.140** (0.046)	-0.139** (0.045)	-0.140** (0.045)
Mistake + Loyalty	-0.265*** (0.040)	-0.263*** (0.040)	-0.264*** (0.040)
Hours worked	0.134** (0.050)	0.132** (0.050)	0.133** (0.050)
Quality of headlights	0.257*** (0.042)	0.256*** (0.042)	0.257*** (0.042)
Environmental certificates	0.294*** (0.044)	0.292*** (0.043)	0.295*** (0.044)
Observations	9672	9672	9672
Vignette FEs	X	X	X
Dilemma FEs	X	X	X
Individual controls	X	X	X

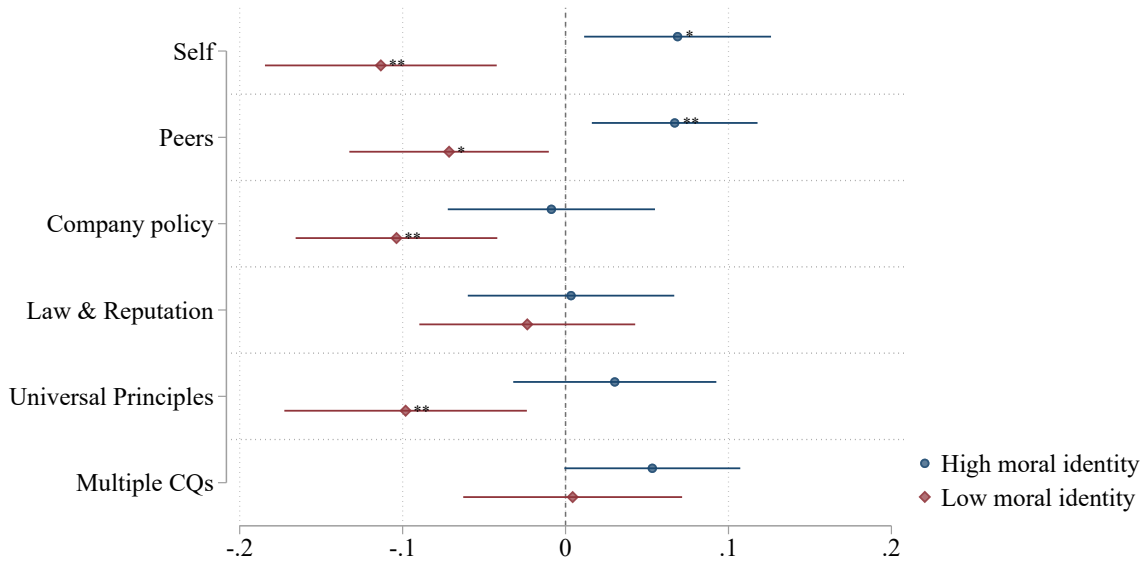
Note. All models present marginal effects of pooled probit regression models. The dependent variable is binary, capturing the ethical choice in each dilemma. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Standard errors are clustered at the individual level. * p < 0.05; ** p < 0.01; *** p < 0.001.

Table A16: Study 2 – Effects of categories by high and low moral identity (pooled probit)

	(1)	(2)
	High Moral Identity	Low Moral Identity
Self	0.073* (0.032)	-0.113** (0.035)
Peers	0.079** (0.028)	-0.068* (0.030)
Company policy	-0.009 (0.035)	-0.097** (0.031)
Law & Reputation	0.005 (0.035)	-0.019 (0.032)
Universal Principles	0.029 (0.035)	-0.093* (0.037)
Multiple CQs	0.058 (0.030)	0.000 (0.032)
Observations	8,880	8,880
Controls	X	X

Note. Both columns present marginal effects of a single pooled probit regression model. The dependent variable is binary, capturing the ethical choice in each dilemma. We interact a dummy for high moral identity with each CQ category and generate marginal effects for low and high moral identity. Individual controls include age, gender, social preferences, moral identity, protected values, self-efficacy, need for cognition, HEXACO traits, and risk preferences. Standard errors are clustered at the individual level. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Figure A2: Study 2 – Effects of CQs by moral identity (pooled probit)



Note. We present marginal effects of individual CQs from one pooled probit regression. The dependent binary variable is whether participants made an ethical choice. We present interactions with a high moral identity dummy in blue and with a low moral identity dummy in red. Coefficients capture the effect relative to the control condition (Wave 1) in the absence of CQs. We include situation and vignette fixed effects, as well as age, gender, social value orientation, HEXACO as additional individual controls. We also control whether an individual has a manager, is in a managerial role and how long they have been employed. Regression results in tabular form are presented in Appendix A7. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Bars indicate 95% confidence intervals.

A.2 Open-text justifications

As additional, suggestive evidence on the mechanism underlying our results, we draw on a separate pilot study for Study 2 ($N = 562$) in which we collected open-text justifications for participants’ choices. Participants faced and answered a subset of CQs before making binary decisions in three dilemma vignettes, and after each decision an open-text question asked for the reasons behind it. These justifications speak to the contrast between our two mechanisms. Under moral agency, the reference standard $\Delta r_i(a)$ is fixed, so justifications should be similar regardless of the action chosen. Under motivated moral reasoning, $\Delta r_i(a)$ is endogenous to $\Delta v_i(a)$, so those favouring the unethical action adjust their perceived approval to fit it. The latter implies an asymmetry: justifications for the unethical action should draw on whichever rationale the vignette makes available, whereas justifications for the ethical action rest on more stable standards.

Using keyword dictionaries, we code each open-text response for two types of rationale, each as a binary indicator equal to one if any corresponding expression appears. The first *dissolves* the moral cost of the action (denying harm, citing a superior’s instruction, or invoking the firm’s survival); the second *highlights* it (explicit reference to the deception or harm, or adoption of the customer’s perspective). Restricting attention to participants who received contemplation questions ($N = 412$; 437 unethical versus 670 ethical decisions), the justifications given for the unethical action differ markedly from those given for the ethical action (Table A17): those choosing the unethical action refer significantly more often to a cost-dissolving rationale, and significantly less often to naming the harm or taking the customer’s perspective.

Table A17: Study 2 pilot — Justification content by chosen action

	Unethical action	Ethical action	z
Cost-dissolving rationale	52%	17%	12.3***
Names the harm	8%	24%	−6.6***
Customer’s perspective	23%	39%	−5.5***

Note. Share of open-text justifications coded as containing each rationale among pilot participants who received contemplation questions ($N = 412$), shown separately for participants who chose the unethical versus the ethical action. z -statistics are from two-sample tests of proportions. *** $p < 0.001$.

We stress that this evidence is only suggestive. The pilot does not vary CQs within-subject, so we cannot observe whether an individual’s justifications shift when contemplation questions are introduced, and the observed asymmetry does not differ between the CQ and control conditions, so we cannot attribute it to contemplation questions specifically. The pattern is nonetheless consistent with motivated moral reasoning, under which individual valuations

$v_i(a)$ are reinforced through interaction with CQs.

B Overview of Scales and Questionnaires

All scales below are translated from German

B.1 Study 1

Integrated within Main Study

Decision Difficulty (Hanselmann & Tanner, 2008)

Please now think of the decision you made at the end of the situation. To what extent do the following characteristics apply from your point of view? (7-point Likert scale, 1 – Does not apply at all; 7 – Applies very much)

1. I found it difficult to make the decision.
2. I was very uncertain about this decision.
3. It took me a long time to make a decision.

Decision Stress (Hanselmann & Tanner, 2008)

Please now think of the decision you made at the end of the situation. To what extent do the following characteristics apply from your point of view? (7-point Likert scale, 1 – Does not apply at all; 7 – Applies very much)

1. This decision stressed me.
2. I found this decision burdensome.
3. I felt overwhelmed by this decision.

Decision Avoidance (based on Janis & Mann, 1977)

Please now think of the decision you made at the end of the situation. To what extent do the following characteristics apply from your point of view? (7-point Likert scale, 1 – Does not apply at all; 7 – Applies very much)

1. I would have preferred to avoid the decision.
2. I would have preferred to leave the decision to others.
3. I would have preferred to postpone the decision.

Separate Questionnaire to Elicit Personality Traits

Moral Identity (Black & Reynolds, 2016)

Please indicate for all the statements listed to what extent you agree with each one. (5-point Likert scale, 1 - Do not agree at all; 5 - Strongly agree)

1. I try to be honest in most things I do.
2. Not hurting other people is one of the rules I live by.
3. It is important to me to treat other people fairly.
4. I want other people to know that they can rely on me.
5. I always act in a way that benefits other people the most and harms them the least.
6. If an action would harm another person, I try to avoid it, even if no one would ever find out.
7. One of the most important things in life is to do what you know is right.
8. Once I have decided what is right, I make sure I do it.
9. As long as I decide to do something that helps me, it doesn't matter much if other people are harmed.
10. It is okay to do something you know is wrong if the reward is great.

Falk Survey Questions (Falk et al., 2023)

Risk

General Assessment: How do you assess yourself personally? Are you generally a risk-taking person or do you try to avoid risks? (Likert scale 0 to 10: 0 = not at all willing to take risks, 10 = very willing to take risks)

Risk (Gneezy & Potters, 1997)

Imagine you receive 80 Euros from us. You can invest part of this money in a lottery. In the lottery you have a 50% chance of winning.

How does the payout work?

- You roll a die. If you roll a four or higher, you win. The amount you invested in the lottery will be multiplied by 2.5 and paid out to you. If you roll a three or lower, you lose your investment. You may keep any amount you did not invest in any case.

What amount would you like to invest?

Table B18: Possible Payouts

Your Investment	Roll: 1, 2, 3	Roll: 4, 5, 6
0 Euro	80	80
10 Euro	70	95
20 Euro	60	110
30 Euro	50	125
40 Euro	40	140
50 Euro	30	155
60 Euro	20	170
70 Euro	10	185
80 Euro	0	200

Altruism

How do you assess your willingness to share with others without expecting anything in return when it comes to charitable causes? (Likert scale 0 to 10: 0 = not at all willing, 10 = very willing)

Trust

How do you assess yourself personally? "If I am not otherwise convinced, I assume that people have the best intentions." (Likert scale 0 to 10: 0 = does not describe me at all, 10 = describes me very well)

Positive Reciprocity

How do you assess yourself personally? "If someone does me a favor, I am willing to return it." (Likert scale 0 to 10: 0 = not at all willing, 10 = very willing)

Self-Efficacy (Hannah & Avolio, 2010)

I am confident that I can...

1. ...admit a mistake.
2. ...resist others who demand something immoral from me.
3. ...not let myself be led astray from the right path.
4. ...handle difficult situations competently.

Protected Values (Gibson et al., 2013)

Evaluation of Managerial Decisions

Due to their performance-based compensation, managers have the incentive to alter information in public reports. What do you think of managers altering company information in reports to their advantage? Please select the response that applies to you. (7-point Likert scale, 1 – not immoral at all; 7 – very immoral)

1. Immoral
2. Praiseworthy
3. Shameful
4. Outrageous
5. Acceptable

Value of Honesty in this Situation

Due to their performance-based compensation, managers have the incentive to alter information in public reports. Some see this as a violation of the principle of honesty, others as the protection of personal interests. What do you think about the value of honesty in such a situation?

Regarding the value "Honesty"... (7-point Likert scale, 1 – do not agree at all; 7 – strongly agree)

1. ... it is something that should not be sacrificed, no matter how great the (financial or other) benefit.
2. ... it is something that cannot be expressed in monetary terms.
3. ... I think it is right to make cost-benefit considerations.
4. ... I can be flexible if the situation requires it.
5. ... it is about things or values that are inviolable.

HEXACO (de Vries, 2013)

Please indicate to what extent you agree with the following statements. (5-point Likert scale, 1 = do not agree at all; 5 = strongly agree)

1. I can look at a painting for a long time.
2. I make sure that things are always in the right place.
3. I remain unfriendly toward someone who was mean to me.
4. Nobody wants to talk to me.

5. I am afraid of feeling pain.
6. I find it difficult to lie.
7. I find science boring.
8. I procrastinate on complicated tasks as long as possible.
9. I often criticize.
10. I easily make contact with strangers.
11. I worry less than others.
12. I would like to know how I can make a lot of money dishonestly.
13. I have a lot of imagination.
14. I work very precisely.
15. I quickly agree with others.
16. I like to talk to others.
17. I can cope well on my own to overcome difficulties.
18. I want to be famous.
19. I like people with strange ideas.
20. I often do things without really thinking.
21. Even when treated badly, I remain calm.
22. I am rarely lively.
23. I cry at sad or romantic movies.
24. I feel entitled to special treatment.

Need for Cognition (Beißert et al., 2014)

The following statements may apply to you to a greater or lesser extent. Please indicate for each statement the extent to which it generally applies to you personally. (7-point Likert scale, 1 – does not apply at all; 7 – fully applies)

1. It is enough for me to simply know the answer without understanding the reasons for the answer to a problem.
2. I like it when my life is full of tricky tasks I have to solve.
3. I would prefer more complicated problems to simple ones.
4. I primarily think because I have to.

B.2 Study 2

Integrated within Main Study

Decision Confidence (Combination of 3 Scales in Study 1)

Please now think of the decision you made at the end of the situation. To what extent do the following characteristics apply from your point of view? (7-point Likert scale, 1 – Does not apply at all; 7 – Applies very much)

1. I am confident that I made the right decision.
2. I feel up to such situations.
3. I would have preferred to leave the decision to others.
4. This decision feels right.
5. I found this decision burdensome.

Emotions (Krohne et al., 1996 - PANAS)

Please now think of the decision you made at the end of the last situation. How did you feel afterwards?

1. Guilty
2. Nervous
3. Ashamed
4. Distressed
5. Confused
6. Proud
7. Strong
8. Determined
9. Excited
10. Joyful

Moral Disengagement (based on Barsky, 2011)

Please now think of the decision you made at the end of the last situation. To what extent do the following statements apply to you?

1. If it helps employees do their job, it is okay to deceive others.
2. It is okay to stretch the truth to protect the company.
3. If an employee has to stretch the truth to get their job done, they cannot be blamed for lying.

4. Employees cannot be held responsible for misconduct if they feel pressured by others to do it.
5. Employees cannot be blamed for exaggerating the truth if all other employees are doing it.

Separate Questionnaire to Elicit Personality Traits

Regulatory Focus (Crowe & Higgins, 1997)

In this series of questions, you will be asked about certain events in your life. (5-point Likert scale, 1 – never or rarely; 5 – very often)

1. Compared to most people, are you usually not able to achieve in life what you want?
2. As a child, did you ever “cross boundaries” by doing things your parents did not tolerate?
3. How often have you been motivated by achieving goals to work even harder?
4. Did you often annoy your parents during your childhood?
5. How often did you follow your parents’ rules and regulations?
6. As a child, did you ever behave in a way that your parents found reprehensible?
7. Do you often succeed at various things you try?
8. I have sometimes gotten into trouble because I was not careful enough.
9. When I want to achieve goals that are important to me, my performance is often not as good as I would like.
10. I have the impression that I have made progress regarding my personal success in life.
11. I have very few hobbies or interests that I can get excited about or that motivate me to strive for them.

Moral Identity (Black & Reynolds, 2016)

As in Study 1

Falk Risk (Falk et al., 2023)

As in Study 1

Social Value Orientation (Murphy et al., 2011)

In this task, imagine that you are randomly paired with another person whom we will refer to as the other person. This other person is someone you do not know and remains anonymous

to both sides. All of your decisions are also anonymous. You will make a series of hypothetical decisions about the allocation of Euros between yourself and this other person.

Please indicate for each of the following questions the allocation you most prefer. There are no right or wrong answers, this is only about personal preferences.

Table B19: Money distributions between you and the other person

You receive	Other receives
85	85
85	76
85	68
85	59
85	50
85	41
85	33
85	24
85	15

HEXACO (de Vries, 2013)

As in Study 1

Moral Motivation (Krettenauer & Casey, 2015)

Please indicate for each reason in the three questions below how important you consider it to be for yourself: (5-point Likert scale, 1 = not important at all; 5 = very important)

DoGood

How important is each of the following reasons why you might do something good for someone?

1. Because I want others to be nice to me.
2. Because I want other people to think I am a nice person.
3. Because I think it is good to do nice things for others.
4. Because I am a nice person.

Steal

How important is each of the following reasons why you would not take something that does not belong to you?

1. Because if I get caught, I could get into trouble.
2. Because other people would be disappointed in me.
3. Because I consider stealing to be unjust.
4. Because I do not see myself as a person who is unjust.

Truth

How important is each of the following reasons why you would decide to tell the truth if you had the opportunity to lie?

1. Because I do not want to get into trouble if I lie.
2. Because I do not want people to think of me as a liar.
3. Because I think it is right to tell the truth.
4. Because I see myself as an honest person.

C Screenshots

This appendix contains screenshots documenting the study's implementation. Each screenshot is annotated with a number in the top right corner indicating its sequence of implementation. The numbering restarts at the beginning of each page.

Table C20 outlines the order which participants followed. The CQ intervention screens showcase the differences in interacting with CQs between Study 1 and Study 2.

Table C20: Order of tasks completed by participants

Order	Phase	Content
1	Welcome	Welcome screen and informed consent
2	Instructions	Task instructions and selection of profile photo
3	Vignette decision	Repeated 4× (randomly drawn Vignette × Scenario combination). Each vignette presents: • Vignette × Situation content • CQ intervention screen (absent in control condition) • Decision screen • Decision certainty rating (after each vignette)
4	Post-survey	Post-intervention survey

Figure C3: Start screens



Universität Zürich

zweischen
Wirtschaft Kultur Politik



zeppelin universität

zweischen
Wirtschaft Kultur Politik

1

Welcome!

We are very pleased that you would like to take part in the 2nd part of this 3-part scientific study conducted by the University of Zurich and Zeppelin University Friedrichshafen.

Background of the study:
In the work context, employees are confronted with various situations. We are interested in how you perceive such situations and how you would act in them. Through your participation, you make an important contribution to scientific research!

The survey takes approximately 15 minutes.

Please read the information on the next page carefully, and then decide whether you would like to participate in this study.

Thank you very much for your interest and your support!

Prof. Dr. Carmen Tanner and Prof. Dr. Alexander F. Wagner

University of Zurich
Department of Finance

Plattenstrasse 32
CH-8032 Zurich

Contact: malte.baader@df.uzh.ch

Zeppelin University
Chair of Business Psychology
and Leadership Ethics
Am Seemooser Horn 20
DE-88045 Friedrichshafen

Next



Universität Zürich

zweischen
Wirtschaft Kultur Politik



zeppelin universität

zweischen
Wirtschaft Kultur Politik

2

Important Information About the Study

What to expect: In the first part of the study, you will be asked some questions about yourself. Afterward, you will be confronted with various situations in which you must make a decision and explain it in more detail.

Voluntary participation: Participation in this study is completely voluntary. You can stop the survey at any time by simply closing this window in your web browser. This will not result in any disadvantages for you.

Anonymity: All information is collected in anonymized form, treated strictly confidentially, and used exclusively for research purposes. Under no circumstances is it possible to draw conclusions about your identity.

Data protection: Your data is collected in anonymized form. Since the survey is completely anonymized, subsequent deletion of individual data records is not possible. Consent to data collection and processing is therefore irrevocable.

Personal perspective: Since only your personal perspective counts, there are no right or wrong answers. Please read the instructions carefully, and answer spontaneously and openly.

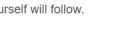
Conditions of participation: To take part in the study, you must be at least 18 years old. In addition, you should currently be employed and not work exclusively on a self-employed basis.

Contact: If you have any questions about this study, you can contact the study management at any time by email at malte.baader@df.uzh.ch.

Have you read and understood the above information carefully, and would you like to take part in this study?

Yes

No



Universität Zürich

zweischen
Wirtschaft Kultur Politik



zeppelin universität

zweischen
Wirtschaft Kultur Politik

3

Thank you for your willingness to take part in this study! First, some questions about yourself will follow.

What is your gender?

Female

Male

Other

Prefer not to say

How old are you?

Next



Universität Zürich

zweischen
Wirtschaft Kultur Politik



zeppelin universität

zweischen
Wirtschaft Kultur Politik

4

Notes on the Study

During the study, we will present you with **two realistic scenarios from everyday company life**. To continue within a situation, simply click the **"Next"** button. With the **"Back"** button, you can return to previous sections within a situation at any time. At the end of each situation, you make a decision between two courses of action. Afterward, you will answer a few brief questions about the respective situation and your decision.

For the study to deliver meaningful results, it is essential that you put yourself into each situation as fully as possible. Therefore, you will shortly have the opportunity to select a profile picture that represents your role in all scenarios.

Important note on participating in the survey
In addition to your participation reward, you will receive an **additional financial compensation**, the amount of which depends on your decisions. At the end of the study, we will randomly select one of your four decisions, which will then determine the amount of your additional payment.

In each situation, we will explain to you how your decision is linked to the additional compensation. During the study, we refer to "points" when we talk about the payment. At the end, your accumulated points will be converted at the **conversion rate of 100 points = 1 euro** and paid out to you via Bilendi in addition to your participation reward.

Please answer the following comprehension question.
In addition to the participation reward, what is your payment if, in the randomly selected situation, you make a decision worth 300 points?

3,00€

1,50€

2,75€

Before you begin with the situations, we will describe the general company environment on the following page.

Figure C4: Start screens cont'd

1

Description of the Company

All situations take place within the automotive group "**McMenba Automobile**". It is a German group that operates and produces internationally.

Due to its size, the group comprises numerous departments – from production and design to sales and strategy. All areas work closely together to develop the best possible vehicles and successfully bring them to market.

During the two scenarios that we present to you on the next pages, your role within the company will change, but all positions are within **McMenba Automobile**.

On the next page, you have the opportunity to select a **profile photo** that represents you throughout the entire study.

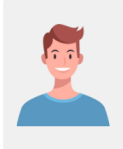
Back

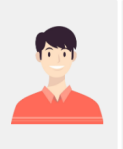
Next

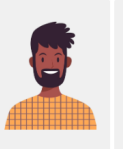
2

Profile Photo

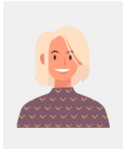
Before we begin with the situations, please click on the person below with whom you feel best represented

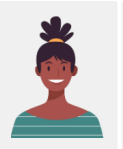


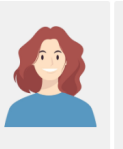














Next

32

Vignette: Car accessories

Figure C5: Workplace scenario: Self

1

Description

You are a salesperson, and your task is to advise customers and sell them cars.

Your ideal goal is to sell many high-quality cars.

To encourage your sales, your salary consists of a fixed portion and an additional commission depending on the number and value of the cars sold.

You are on the phone with a customer...



"Good day, Mrs. Schuster. I have found a great car for you. Based on your information and wishes, I recommend our model CTS."



"Thank you very much for your effort. The car is actually exactly what I am looking for. We had already briefly discussed the additional equipment. I would prefer if it remained with a basic configuration and the costs were kept as low as possible. My maximum budget is €30,000."

2

You are preparing the offer...



"I could increase my commission if I send Mrs. Schuster the offer for the premium model. That would also be within her budget, but it wouldn't be the cheapest option." (+500 points)

"Mrs. Schuster will surely not notice that there is actually also a cheaper model."

"Mrs. Schuster actually wants the basic model to keep the costs low. But then I would have a lower commission." (+100 points)

Figure C6: Workplace scenario: Loyalty + Mistake

1

Description
 You are a salesperson, and your task is to advise customers and sell them cars.
 Your ideal goal is to sell many high-quality cars.
 To encourage your sales, your salary consists of a fixed portion and an additional commission depending on the number and value of the cars sold.
You are on the phone with a customer...



"Good day, Mrs. Schuster. I have found a great car for you. Based on your information and wishes, I recommend our model CTS."

"Thank you very much for your effort. The car is actually exactly what I am looking for. We had already briefly discussed the additional equipment. I would prefer if it remained with a basic configuration and the costs were kept as low as possible. My maximum budget is €30,000."



2

Since the basic model is sufficient for Mrs. Schuster, you send her the corresponding offer by email. She accepts the offer.
 In the last meeting, your supervisor presented your sales statistics to the management and praised you for your highly dedicated and successful work. They are a bit confused, however, because the statistics listed the premium model for Mrs. Schuster, not the basic model. So, after the meeting, you check again the offer you had sent to Mrs. Schuster.



"Oh no, I sent Mrs. Schuster the offer for the premium model and not for the basic model."

"Maybe I should ask my boss what I should do now."

You go to your supervisor to discuss the next steps...

"Hello Mr. Schneider, I made a mistake in customer service. I accidentally sold a customer the premium model instead of the basic model, even though she clearly told me that I should keep the costs low. How should I proceed?"



"We need to drastically strengthen our sales team's revenue. To support your own career and that of your colleagues, I can only strongly advise leaving everything as it is. If we as a team generate the highest revenues this year, it will put us in a good position for the next career steps. The team is counting on you."

3

You reflect...



"If I leave everything as it is, I would get a higher commission. The premium model was also within her budget, but it was not the cheapest option." (+500 points)

"Mrs. Schuster will surely not notice that there is actually also a cheaper model."

"Mrs. Schuster actually wants the basic model to keep the costs low. But then I would have a lower commission." (+100 points)

"My boss already mentioned the sale in front of everyone in the meeting. It would be really embarrassing now if management found out that there had been a sales mistake."

"My boss basically demanded that I sell the higher-priced models and not disclose the mistake, since it benefits the team and the company. And I also have a responsibility toward my teammates."

Figure C7: CQ interventions

Study 1	Study 2								
The following decision options are available to you: A) I sell Mrs. Schuster only the basic package (100 points) B) I sell Mrs. Schuster the premium package (500 points) Before you make a final decision, please consider the following: Would I like the person I see in the mirror after making this decision? Back Next	The following decision options are available to you: A) I sell Mrs. Schuster only the basic package (100 points) B) I sell Mrs. Schuster the premium package (500 points) Which of the decision options would... <table><thead><tr><th></th><th>Option A</th><th>Option B</th><th>I don't know</th></tr></thead><tbody><tr><td>... make me like myself when I see myself in the mirror</td><td>☐</td><td>☐</td><td>☐</td></tr></tbody></table> Back Next		Option A	Option B	I don't know	... make me like myself when I see myself in the mirror	☐	☐	☐
	Option A	Option B	I don't know						
... make me like myself when I see myself in the mirror	☐	☐	☐						

Figure C8: Decision screen

1	2
Please now think about the situation described before you can make your decision on the next page. This decision may have an influence on your additional financial remuneration. Back Next	 What do you decide? ☐ I sell Mrs. Schuster the premium equipment (500 points) ☐ I do not sell Mrs. Schuster the premium equipment (100 points)

Vignette: Hours worked

Figure C9: Workplace scenario: Loyalty

Description

You are employed as a worker in vehicle production. In addition to your fixed monthly salary, you are also paid for overtime. Your usual working time is 6 hours per day. Ideally, you clock your hours with a machine. However, you can also manually enter the hours afterward if you forgot to clock in or out.



"Work in production is really exhausting, especially when you look at the poor pay. With how hard we work, we actually deserve much more."

1

Shortly before the end of your 6-hour shift, a colleague approaches you.

"Hello, I just talked with the others in the team. We all agree that we really work hard and are paid too poorly. That's why we want to each record 2 extra hours in accounting today. If we all do it together, no one will notice. People will just assume that we, as a whole team, had to work overtime. It's just important that we all stick together."



3

You consider...



"Today was exhausting again. I could just skip clocking out today and tomorrow manually record 8 hours instead of 6. That way I would have more pay." (+500 points)

"Accounting is already overloaded and busy with other problems anyway. Nobody will notice if I once record a higher time. Besides, it even looks good if I do overtime. It shows that I take the job seriously and am willing to work longer for the company."

"But in reality, I only worked 6 hours. I should also record it that way." (+100 points)

"But I don't want to let my teammates down. They would surely be upset and, in the worst case, could get into trouble if I don't go along with it."

Figure C10: Workplace scenario: Mistake

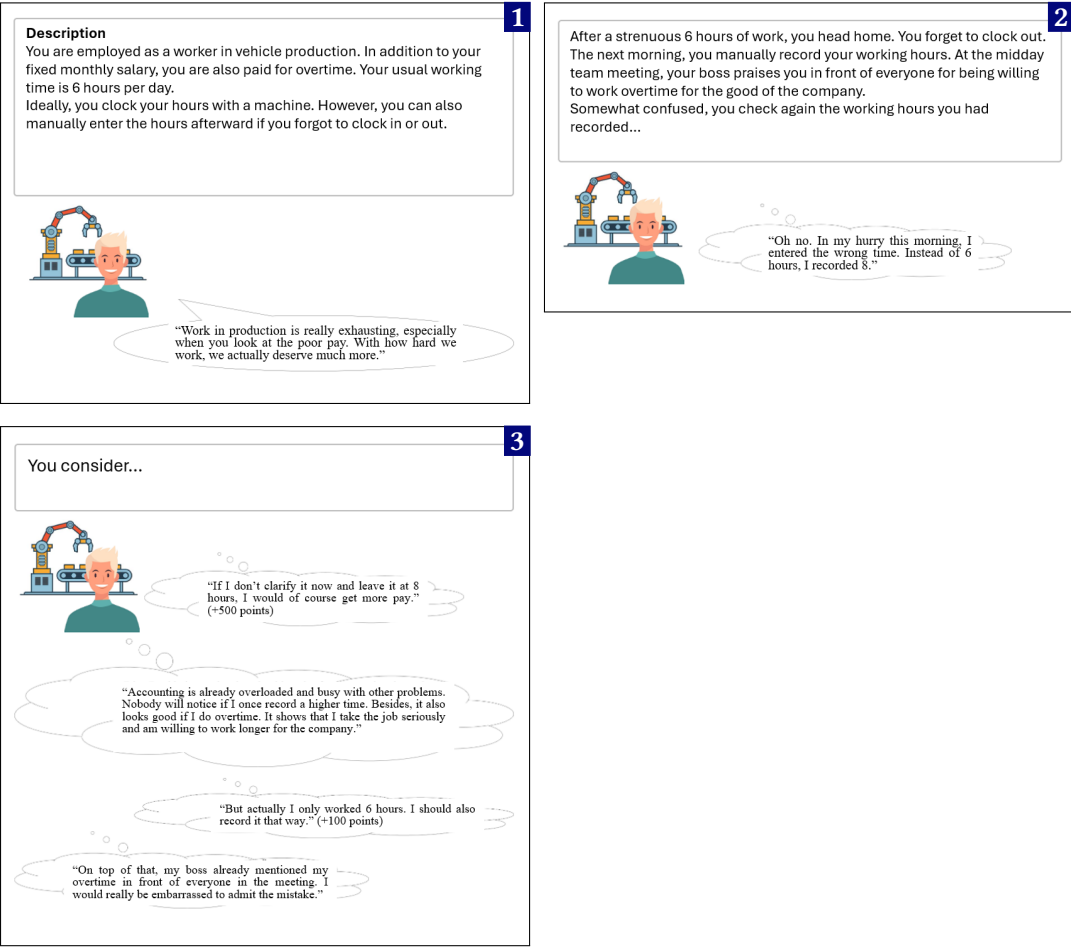


Figure C11: CQ interventions

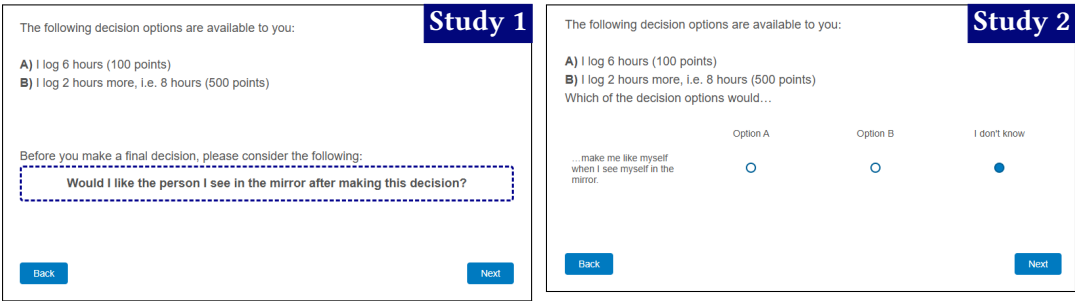
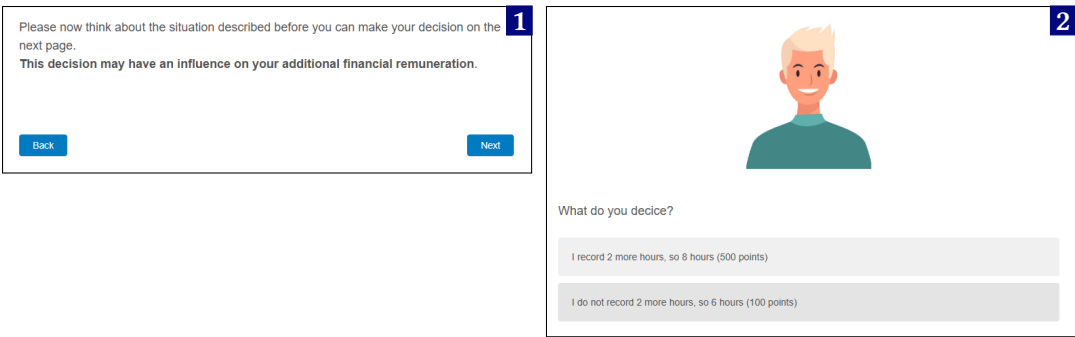


Figure C12: Decision screen



Vignette: Headlights

Figure C13: Workplace scenario: Mistake + Loyalty



Figure C14: Workplace scenario: Loyalty

1

Description
 You work in quality assurance and are mainly responsible for checking vehicle lighting. You and your colleagues are to test the daytime running lights from a supplier for a new car model that is scheduled to be launched next year.
 You receive a call from the supplier...



"Hello, I just wanted to let you know that we have just shipped the order. Hopefully, the lights will arrive with you soon so that you can quickly start the tests. We would also like to invite you on a vacation if everything is handled quickly and without problems."

2

When the supplier's daytime running lights arrive, you begin testing them. The majority of the tests are already passed. However, you now notice a small problem: the durability seems to be 10% shorter than the criteria require. You call the supplier to report the problem...



"That is of course very unfavorable regarding the durability. Luckily, it is not something critical. For our company, a revision would not be manageable. Even though we have been producing high-quality products for decades, financially we are not doing well at the moment. The pandemic caused heavy losses, and we are absolutely dependent on the order with you. Our companies have been working together successfully for 20 years, and there have never been problems before, so I kindly ask you to overlook the issue this time. This should also be in your long-term interest, since we as your supplier want to continue to exist and you won't have to look for new suppliers."

3

You think to yourself...



"The supplier's tests have been good for years, and the problem with the daytime running lights only concerns durability. If I record the tests as passed, I'll even get a vacation from the supplier." (+500 points)

"The problem is not directly noticeable, and no one else will carry out such tests. Therefore, no one can hold me responsible for it."

"On the other hand, it would of course not be appropriate to record the tests as passed. Even if a safety risk is unlikely, we must be accurate in such matters." (+100 points)

"The supplier's situation really does sound bad. If I record the test as failed, it would definitely have drastic effects on their company, and our relationship with the supplier could also suffer."

Figure C15: CQ interventions

The following decision options are available to you:

A) I correct the test, record it as failed, and forgo the vacation (100 points)
B) I do not correct the test and record it as passed (500 points)

Before you make a final decision, please consider the following:

Would I like the person I see in the mirror after making this decision?

Back Next

The following decision options are available to you:

A) I correct the test, record it as failed, and forgo the vacation (100 points)
B) I do not correct the test and record it as passed (500 points)

Which of the decision options would...

	Option A	Option B	I don't know
... make me like myself when I see myself in the mirror.	☐	☐	☐

Back Next

Figure C16: Decision screen

Please now think about the situation described before you can make your decision on the next page.
This decision may have an influence on your additional financial remuneration.

Back Next

What do you decide?

I do not correct myself and record the test as passed (500 points)

I correct myself and record the test as not passed (100 points)

Vignette: Environmental certificates

Figure C17: Workplace scenario: Mistake

1

Description
You are an employee in the purchasing department of the design division. Your task is to select components for the interior fittings. Recently, the government passed a new law that grants a tax advantage to vehicles made from sustainable products. If you quickly find a supplier who can offer sustainable fittings at the same cost, you will receive a bonus. In your search for a potential producer, you first contact a supplier you have been working with for a long time.



"Hello Ms. Müller, my company definitely wants to take advantage of the government's possible tax benefit, and I am currently looking for a supplier of sustainable products. For us, it is especially important that the costs of our production do not increase. Could you possibly help me with that?"



"Of course, we also offer all our products from sustainable production. We have producers in various countries. I'll gladly send you a list of possible candidates and the corresponding costs."

2

After receiving the list from Ms. Müller, you review the various producers and costs and choose a producer that meets your requirements. In a meeting with management, your boss then highlights you as an example of how the company can be sustainable without incurring additional costs. A few days after the meeting, you do further research about the chosen producer...



"Oh no, it looks like the producer does have certificates that are sufficient for the tax benefit, but in reality, they are not working in a truly sustainable way. With more trustworthy producers, however, the costs would be significantly higher."

3

You consider...



"If I simply stick with the producer with the questionable certificates, then I'll also get my promised bonus." (+500 points)

"No one will find out that the producer is actually not working sustainably."

"I should withdraw my recommendation and officially state that it is not possible to get sustainable fittings at the same price. But that would also mean I won't get a bonus." (+100 points)

"But my boss has already praised me for the solution in front of management, and it would be really uncomfortable to withdraw my recommendation now. People will think that I don't know how to do my job."

Figure C18: Workplace scenario: Self

1

Description

You are an employee in the purchasing department of the design division. Your task is to select components for the interior fittings. Recently, the government passed a new law that grants a tax advantage to vehicles made from sustainable products. If you quickly find a supplier who can offer sustainable fittings at the same cost, you will receive a bonus. In your search for a potential producer, you first contact a supplier you have been working with for a long time.



"Hello Ms. Müller, my company definitely wants to take advantage of the government's possible tax benefit, and I am currently looking for a supplier of sustainable products. For us, it is especially important that the costs of our production do not increase. Could you possibly help me with that?"



"Of course, we also offer all our products from sustainable production. We have producers in various countries. I'll gladly send you a list of possible candidates and the corresponding costs."

2

After receiving the list from Ms. Müller, you review the various producers and costs. In your research on the sustainability certificates of some producers, you find that while these certificates are sufficient for the tax advantage, the producers are actually not working in a truly sustainable way. Unfortunately, producers with such certificates are the only ones whose costs would not increase production expenses.



"If I choose one of the producers with the questionable certificates, then I'll also get my promised bonus." (+500 points)

"No one will find out that the producer is actually not working sustainably."

"I should report that it is not possible to get sustainable fittings at the same price. But that would also mean I won't get a bonus." (+100 points)

Figure C19: CQ interventions

Study 1

The following decision options are available to you:

A) I report the new findings and state that it is not possible to find a sustainable producer within the budget (100 points)

B) I do not report the new findings and stick with one of the producers that do not operate sustainably (500 points)

Before you make a final decision, please consider the following:

Would I like the person I see in the mirror after making this decision?

BackNext

Study 2

The following decision options are available to you:

A) I report the new findings and state that it is not possible to find a sustainable producer within the budget (100 points)

B) I do not report the new findings and stick with one of the producers that do not operate sustainably (500 points)

Which of the decision options would...

Option A

Option B

I don't know

... make me like myself when I see myself in the mirror.

☐

☐

☐

BackNext

Figure C20: Decision screen

1

Please now think about the situation described before you can make your decision on the next page.

This decision may have an influence on your additional financial remuneration.

BackNext

2



What do you decide?

I don't report the new findings and stick with the producer who doesn't operate sustainably (500 points)

I report the new findings and don't stick with the producer who doesn't operate sustainably (100 points)

Post intervention survey

Figure C21: Decision certainty and completion - After each vignette

1

How certain were you about the decision?

Not at all sure
1

2

3

Moderately sure
4

5

6

Very sure
7

Back

Next

2

You have now completed the situation.
To continue with the study, please click on "Next".

Next

Figure C22: Post intervention survey - After last vignette

1

Please now think about the decision you made at the end of the last situation.

How did you feel afterwards?

	not at all 1	a little 2	moderately 3	predominantly 4	completely 5
confused	☐	☐	☐	☐	☐
distressed	☐	☐	☐	☐	☐
ashamed	☐	☐	☐	☐	☐
enthusiastic	☐	☐	☐	☐	☐
guilty	☐	☐	☐	☐	☐
nervous	☐	☐	☐	☐	☐
proud	☐	☐	☐	☐	☐
excited	☐	☐	☐	☐	☐
strong	☐	☐	☐	☐	☐
determined	☐	☐	☐	☐	☐

Next

2

Please now think about the decision you made at the end of the last situation.

To what extent do the following statements apply to you?

	does not apply at all 1	2	applies moderately 3	4	applies very much 5
I am confident that I have made the right decision.	☐	☐	☐	☐	☐
I would have preferred to leave the decision to others.	☐	☐	☐	☐	☐
I found this decision burdensome.	☐	☐	☐	☐	☐
I feel up to such situations.	☐	☐	☐	☐	☐
This decision feels right.	☐	☐	☐	☐	☐

Next

3

To what extent do the following statements apply to you?

	does not apply at all 1	2	applies moderately 3	4	applies very much 5
If it helps employees get their work done, it is acceptable to deceive others.	☐	☐	☐	☐	☐
When an employee has to stretch the truth in order to get their job done, they cannot be accused of lying.	☐	☐	☐	☐	☐
Employees cannot be accused of exaggerating the truth if all other employees do so.	☐	☐	☐	☐	☐
Employees cannot be held responsible for misconduct if they feel that they were pressured into doing it by others.	☐	☐	☐	☐	☐
It is acceptable to stretch the truth in order to protect the company.	☐	☐	☐	☐	☐

Next

4

You have now completed 90% of our questionnaire.

A few questions about your current company now follow. If you are not currently employed, please base the questions on the company where you were most recently employed.

Back

Next

Figure C23: Post intervention survey cont'd

1

Do you currently work full-time or part-time?

If you are not currently employed: Please base your answer on your most recent job.

Full-time

Part-time at %

How many years have you been working?

Less than one year

1 year to under 3 years

3 to under 6 years

6 to under 10 years

10 to under 15 years

15 to under 20 years

20 years or more

3

Thank you very much for the information you have provided so far!

Before we conclude this survey, it is important to us at this point to emphasise that we can only include responses from people who have completed the questionnaire attentively and seriously. Otherwise the data is distorted and unusable for the analysis.

Please therefore answer this question truthfully now. Regardless of your answer, you will receive the incentive for this study in any case. Thank you very much for your assistance!

Yes, I answered the questionnaire attentively and to the best of my knowledge and belief. My responses may be used for the analysis.

No, I did not answer the questionnaire attentively and seriously. My responses should preferably not be used for the analysis.

[Next](#)

2

Notes: If you hold several jobs at the same time: Please answer the questions with reference to the job with the highest number of hours.

If you are not currently employed: Please base the question on your most recent job.

How long have you been working at the company where you are currently employed?

Less than one year

1 year to under 3 years

3 to under 6 years

6 to under 10 years

10 years or more

How many employees does the company you currently work for have?

Fewer than 50 employees

50 to 249 employees

250 to 10,000 employees

More than 10,000 employees

[Back](#)
[Next](#)

4

You have now completed 100% of our questionnaire.

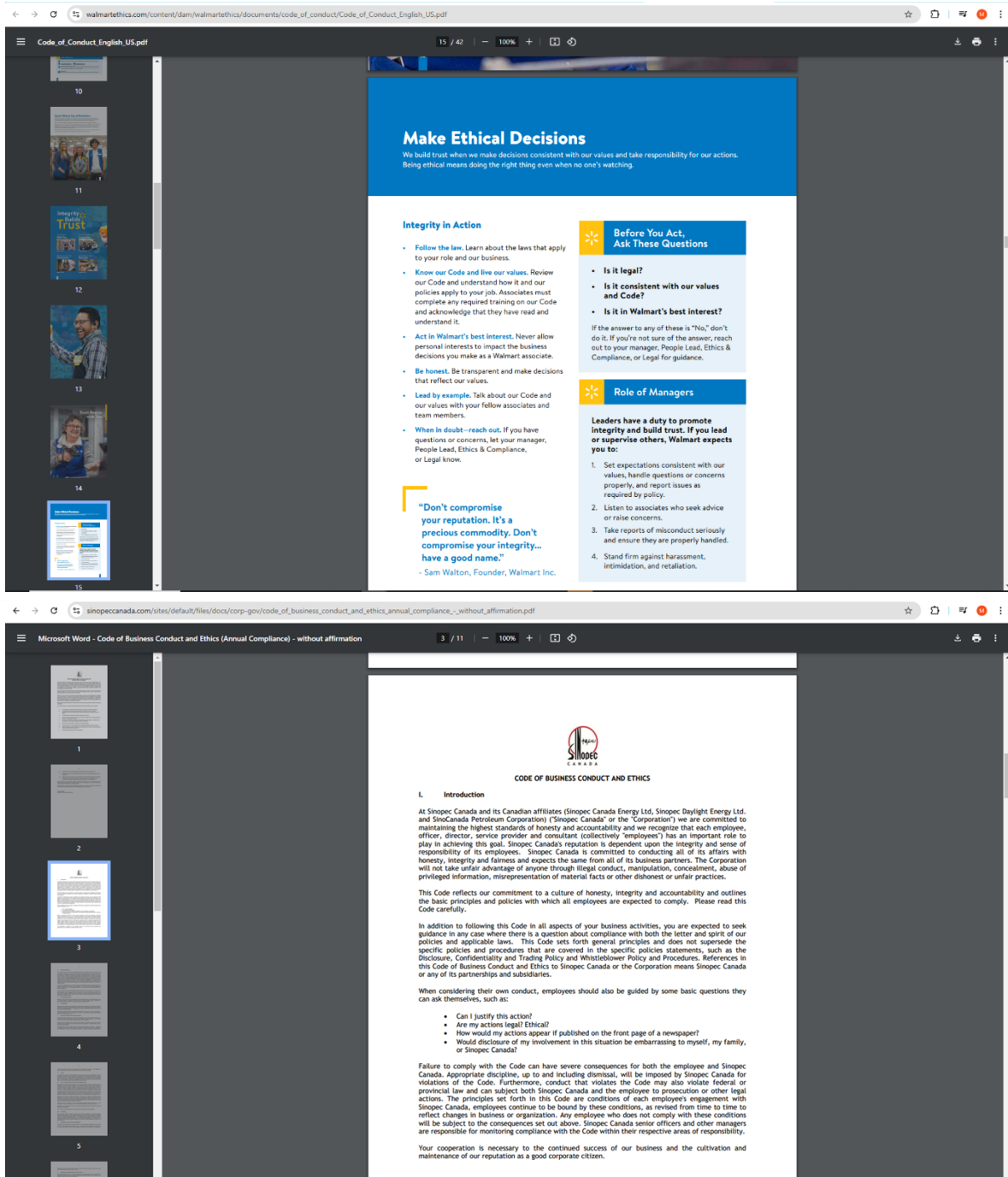
Do you have any **comments** on this survey or anything you would like to tell us? You can also reach the study team at malte.baader@df.uzh.ch.

Please click on "Next" to complete the survey.

[Next](#)

D Examples of Contemplation Questions

Figure D24: Contemplation questions example at Walmart (top) and Sinopec (bottom)



E Pre-registrations

Below we include pre-registrations for Study 1 (<https://osf.io/gk2er>) and Study 2 (<https://osf.io/g3ewa>) respectively. Comments within the pre-registrations highlight deviations from the pre-registration within the paper.

E.1 Study 1

Hypotheses

1. On average, CQs reduce unethical choices compared to an absence of CQs.
2. Along with the effects on unethical decisions, we also expect CQs to reduce conflict perception, decision uncertainty, stress and decision avoidance. CQs are effective across studied dilemma situations (self, mistake, loyalty – see design plan for definitions).
3. CQs differ in their effectiveness. Among the four categories of CQs under evaluation, we expect Universal Principles > Inner Voice > Law & Public > Policies > Peers (with “>” indicating greater effectiveness). This order is based on a previously conducted survey, thus the final order might differ. Thus, whilst expecting heterogeneity by category, we do not have a strong prior regarding the order.
4. Additionally, we also expect heterogeneity in the effectiveness of CQs within each category.
5. Measured personality traits in wave 1 (see below) serve as moderators in the effectiveness of CQs.

E.1.1 Design Plan

Study type

Experiment — A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Is there any additional blinding in this study?

No data

Study design

We have a between subject design with 1 factor (CQ) with 16 levels. 15 different CQs and a control group where the CQ is absent. We conduct our study using incentivised vignettes. We employ a total of 4 distinct vignettes, each in 4 variations to capture a range of dilemma situations. The first variation is labelled *self*, where the vignette describes a scenario with a temptation of a higher payoff. The second is *mistake*, where a mistake has occurred and correcting it would be ethical. The third is a conflict of *loyalty*, where within the vignette subjects receive pressure from another individual. Lastly, we combine mistake and loyalty in the fourth variation. This yields a total of 4×4 vignettes that all have the same incentive structure (higher payoff for unethical decision). Each vignette is followed by a binary choice to behave ethically or unethically. At the point of decision, we present participants with a randomly drawn CQ to examine its effect on decision making. Each participant experiences four different vignettes in four different variations with a unique CQ. In addition to choice data, we also collect information on decision certainty, decision avoidance and stress as secondary outcome variables after each decision.

Randomization

We will use simple randomisation, where each participant will be randomly assigned to one of the 15 CQs (or control) and 4 of the vignettes according to the design described above. Using Qualtrics, the randomisation ensures somewhat even distribution across treatments.

E.1.2 Sampling Plan

Existing Data

Registration prior to creation of data

Explanation of existing data

No data

Data collection procedures

Participants will be recruited through the panel service provider Bilendi Respondi. Participants reside in Germany, are at least 18 years old and currently in employment (at least 6 months in the past 2 years). To study personality traits, we collect data across two waves. First, we conduct a number of survey measures (see details below), and then the main study one week later. In total, participants receive €3 for participating in the study and can earn an additional €5 depending on their decisions. We aim at a mean payment of €15 per hour. Bilendi Respondi will recruit participants via E-Mail. The study is programmed using Qualtrics and is accessed directly by the participants.

Deviation from pre-registration. Eligibility for participations additionally required participants to be aged ≤ 67 and not self-employed.

Sample size

Our target sample size is 1,920 participants evenly distributed across dilemma situations and CQs. To achieve this sample, we first recruit 2,500 participants in wave 1, hoping to have 2,000 remaining in wave 2 accounting for attrition.

Sample size rationale

Our aim is to sample 30 observations per possible combination of dilemma and CQ. Since we are utilising 4 dilemmas in 4 variations, we have 16 dilemmas. In addition, we examine 15 CQs (+ 1 control), yielding $16 \times 16 = 256$ possible combinations. However, since each subject completes 4 separate situations, we require $256 / 4 \times 30 = 1,920$ participants. This allows us to compare categories of dilemmas and CQs with substantive statistical power whilst also enabling us to compare single combinations of dilemmas and CQs.

Stopping rule

No data

E.1.3 Variables

Manipulated variables

We manipulate the presence of CQs; this is our exogenous treatment variation.

Measured variables

The main outcome variable is the share of ethical decisions per dilemma variation. In addition, we also conduct survey questions regarding decision certainty (3 items), decision avoidance (3 items) and stress (3 items) after every decision. We also study the effect of CQs on lying in a die incentivised die roll task. In wave 1 we measure self-efficacy (3 items), risk (2 items), social preferences (3 items), cognitive skill (2 items), moral identity (20 items), protected values (12 items), need for cognition (4 items), the HEXACO (24 items) and additional individual and company demographics.

Indices

We will treat the utilised scales as they have been used within the literature. In case there are no clear guidelines from the literature, we take the mean of the survey items regarding the different dimensions outlined in the previous question. We allow ourselves to exclude items as a result of factor analyses or substantive cross-correlations.

E.1.4 Analysis Plan

Statistical models

We will use Chi-square tests to analyse the effect of CQs on the share of ethical decisions. Regarding our secondary outcome variables, we utilise Wilcoxon rank-sum tests to explore differences in the reported Likert scales. We will do so in aggregate across all CQs and then examine CQs based on their categories. Lastly, we will also examine each CQ individually to explore additional heterogeneities. Similarly, we will also examine differences across dilemma situations (pooled by category and then individually). Whenever we pool data, we create means of the outcome variables in line with the level of aggregation. For more detailed analyses, we utilise Probit regressions with choice as the dependent variable and errors clustered on the individual level to control for individual demographics and study other individual-level variables (risk, social preferences, lying aversion). For the secondary outcome variables, we use Tobit regressions, to account for censoring at upper (7) and lower (1) bounds. Dummies for CQ and dilemmas (depending on level of aggregation) capture our treatment effects. To explore personality traits, we regress measured characteristics on ethical choice or perception using probit and Tobit specifications. To examine moderation, we explore mean personality traits between treatment and control group conditional on their ethical choice.

Deviation from pre-registration. The paper reports random-effects probit models (and linear random-effects models for the secondary outcomes) as the primary specification, rather than pooled probit / Tobit with individual-clustered standard errors. We report the pre-registered pooled-probit models as robustness checks in Appendix A.1; results are materially unchanged.

Transformations

No data

Inference criteria

We will use the standard $p < 0.01$, $p < 0.05$ and $p < 0.1$ criteria for determining if any of the tests suggest that the results are significantly different from zero. We will report all relevant levels of significance in the article.

Deviation from pre-registration. We decided to adopt a more stringent significance criteria in the paper at the $p < 0.05$, $p < 0.01$ and $p < 0.001$ levels (rather than $p < 0.01$, $p < 0.05$ and $p < 0.1$).

Data exclusion

We exclude data when there is clear evidence that participants did not complete the study thoroughly. Left/Right-clicking on surveys, excessively fast/slow completion or failed attention checks will lead to exclusion. However, we will also examine the results including all participants to explore robustness. If participants indicated a complication when interacting with the study (using our feedback option or reaching out to us via email), we will exclude them from the analysis.

Missing data

If a participant does not complete the study, they are excluded from the analysis.

Exploratory analysis

1. To benchmark our vignettes, we also expect CQs to increase honesty in the incentivised die roll task. Since some of the CQs are worded explicitly for company contexts, it is unclear whether these will also be effective in a die roll task.
2. In addition to examining effects of CQs on the mentioned outcome variables, we will also examine whether the secondary outcomes can serve as moderators. We conjecture that

CQs only affect ethical decision making when there are substantive changes in decision certainty, decision avoidance or stress.

3. We further conjecture that stress might increase in the presence of CQs. Thus, despite a shift towards ethical behaviour, there may be a cost in terms of stress. To study this, we also conduct a ‘cost-benefit’ analysis to study “costs” of more ethical decisions.
4. Lastly, since we previously collected hypothetical survey data across two studies, asking about the perceived utility of different CQs, we will compare our experimental findings with the predictions of this survey. Moreover, we also collected information on which CQs are used in various international companies. This information will also be compared to the experimental effectiveness in our study presented above.

Other

No data

E.2 Study 2

Hypotheses

1. Irrespective of the presence of CQs, participants with pro-ethical responses behave more ethically, and vice versa for pro-unethical responses.
2. CQs do not affect ethical decisions on average without taking pro-ethical or pro-unethical responses into account.
3. Pro-ethical responses to CQs are associated with an increase in ethical decisions, whereas pro-unethical responses lead to a decrease in ethical behaviour.

E.2.1 Design Plan

Study type

Experiment — A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Is there any additional blinding in this study?

No

Study design

We employ a within-participant design using 1 factor (CQ) with 13 levels. 5 individual CQs or 5 CQs together (6 levels), both in a positive or negative frame ($\times 2$ level) and a control group (1 level) where the CQ is absent. As in Study 1, we conduct our study using incentivised vignettes. We include a within-participant design where each person first encounters 4 vignettes in 4 distinct situations as in Study 1. In Wave 1, all participants make a binary decision (with higher payoff for unethical decision) without facing any CQs. In Wave 2, participants are then randomly assigned to one of the 13 treatment groups (5 $\times$ positive single CQ, 5 $\times$ negative single CQ, 1 $\times$ positive 5 CQs, 1 $\times$ negative 5 CQs, or Control condition) and face parallel vignettes with another 4 ethical decisions. The set-up of the vignettes is described below in section ‘data collection procedures’. At the end of each of the two waves we also collect data on moral disengagement, decision confidence and emotions. Lastly, in a third wave, we elicit data on personality traits.

Randomization

We will use simple randomisation in Wave 2, where each participant will be randomly assigned to one of the 13 CQ- or control conditions. Using Qualtrics, the randomisation ensures somewhat even distribution across treatments.

E.2.2 Sampling Plan

Existing Data

Registration prior to creation of data

Explanation of existing data

We have two sources of existing data. First, we conducted Study 1, registered in the same OSF project. We did not find any statistically significant effect of CQs. We consider a lack of

engagement with the CQs to be one possible explanation for this finding. Consequently, the evaluation of CQs the participants were confronted with is a key part of Study 2. Motivated by this we already conducted a pilot study, where we asked participants to respond to CQs, allowing us to classify them as pro-ethical or pro-unethical and refine our methodology for this registered study.

Data collection procedures

Participants will be recruited through the panel service provider Bilendi Respondi. Participants reside in Germany, are at least 18 years old and currently in employment (at least 6 months in the past 2 years). Overall, we collect data across three waves. In both, Wave 1 and 2, we conduct our within-participant design. In each wave participants face 4 distinct ethical dilemmas and corresponding decisions. The dilemmas in each Wave are parallel vignettes (labelled A and B) that half of the participants experience in order AB and the other half in order BA. This allows us to control for potential effects due to differences in the parallel vignettes. In a subsequent Wave 3, we collect data on personality traits.

Sample size

Our target sample size is 100 observations for each treatment condition. Since we have 13 different conditions, this amounts to 1,300 observations in total. Considering attrition between waves of around 20% and using 3 waves, we invite 2,050 participants in Wave 1 to ensure our target of 100 per treatment by Wave 3.

Sample size rationale

No data

Stopping rule

No data

E.2.3 Variables

Manipulated variables

We manipulate the presence of CQs; this is our exogenous treatment variation. 5 CQs are presented either individually (one for each of 5 identified categories) or as 5 simultaneously.

We also vary whether presented CQs are positively or negatively framed. In the case of 5 CQs, all CQs are framed either positively or negatively.

Deviation from pre-registration. As positive and negative CQ framing showed no differential effects, the 13 treatment conditions are pooled across frame in the analysis (five single-CQ conditions, one multiple-CQ condition, and control).

Measured variables

The main outcome variable is the share of ethical decisions in each treatment condition. In Wave 2, we also elicit participants' responses to the CQ they faced. This will allow us to categorise participants into pro-ethical and pro-unethical types in treatment and control conditions. In addition, at the end of Wave 1 and Wave 2 we also conduct survey questions regarding moral disengagement, decision confidence and emotional affect after the final decision of each wave. In Wave 3, we collect data on moral identity, regulatory focus, social value orientation and HEXACO as personality traits. Lastly, in addition to demographics, we also ask participants at the end of Wave 2 to remember the CQ they faced as a manipulation check.

Indices

We will treat utilised scales as they have been used within the literature. In case there are no clear guidelines from the literature, we take the mean of the survey items regarding the different dimensions outlined in the previous question. We allow ourselves to exclude items as a result of factor analyses or substantive cross-correlations.

E.2.4 Analysis Plan

Statistical models

We will use Chi-square tests to analyse the effect of CQs on the share of ethical decisions. This will also be used when examining pro-ethical and pro-unethical respondents separately. Regarding secondary outcome variables, we utilise Wilcoxon rank-sum tests to explore differences in the reported Likert scales. We will do so in aggregate across all CQs and then examine CQs across treatment conditions. Meaning, we split positively and negatively framed CQs, as well as single and multiple CQs. We also examine each CQ treatment individually. Since a key focus is pro-ethical and pro-unethical responses to CQs, this constitutes an endogenous sample split. We therefore utilise the split in pro-ethical and pro-unethical from Wave 2, to also split the sample in the control condition in Wave 1. This ensures that any differences

between control and treatment conditions are not due to unobserved individual characteristics, but because of our CQ manipulation. Whenever we pool data, we create means of the outcome variables in line with the level of aggregation. For more detailed analyses, we utilise Probit regressions with ethical choice (= 1) as the dependent variable and errors clustered on the individual level to control for individual demographics and study other individual-level variables. For secondary outcome variables, we use Tobit regressions, to account for censoring at upper and lower bounds. Dummies for CQs (depending on level of aggregation) capture our treatment effects. To explore personality traits, we regress measured characteristics on ethical choice or perception using probit and Tobit specifications. To examine moderation, we explore mean personality traits between treatment and control group conditional on their ethical choice.

Deviation from pre-registration. The confirmatory analysis reported in the paper moderates the effect of CQs by *moral identity*, rather than by the pre-registered split into pro-ethical and pro-unethical responses to the CQs. Since responses to CQs are endogenous to the decision-making process, we decided to focus on moderation by personality traits instead. In the data we observe a significant correlation between pro-ethical and pro-unethical responses to positively framed CQs and moral identity ($r = 0.19$, $p < 0.001$), which is why we opted for moderation by moral identity in the end. We restrict this correlation to positively framed CQs because response behaviour for negatively framed CQs is inconsistent due to double negatives. This moral identity moderation, and the accompanying decomposition of switching behaviour across waves (Section 5.4), were not pre-registered; robustness using honesty humility and social value orientation is reported in the Appendix. As in Study 1, the primary models are random-effects probit rather than pooled/clustered probit (reported as robustness), and secondary outcomes use linear rather than Tobit specifications.

Transformations

No data

Inference criteria

We will use the $p < 0.01$, $p < 0.05$ and $p < 0.1$ criteria for determining if any of the tests suggest that the results are significantly different from zero. We will report all relevant levels of significance in the article.

Deviation from pre-registration. We decided to adopt a more stringent significance criteria in the paper at the $p < 0.05$, $p < 0.01$ and $p < 0.001$ levels (rather than $p < 0.01$, $p < 0.05$ and $p < 0.1$).

Data exclusion

We exclude data when there is clear evidence that participants did not complete the study thoroughly. Left/Right-clicking on surveys, excessively fast/slow completion or failed attention checks will lead to exclusion. However, we will also examine the results including all participants to explore robustness. If participants indicated a complication when interacting with the study (using our feedback option or reaching out to us via email), we will exclude them from the analysis.

Missing data

If a participant does not complete the study, they are excluded from the analysis.

Exploratory analysis

Since the focus of Study 2 is on the engagement with the CQs, we consider analyses related to this as our primary objective, registered under “hypotheses” above. Here we list analyses that consider other treatment manipulations we include (i.e. positive, negative framing and multiple CQs) and possible mechanisms we aim to explore.

1. We expect differences in positively and negatively framed CQs. Positive framing is expected to lead to a more pronounced effect of CQs compared to a negative frame.
2. We also expect pro-ethical respondents facing multiple CQs to make more ethical decisions relative to no CQs or single CQ-Treatments.
3. We hypothesise moral disengagement to be an underlying mechanism in explaining a reduction in ethical decisions for participants with pro-unethical responses. Thus, higher levels of moral disengagement for pro-unethical responses and pro-ethical responses combined with unethical decisions.
4. Elicited personality traits allow us to characterise participants that respond pro-ethically and pro-unethically. We will also analyse them as potential moderators in the effect of CQs.

Other

No data

E.3 Outcomes of Pre-registered Hypotheses

The table below reports the outcome of each pre-registered main hypothesis. Statistics match the main-text analyses where available and the pre-registered test otherwise.

#	Pre-registered prediction	Result	Verdict
Study 1			
H1	CQs increase ethical (reduce unethical) choices on average	No effect ($z = 1.34$, $p = 0.18$)	Not supported
H2	CQs reduce conflict/uncertainty, stress, and decision avoidance	No reduction (Wilcoxon: stress $z = -1.6$; difficulty $z = -1.8$; avoidance $z = -0.7$; all $p > 0.07$)	Not supported
H3	Effectiveness differs across categories, in a predicted order	No joint difference ($\chi^2(4) = 8.4$, $p = 0.08$)	Not supported
H4	Effectiveness differs within categories	No joint heterogeneity ($\chi^2(15) = 23.9$, $p = 0.07$); within-category n.s. except Company Policy ($\chi^2(2) = 9.7$, $p = 0.008$)	Not supported
H5	Personality traits moderate CQ effectiveness	Moral identity $\times$ CQ n.s. ($z = 1.3$, $p = 0.18$)	Not supported
Study 2			
H1	Pro-ethical responders act more ethically (and vice versa)	0.74 vs. 0.69 ethical share ($z = 5.6$, $p < 0.001$)	Supported
H2	No average effect of CQs on ethical decisions	$z = -1.81$, $p = 0.07$	Supported
H3	Pro-ethical responses raise, pro-unethical responses lower ethical choices	CQ effect +5.6pp (pro-ethical) vs. -6.6pp (pro-unethical), both $p < 0.001$; interaction $z = 7.2$, $p < 0.001$	Supported

E.4 Secondary and Exploratory Measures

Study 1

Die-roll benchmark. As a benchmark outside the workplace vignettes, we pre-registered whether CQs increase honesty in an incentivized die-roll task. They do not: 10.3% of participants exposed to a CQ misreport the die versus 5.8% in the control, a difference that is not significant and, if anything, runs in the opposite direction ($\chi^2(1) = 2.9$, $p = 0.09$; the control cell is small, $n = 139$). We find no differences across CQ categories ($\chi^2(5) = 3.3$, $p = 0.66$).

There are also very low levels of dishonesty (9.96%) in the die roll task across the sample, likely related to low incentives for this task.

Secondary outcomes as moderators. We further conjectured that CQs affect ethical choices only when they shift decision certainty, stress, or avoidance. Interacting CQ presence with each secondary outcome yields no significant moderation (joint $\chi^2(3) = 2.8, p = 0.42$).

Study 2

Moral disengagement. We pre-registered an exploratory hypothesis with moral disengagement as a possible mechanism. State moral disengagement is higher for unethical choices (mean 2.82) than for ethical decisions (2.10, difference $z = -29.6, p < 0.001$), but it is unaffected by CQs ($p = 0.41$). Moral disengagement thus tracks the decision itself rather than the initial inclination of participants.

Regulatory focus. Consistent with our conjecture that traits both characterise responders and moderate CQ effects, prevention focus predicts ethical choices ($z = 2.5, p = 0.014$), is higher among pro-ethical responders ($p = 0.006$), and moderates the CQ effect much like moral identity (prevention $\times$ CQ, $z = 2.8, p = 0.005$). Promotion focus does not significantly predict ethical choices ($p = 0.11$), distinguish responder types, or moderate the CQ effect.

Moral motivation. The four moral-motivation subscales (Krettenauer & Casey, 2015) enter our regressions as individual controls and are not analysed as focal moderators.

Recall. Overall, 53% of treated participants recalled the CQ. Because no comparable recall measure was collected in Study 1 and recall was associated with moral identity, we do not use recall as a measure of differential engagement across the two studies.

F The Role of Detectability

Since it is plausible that the probability of detection affects ethical decision making as well as interacts with the effectiveness of specific CQs, we tested these dimensions, i.e., the chance of having the unethical action being observed in a distinct sub-study. Here, we are particularly interested in whether specific CQs that relate to the public (i.e., newspaper question) have a different effect than CQs that relate to other reference groups. Our experimental design follows a within-subject set-up as utilized in Study 2. In Wave 1, participants face ethical dilemmas without exposure to CQs, in a subsequent wave (Wave 2) they face a parallel set of vignettes, now interacting with a single or multiple CQs. Participants either face the *self* CQ, the *law and reputation* CQ or all 5 CQs simultaneously. The vignettes are identical to Study 2. See the main text for details.

Crucially, we now vary the chance of their action being observed. In the low detectability scenario, participants have a 1% chance that their action is being shown to another person. In the high detectability case, this likelihood is 25%. We implement this treatment variation for real, by recruiting additional participants that observe the action of our "decision sample" with the above probabilities.

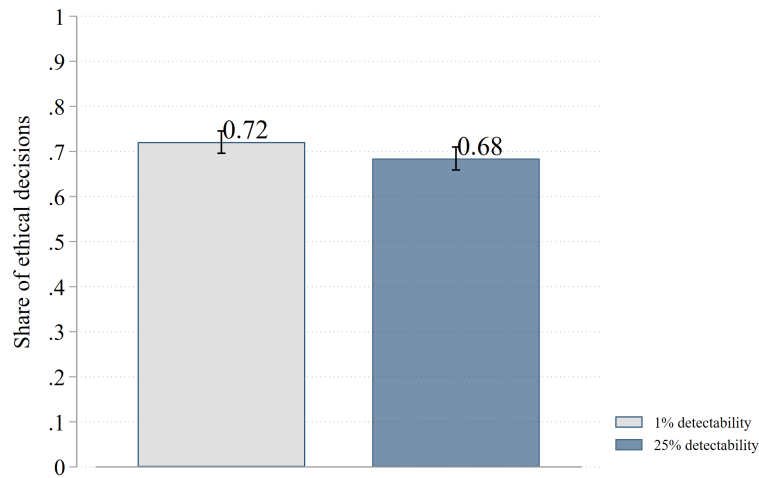
This leaves us with a 3x2 treatment design, with three different CQ treatments and two detectability levels. Our hypotheses are that higher levels of detectability increase the share of ethical decisions. Moreover, we expect differential effects of the reputation CQ to the self CQ.

We again recruit our sample via Bilendi (Oct to Nov 2024) following the same inclusion criteria from Study 1 and Study 2. In total, we recruited $N = 323$ participants across the two waves, with each participant making four decisions. This leaves us with around $N = 50$ participants and 200 decisions per treatment.

F.1 Results

We do not find a statistical difference in the share of ethical decisions between the low and high detectability treatment. If at all higher detectability leads to a marginal reduction in ethical decisions (see difference between blue and gray).

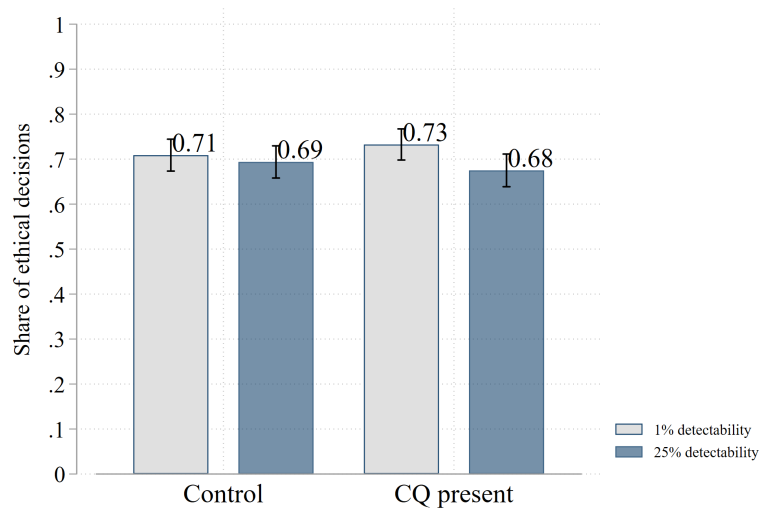
Figure F25: The effect of detectability on the share of ethical decisions



Note. Here we do not differentiate across treatment effects of CQs. The figure solely indicates the effects of detectability on the share of ethical decisions.

When examining differences in effects of CQs based on detectability, we also find no differential effects (compare gray bars and blue bars). Below all CQs are pooled, but examining the three different groups separately also shows no significant effects of CQs across levels of detectability.

Figure F26: The effect of detectability with and without CQs



Note. We now split the sample by Waves (i.e. the presence of CQs).

About ECGI

ECGI is an international non-profit membership organisation that provides a platform for debate and dialogue among academics, policymakers, and business leaders, with a focus on major corporate governance, ESG, and stewardship issues.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.

www.ecgi.global

ECGI Working Paper Series in Finance

Editorial Board

Editor

Nadya Malenko, Professor of Finance, Boston College

Consulting Editors

Renée Adams, Professor of Finance, University of Oxford

Franklin Allen, Professor of Finance and Economics, Imperial London

Julian Franks, Professor of Finance, London Business School

Mireia Giné, Professor of Finance, IESE Business School

Marco Pagano, Professor of Economics, Facoltà di Economia Università di Napoli
Federico II

Series Manager

Asif Malik, ECGI Working Paper Series Manager

<https://www.ecgi.global/publications/working-papers>