WORKING PAPER

MORAL CLEANSING: THE ROLE OF SALIENCY IN MISBEHAVIOR

Matthieu Légeret

UNIVERSITY OF LAUSANNE

Quartier de Chamberonne

1015 Lausanne

Switzerland

May 2021

PRELIMINARY AND INCOMPLETE

Abstract

Although moral cleansing has frequently been documented, existing evidence remains mixed and the prerequisite for its occurrence remain unclear. We argue that the saliency of moral misbehavior is decisive: we hypothesize that moral cleansing is only triggered if people strongly deviate from their own moral identity. We test our conjecture using an incentivized experiment. We find that moral cleansing indeed occurs following salient deviations. Moreover, we also observe that cleansing behaviors depend on participants' initial morality.

1. Introduction

People often engage in morally inconsistent behaviors that seem paradoxical. For example, individuals who engage in morally questionable actions sometimes follow up on these actions with particularly moral behavior. Such inconsistent patterns may be the consequence of an attempt to recover ones moral identity ensuing a misbehavior. Individuals who realize that their past behavior conflicts with their own moral code tend to experience disgust (e.g., Chapman & Anderson, 2013) and may respond with what has been termed "moral cleansing" (Lee & Schwarz, 2021, Gneezy, Imas, & Madarász, 2014, Sachdeva, Iliev, & Medin, 2009).[1]

Although ample evidence for morally inconsistent behavior exists (e.g., Effron, Cameron, & Monin, 2009; Merritt, Effron, & Monin, 2010; Zhong & Liljenquist, 2006), investigations of the underlying mechanisms yielded mixed results: the same people who engage in moral cleansing in some cases don't do so in other situation, but the reasons remain unclear (Dolan & Galizzi, 2015). We argue that one factor which is decisive for whether moral cleansing is triggered or not is the saliency of the misbehavior. We hypothesize that large deviations from one's moral identity require repair through moral cleansing, whereas small deviations remain unnoticed or are tolerated and therefore tend to lead to continued misbehavior.

We test our conjecture with an incentivized experiment in which we first measure participants initial morality before we expose them to a stimulus that aims at triggering unexpected reactions. In the final phase we then observe participants' behavior once the stimulus has been removed. Our treatments vary the intensity of the stimulus to trigger deviations of different magnitude and saliency.

---

[1] The opposite pattern—moral behavior followed by immoral actions—is called "moral licensing".

2. Theoretical background

While individuals like to believe they are consistent, especially when it comes to morality, research shows that they systematically deviate from their own moral code (Bazerman & Gino, 2012; Bazerman & Sezer, 2016; Bazerman & Tenbrunsel, 2011). As such, individuals can be described as boundedly ethical (see also Chugh & Bazerman, 2007; Chugh, Bazerman, & Banaji, 2005). Echoing the seminal work of Herbert Simon (1955, 1956), bounded ethicality assumes that individuals are only boundedly aware of their own limitations as well as the moral aspect of certain situations[2]. For instance, individuals believe they are moral and objective (Messick & Bazerman, 1996; Tenbrunsel, 1998). Being boundedly ethical then serves a self-preservation purpose – it protects the moral self (Chugh et al., 2005). However, this distorted self-image is exactly what prevents individuals from seeing relevant information when making moral decisions (Bazerman & Sezer, 2016; Tenbrunsel & Messick, 2004; Zhang, Fletcher, Gino, & Bazerman, 2015), as they tend not to see when a moral decision involves conflicts of interests (Moore, Tetlock, Tanlu, & Bazerman, 2006) and fail to focus on morally relevant information (Bazerman & Sezer, 2016; Chugh & Bazerman, 2007). While the framework of bounded ethicality places the moral self and moral self-image at the centre of its explanations, it says little about the psychological mechanisms underlying misbehaviours nor the reaction to misbehaviours.

Part of the psychological literature on (im)moral behaviours has focused on individuals' reactions to their own misbehaviours. Indeed, for most individuals such misbehaviours come as a paradox between one's moral identity and behaviour and have to be rationalized one way or another (Tenbrunsel & Messick, 2004). There are various ways through which individuals

---

[2] While certain behavioral patterns appear to be internally inconsistent, scholars have argued that they may result from the application of morally neutral decision strategies to moral contexts and can therefore be understood by taking the environment into account (see Fleischhut & Gigerenzer, 2013).

deal with identity conflicts ranging from psychological to behavioural. Some scholars have focused on how individuals detach themselves from their decisions, their outcomes, or their victims (e.g., Bandura, 1990, 1999, 2002; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996), but the part of the literature that is most relevant for this work has focused on how individuals compensate for their own misbehaviours (moral cleansing[3]). Moral cleansing is used to describe the behaviour of an individual who behaves morally after having behaved immorally, in order to compensate for the previous behaviour (Sachdeva et al., 2009; Stone, Aronson, Crain, Winslow, & Fried, 1994; Zhong & Liljenquist, 2006). However, while moral cleansing has been experimentally observed in some instances (e.g., Gneezy, Imas, & Madarász, 2014), there were also several failures to replicate the phenomenon, indicating that the mechanism triggering the behaviour has not yet been fully understood (Dolan & Galizzi, 2015).

To investigate the mechanisms behind moral cleansing, we propose to distinguish between salient and inconspicuous misbehaviours. As saliency is likely to draw attention, it may act as a stimulus and trigger a reaction from the individual. For example, the participants in Milgram's (1963, 1965) experiments on obedience to authority, faced a very salient misbehaviour as they were instructed to punish another individual with potentially harmful and even fatal electric shocks. As a reaction, several participants verbally expressed their discomfort and even lied to the experimenter to avoid hurting others. However, if a misbehaviour is inconspicuous then individuals simply cannot react, due to the lack of stimuli. This second case is what the framework of bounded ethicality qualifies as a blind spot (Chugh et al., 2005), or numbness (Bazerman & Sezer, 2016). We therefore argue that one's moral identity serves as a reference point. If an individual believes they are highly moral and

---

[3] While part of the literature on moral cleansing focuses on symbolic physical manifestations such as washing one's hands with soap or using a cleaning wipe after recalling a misbehavior (e.g., Zhong & Liljenquist, 2006), I focus on behavioral cleansing in a broader sense as I include any form on compensation for pasts immoral behaviors.

misbehaves, the discrepancy between their identity beliefs and actual behaviour will be large and therefore salient. This salient difference between beliefs and reality can be thought of as a moral "burn" and will stimulate the need for cleansing behaviours. On the contrary, if the misbehaviour is inconspicuous, then this behaviour may be taken as a signal of one's identity (Akerlof & Kranton, 2000, 2005; Ariely & Norton, 2008; Bem, 1972). Consequently, it will serve as a new reference point and the moral identity of the individual will be shifted. In turn, this shift of reference point will increase the likelihood of repeating the past behaviour, although the latter will no longer be considered as such due to the identity being adapted. Therefore, we hypothesize that salient moral dissonance has a positive impact on the likelihood of future moral behaviours. This behavioural pattern – a misbehaviour triggering a better behaviour – is a behavioural expression of moral cleansing.

Hypothesis 1: A salient moral dissonance is more likely to lead to a subsequent moral behaviour than an inconspicuous moral dissonance.

Conversely, we hypothesize that a moral dissonance that is inconspicuous has a negative impact on the likelihood of future moral behaviours.

Hypothesis 2: An inconspicuous moral dissonance is more likely to lead to a subsequent misbehaviour than a salient moral dissonance.

To summarize, we posit that prior misbehaviours can lead to either moral cleansing or subsequent misbehaviours. We propose that the saliency of the prior misbehaviour is what drives the two opposing behavioural patterns. The role of saliency in triggering subsequent misbehaviors is summarized in Figure 1.

FIGURE 1. Theoretical Model for the Link between Prior and Subsequent Misbehaviours, Moderated by the Saliency of the Moral Dissonance
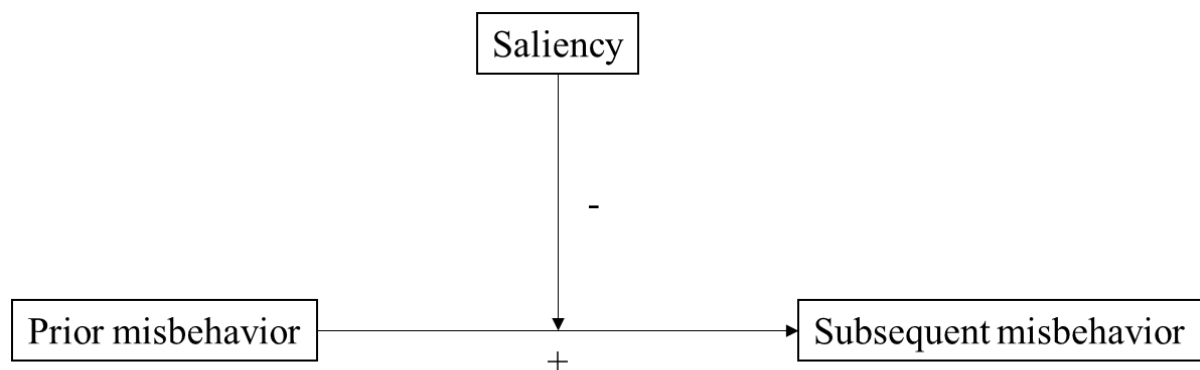
3. Experimental design and procedures

We recruited 615 participants on Amazon Mechanical Turk. Participants completed 30 rounds of the effort task described by Abeler, Falk, Goette, and Huffman (2011). Participants were paid a fixed compensation of 6 USD, for an average completion time of 50 minutes (7.12 USD per hour).

The design had three parts of 10 rounds. The task consisted in counting the number of ones in a table randomly filled with ones and zeros. In each round participants could choose between a large and a small table. For the participant a part of the experiment was completed once 10 tables had been correctly completed (irrespective of the type of table). Completing small tables therefore had the advantage that the participants could complete the experiment faster for the same payment. However, completing a table also had an externality on a charity: completing a large table increased the donation to a charity by 10 cents, whereas completing a small table removed 1 cent from the donation. In the first part of the experiment all participants faced the same conditions: small tables contained 6 lines, large tables contained 9 lines. In the second part, participants were randomly assigned to one of three experimental conditions. In the control condition (C) the size of small and large tables remained as in part 1. In the small

(S) and high (H) temptation treatments, the size of the small tables was reduced to 5 and 3 lines, respectively. In the third part the size of small tables was reverted to 6 lines for all participants. See Appendix A for the screenshots of the experimental tasks for each condition.

4. Results

3.1 General pattern

Figure 1 provides an overview of participants choices in all three parts. We test the statistical significance of observed effect with an OLS estimation in which we regress high effort (measured as the proportion of large tables) in each part of the experiment on indicator variables for treatments (Low, High), parts of the experiment (Part 2, Part 3) and the interaction effects of those variables (see column (1) of Table 1).

In part 1—which is the same for all participants—the proportion of large tables is very similar in all three conditions (C: 0.39, L: 0.37, H: 0.39, none of the differences are statistically significant). In the control question high effort drops insignificantly in part 2 (0.35, $p = 0.274$) and the remains roughly constant in part 3 (0.34).

In the low temptation condition there is an intermediate drop in the proportion of high effort when the participants face a slightly stronger temptation to pick the small tables in part 2 (0.30, $p = 0.038$). As expected, this effect is not reversed in part 3 when incentives return to the initial level (0.29). Although this pattern is observationally consistent with our prediction, a diff-in-diff test reveals that neither the decrease from part 1 to part 2 is not significantly larger than the pure time trend in the control condition ($p=0.283$).

In the high temptation condition we see a more pronounced drop in high effort in part 2 (0.23, $p < 0.001$). This effect is partially undone in part 3 when incentives are set back to their initial level (0.33, $p = 0.008$). The diff-in-diff analysis shows that decrease in part 2 of the high temptation conditions is significantly larger than those in the low temptation condition

($p=0.004$) and the control condition ($p=0.000$). Moreover, also the increase from part 2 to part 3 is significantly different from the developments in the other two conditions ($p < 0.001$ in both cases).

Observationally, the overall pattern of these results is consistent with our hypothesis. However, from a statistical point of view, only the effects in the high temptation condition reach significance. However, an F test for individual fixed effects indicated that individuals vary greatly in terms of their initial effort provision as well as their reactions to the treatments ($p < 0.001$). Therefore, we conducted further analyses to identify types and investigate different reactions to the magnitudes of temptation. The next section therefore uses finite mixture models to investigate whether the general pattern is a combination of interesting type-specific effects.
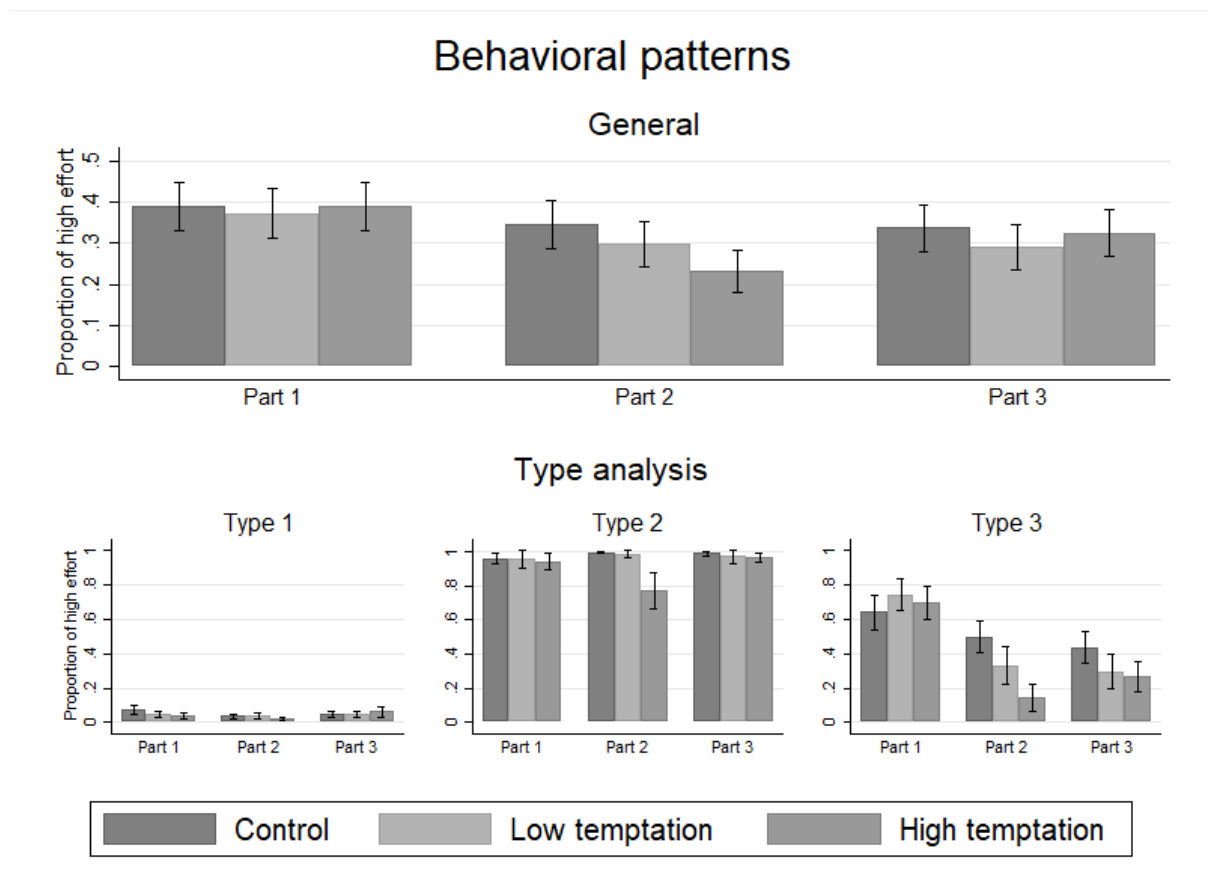

Fig. 1. Average proportion of high effort (large tables) counted by the participants across parts, per treatment. The graphic at the top is the overall pattern, while the three graphics at the bottom

are the patterns for each of the three identified types of individuals. The error bars represent plus/minus one standard deviation.

Table 1

Mean immediate and delayed treatment effects of the choices of high efforts for the whole sample, and the three types identified by the Finite Mixture Model.

	Pooled regressions		Type analysis		
	Model 1	Model 2	Type 1	Type 2	Type 3
Constant	0.41399 ***	0.30698 ***	0.094***	1.01***	0.637***
	(0.01237)	(0.02752)	(0.015)	(0.017)	(0.026)
Low*Part2	-0.02850	-0.02769	0.012	0.04	-0.242***
	(0.02092)	(0.02507)	(0.02)	(0.025)	(0.068)
High*Part2	-0.11353 ***	-0.10930 ***	0.011	-0.157**	-0.408***
	(0.02055)	(0.02479)	(0.02)	(0.057)	(0.057)
Low*Part3	-0.02700	-0.02716	0.024	0.034	-0.267***
	(0.02092)	(0.02507)	(0.025)	(0.033)	(0.068)
High*Part3	-0.01015	0.00025	0.062*	0.048†	-0.284***
	(0.02055)	(0.02479)	(0.028)	(0.028)	(0.067)
Low	-0.01900	-0.01024		0.004	
	(0.01479)	(0.01775)		(0.019)	
High	-0.00173	0.00582		-0.015	
	(0.01453)	(0.01753)		(0.019)	
Part2	-0.00510	-0.00395		-0.005	
	(0.01904)	(0.02267)		(0.019)	
Part3	0.02729	0.04734		0.027	
	(0.02818)	(0.03386)		(0.029)	
Round	-0.00409 ***	-0.00443 **		-0.004***	
	(0.00120)	(0.00145)		(0.016)	
Sex		-0.07974 ***			
		(0.00860)			
Age		0.02585 ***			
		(0.00197)			
N	18450	12450			
R^2	0.01188	0.03334			
Sigma				0.281	
				(0.007)	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.1$

3.2 Type analysis

We identify type-specific patterns using a finite mixture model (FMM) (see Bruhin, Fehr-Duda, and Epper 2010, for a similar approach). Assuming that types may differ in terms of initial moral behavior, as well as immediate, and delayed reactions to the treatments, the FMM cleanly identified three types (see Table 2).

Table 2

Proportions of Type 1, Type 2, and Type 3 individuals.

	Type 1	Type 2	Type 3
Proportion	0.573	0.229	0.197
SE	(0.020)	(0.017)	(0.016)

Type 1. Type 1 individuals display a low initial moral identity, as they rarely exert high effort in part 1 (0.055, averaged across conditions). Type 1 participants have no relevant reactions to the treatments, as they are already exerting the lowest possible amount of effort.

Type 2. Type 2 individuals initially provide almost only high efforts, displaying a really high moral identity (0.95, averaged across conditions). They only react to the high temptation treatment, decreasing their efforts significantly more than in the low temptation treatment ($p=0.002$) and in the control condition ($p=0.001$). Moreover, participants in the high temptation treatment increased their proportion of high efforts significantly more than participants in the low temptation treatment ($p=0.001$) and in the control condition ($p=0.001$). This final proportion in the high temptation treatment was not significantly different from the ones in the control treatment ($p=0.188$) nor the low temptation treatment ($p=0.798$). In

summary, Type 2 participants are only sensitive to high temptations, yet they cleanse fully from their deviation by recovering their initial moral behavior.

Type 3. Type 3 individuals display an intermediate moral identity. They provided a moderate initial level of high effort (0.69, averaged across conditions). Type 3 individuals respond to both the low ($p<0.001$) and the high temptations treatments ($p<0.001$) in part 2, both reactions are significantly different from the drop in the control condition ($p=0.006$ and $p<0.001$, *respectively*), but the reactions are not significantly different ($p=0.197$). In part 3, Type 3 participants in the low temptation treatment did not change efforts significantly differently from participants in the control condition ($p=0.783$), while participants in the high temptation treatment increased their proportion of high efforts significantly more than participants in the low temptation treatment ($p=0.047$) and in the control condition ($p=0.021$).However, the final proportion of high efforts in the high temptation treatment is significantly smaller than the one in the control condition ($p=0.005$), yet not significantly different from the one in the low temptation treatment ($p=0.650$). In consequences, while Type 3 individuals cleansed after facing a salient deviation, they failed to recover their initial morality.

The type-based analysis nicely explains our aggregated results. A challenge in our data set is that 57% of the participants do not react to either treatment, because their initial effort level is already so low that a further decrease is almost impossible (floor effect). On the other extreme, another 23% of our participants exhibit a strong moral identity and therefore only deviate from their initial behavior if the temptation is high. The effects of the low temptation treatment are therefore exclusively driven by the remaining fraction of participants (20%) who respond to both our treatments—largely in the expected way.

5. Discussion and conclusions

The present experimental evidence supports the assumption that salient moral deviations increase the likelihood of moral cleansing, while inconspicuous deviations lead to subsequent misbehaviors. However, this only holds for individuals with a moderate initial moral behavior, as they are sensitive to both high and low temptations. These results emphasize the importance of measuring initial morality in the study of morally inconsistent behavioral patterns such as moral cleansing. We therefore expect future research to further investigate the various degrees of moral cleansing, depending on the magnitude of deviations.

References

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review, 101*(2), 470-492.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics, 115*(3), 715-753.

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *The Journal of Economic Perspectives, 19*(1), 9-32.

Ariely, D., & Norton, M. I. (2008). How actions create–not just reveal–preferences. *Trends in Cognitive Sciences, 12*(1), 13-16.

Bandura, A. (1990). Mechanisms of moral disengagement in terrorism. *Origins of terrorism: Psychologies, ideologies, states of mind, 28*(7), 161-191.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*(3), 193-209.

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education, 31*(2), 101-119.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71*(2), 364.

Bazerman, M. H., & Gino, F. (2012). Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science, 8*, 85-104.

Bazerman, M. H., & Sezer, O. (2016). Bounded awareness: Implications for ethical decision making. *Organizational Behavior and Human Decision Processes, 136*, 95-105.

Bazerman, M. H., & Tenbrunsel, A. E. (2011). *Blind spots: Why we fail to do what's right and what to do about it*: Princeton University Press.

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology, 6*, 1-62.

Bruhin, A., Fehr-Duda, H., & Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica, 78*(4), 1375-1412.

Chapman, H. A., & Anderson, A. K. (2013). Things rank and gross in nature: a review and synthesis of moral disgust. *Psychological Bulletin, 139*(2), 300.

Chugh, D., & Bazerman, M. H. (2007). Bounded awareness: What you fail to see can hurt you. *Mind & Society, 6*(1), 1-18.

Chugh, D., Bazerman, M. H., & Banaji, M. R. (2005). Bounded ethicality as a psychological barrier to recognizing conflicts of interest. *Conflicts of interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy*, 74-95.

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology, 47*, 1-16.

Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of Experimental Social Psychology, 45*(3), 590-593.

Fleischhut, N., & Gigerenzer, G. (2013). Can simple heuristics explain moral inconsistencies?. In R. Hertwig, U. Hoffrage & the ABC Research Group, *Simple heuristics in a social world* (pp. 459-486). New York, NY: Oxford University Press.

Gneezy, U., Imas, A., & Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science, 60*(11), 2645-2658.

Lee, S. W., & Schwarz, N. (2021). Grounded procedures: A proximate mechanism for the psychology of cleansing and other physical actions. *Behavioral and Brain Sciences, 44*.

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass, 4*(5), 344-357.

Messick, D. M., & Bazerman, M. H. (1996). Ethical leadership and the psychology of decision making. *MIT Sloan Management Review, 37*(2), 9.

Moore, D. A., Tetlock, P. E., Tanlu, L., & Bazerman, M. H. (2006). Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *Academy of Management Review, 31*(1), 10-29.

Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371.

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations, 18*(1), 57-76.

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science, 20*(4), 523-528.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99-118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129.

Stone, J., Aronson, E., Crain, A. L., Winslow, M. P., & Fried, C. B. (1994). Inducing hypocrisy as a means of encouraging young adults to use condoms. *Personality and Social Psychology Bulletin, 20*(1), 116-128.

Tenbrunsel, A. E. (1998). Misrepresentation and expectations of misrepresentation in an ethical dilemma: The role of incentives and temptation. *Academy of Management Journal, 41*(3), 330-339.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*(5), 853.

Zhong, C.-B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science, 313*(5792), 1451-1452.

Appendix

Appendix A

Appendix A-1. Choice between the small and large tables as displayed on the participants' screen in the first and third parts.

Part 1/3, decision 1/10

Choose a table.

Table A: decrease of the donation by $0.01														
1	1	0	1	0	0	1	1	0	1	0	1	0	1	1
0	1	0	1	0	1	0	1	1	1	1	1	0	1	0
0	1	0	0	1	1	0	0	0	1	1	0	1	1	1
0	0	0	1	1	0	0	1	0	1	0	1	1	1	1
1	0	0	0	1	1	0	1	1	0	0	1	1	1	1
1	0	0	0	0	1	1	1	1	0	0	0	1	0	1

Table B: increase of the donation by $0.1														
1	1	0	1	0	0	1	1	0	1	0	1	0	1	1
0	1	0	1	0	1	0	1	1	1	1	1	0	1	0
0	1	0	0	1	1	0	0	0	1	1	0	1	1	1
0	0	0	1	1	0	0	1	0	1	0	1	1	1	1
1	0	0	0	1	1	0	1	1	0	0	1	1	1	1
1	0	0	0	0	1	1	1	1	0	0	0	1	0	1
1	1	0	0	1	1	1	1	1	0	1	0	0	0	0
1	0	1	0	1	0	0	0	1	0	1	1	0	1	1
1	1	1	1	0	0	1	1	1	0	1	1	0	1	1

How many ones are in the table you chose?

Next

Appendix A-2. Choice between the small and large tables as displayed on the participants' screen in the second part (control condition).

Part 2/3, decision 1/10

Choose a table.

Table A: decrease of the donation by $0.01														
1	0	1	0	0	0	0	0	1	0	0	1	1	1	0
0	1	1	0	0	1	1	1	0	1	1	0	1	1	1
0	1	1	1	1	1	1	1	0	1	0	1	1	0	1
1	0	0	0	1	0	1	1	1	0	1	0	1	1	0
1	1	0	1	1	1	0	0	1	0	0	1	1	1	0
1	0	0	1	0	1	0	1	0	1	1	0	1	0	0

Table B: increase of the donation by $0.1														
1	0	1	0	0	0	0	0	1	0	0	1	1	1	0
0	1	1	0	0	1	1	1	0	1	1	0	1	1	1
0	1	1	1	1	1	1	1	0	1	0	1	1	0	1
1	0	0	0	1	0	1	1	1	0	1	0	1	1	0
1	1	0	1	1	1	0	0	1	0	0	1	1	1	0
1	0	0	1	0	1	0	1	0	1	1	0	1	0	0
1	1	0	1	1	0	1	0	0	1	0	1	1	1	1
0	0	0	0	0	1	0	0	0	1	1	0	0	0	1
0	1	1	0	0	0	1	0	1	1	0	1	0	1	1

How many ones are in the table you chose?

Next

Appendix A-3. Choice between the small and large tables as displayed on the participants' screen in the second part (low temptation condition).

Choose a table.

Table A: decrease of the donation by $0.01

1	0	1	0	0	0	0	0	1	0	0	1	1	1	0
0	1	1	0	0	1	1	1	0	1	1	0	1	1	1
0	1	1	1	1	1	1	1	0	1	0	1	1	0	1
1	0	0	0	1	0	1	1	1	0	1	0	1	1	0
1	1	0	1	1	1	0	0	1	0	0	1	1	1	0

Table B: increase of the donation by $0.1

1	0	1	0	0	0	0	0	1	0	0	1	1	1	0
0	1	1	0	0	1	1	1	0	1	1	0	1	1	1
0	1	1	1	1	1	1	1	0	1	0	1	1	0	1
1	0	0	0	1	0	1	1	1	0	1	0	1	1	0
1	1	0	1	1	1	0	0	1	0	0	1	1	1	0
1	0	0	1	0	1	0	1	0	1	1	0	1	0	0
1	1	0	1	1	0	1	0	0	1	0	1	1	1	1
0	0	0	0	0	1	0	0	0	1	1	0	0	0	1
0	1	1	0	0	0	1	0	1	1	0	1	0	1	1

How many ones are in the table you chose?

Next

Appendix A-4. Choice between the small and large tables as displayed on the participants' screen in the second part (high temptation condition).

Choose a table.

Table A: decrease of the donation by $0.01

1	0	1	0	0	0	0	0	1	0	0	1	1	1	0
0	1	1	0	0	1	1	1	0	1	1	0	1	1	1
0	1	1	1	1	1	1	1	0	1	0	1	1	0	1

Table B: increase of the donation by $0.1

| 1	0	1	0	0	0	0	0	1	0	0	1	1	1	0
0	1	1	0	0	1	1	1	0	1	1	0	1	1	1
0	1	1	1	1	1	1	1	0	1	0	1	1	0	1
1	0	0	0	1	0	1	1	1	0	1	0	1	1	0
1	1	0	1	1	1	0	0	1	0	0	1	1	1	0
1	0	0	1	0	1	0	1	0	1	1	0	1	0	0
1	1	0	1	1	0	1	0	0	1	0	1	1	1	1
0	0	0	0	0	1	0	0	0	1	1	0	0	0	1
0	1	1	0	0	0	1	0	1	1	0	1	0	1	1

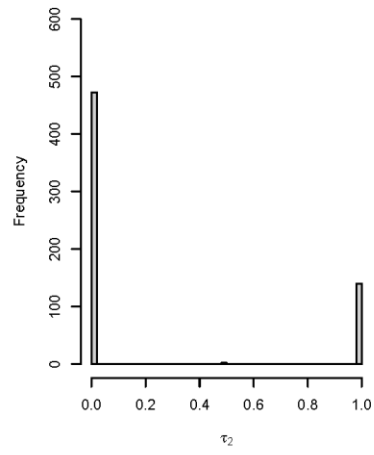How many ones are in the table you chose?

Next

Appendix B

Distributions of the posterior probabilities of belonging to types, based on the results of the

Finite Mixture Models (K = 3)

Pr. of being in Type 1
(K=3 Model)
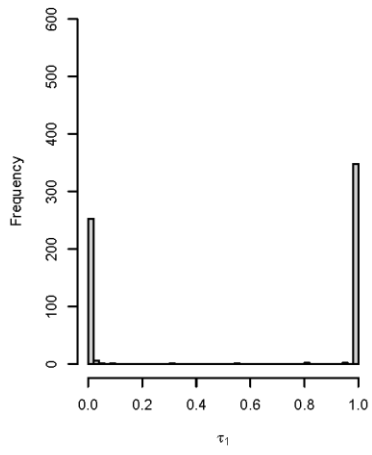
Pr. of being in Type 2
(K=3 Model)

Pr. of being in Type 3
(K=3 Model)