



**UNIL** | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

---

Year : 2021

## THREE ESSAYS IN DECISION MAKING

Légeret Matthieu

Légeret Matthieu, 2021, THREE ESSAYS IN DECISION MAKING

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB\_7499042543705

### **Droits d'auteur**

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

### **Copyright**

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

---

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES  
DÉPARTEMENT DE COMPORTEMENT ORGANISATIONNEL

**THREE ESSAYS IN DECISION MAKING**

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales  
de l'Université de Lausanne

pour l'obtention du grade de  
Docteur ès Sciences Économiques, mention « Management »

par

Matthieu LÉGERET

Directeur de thèse  
Prof. Ulrich Hoffrage

Co-directeur de thèse  
Prof. Christian Zehnder

Jury

Prof. Felicitas Morhart, Présidente  
Prof. Patrick Haack, expert interne  
Prof. Ralph Hertwig, expert externe

LAUSANNE  
2021





UNIL | Université de Lausanne

---

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES  
DÉPARTEMENT DE COMPORTEMENT ORGANISATIONNEL

**THREE ESSAYS IN DECISION MAKING**

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales  
de l'Université de Lausanne

pour l'obtention du grade de  
Docteur ès Sciences Économiques, mention « Management »

par

Matthieu LÉGERET

Directeur de thèse  
Prof. Ulrich Hoffrage

Co-directeur de thèse  
Prof. Christian Zehnder

Jury

Prof. Felicitas Morhart, Présidente  
Prof. Patrick Haack, expert interne  
Prof. Ralph Hertwig, expert externe

LAUSANNE  
2021

## IMPRIMATUR

---

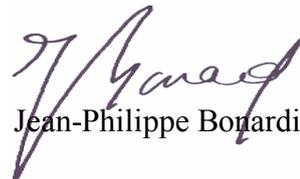
Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Études Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur Matthieu LÉGERET, titulaire d'un bachelor et d'un master en Management de l'Université de Lausanne, en vue de l'obtention du grade de docteur ès Sciences économiques, mention management.

La thèse est intitulée :

### THREE ESSAYS IN DECISION MAKING

Lausanne, le 19 janvier 2021

Le doyen



Jean-Philippe Bonardi



## **Members of the thesis committee**

Prof. Ulrich Hoffrage

Professor, University of Lausanne

Thesis supervisor

Prof. Christian Zehnder

Professor, University of Lausanne

Thesis supervisor

Prof. Patrick Haack

Professor, University of Lausanne

Internal member of the thesis committee

Prof. Ralph Hertwig

Director, Center for Adaptive Rationality (Max Planck Institute for Human Development)

External member of the thesis committee



University of Lausanne  
Faculty of Business and Economics

Ph.D. in Economics  
Subject area "Management"

I hereby certify that I have examined the doctoral thesis of

**Matthieu LÉGERET**

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members  
made during the doctoral colloquium  
have been addressed to my entire satisfaction.

Signature:  Date: January 18<sup>th</sup>, 2021

Prof. Ulrich HOFFRAGE  
Thesis supervisor



University of Lausanne  
Faculty of Business and Economics

Ph.D. in Economics  
Subject area "Management"

I hereby certify that I have examined the doctoral thesis of

**Matthieu LÉGERET**

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members  
made during the doctoral colloquium  
have been addressed to my entire satisfaction.

Signature: \_\_\_\_\_  \_\_\_\_\_ Date: 17.1.2021

Prof. Christian ZEHNDER  
Co-supervisor



University of Lausanne  
Faculty of Business and Economics

Ph.D. in Economics  
Subject area "Management"

I hereby certify that I have examined the doctoral thesis of

**Matthieu LÉGERET**

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members  
made during the doctoral colloquium  
have been addressed to my entire satisfaction.

Signature:



Date: \_\_\_\_17.01.2021\_\_\_\_

Prof. Patrick HAACK  
Internal member of the doctoral committee



University of Lausanne  
Faculty of Business and Economics

Ph.D. in Economics  
Subject area "Management"

I hereby certify that I have examined the doctoral thesis of

**Matthieu LÉGERET**

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members  
made during the doctoral colloquium  
have been addressed to my entire satisfaction.

Signature:



Date: 18.01.2021

Prof. Ralph HERTWIG  
External member of the doctoral committee



## Acknowledgments

As I reached the end of the second year of my Bachelor's studies, I came to realize that I was interested in studying human behaviors. Of course, had I known back then how the journey of my PhD would unfold, I would have thought twice. I have often heard some Professors or even former PhD students claim that the PhD is a process through which you go alone. I do not think that these individuals could be more wrong. These are the same who talk about the PhD as a reward at the end of the adventure. My journey was only made possible because of the people who supported and inspired me. The real reward for me is the friends I have made in these years. They should probably receive part of the title that comes with this thesis.

First and foremost, I would like to thank both of my supervisors, **Christian Zehnder** and **Ulrich Hoffrage**. Although differently, both of them have found ways to motivate me to better myself. I would also like to thank **Patrick Haack** and **Ralph Hertwig**, my internal and external committee members, for the feedback they provided on my thesis chapters. I was truly impressed by the depth of their comments and will strive to improve my work thanks to them.

I would also like to thank my colleagues and friends, who made my everyday a little brighter. As I am trying to keep these acknowledgments short, I would still like to write a few words for each of them (in alphabetical order, with their names in bold so they can find themselves in this long paragraph). Of course, I apologize if I forgot anyone. As I have verbally expressed my gratitude over the past months, my dearest friends know that I will always be thankful. Should the reader be offended by not being mentioned in this page, this reaction probably means they are part of the people who should not be here anyway.

**Afrodite**, thank you for listening to me and supporting me. I have really enjoyed our discussions. I guess I have always felt somewhat like Susie but was never able to defeat Helena. **Aurélien**, thank you for the great discussions we have had about academia, and life in general. I hope we will get the chance to see each other more often, in Toulouse or Lausanne. **Benjamin**, thank you for this great chapter we have co-authored. But also thank you for all the laughs we have had, and the ramen we have eaten together pretending we were meeting for work. **Benoit**, I have always looked up to you. Thank you for everything you have taught me, from beating Psycho Mantis by switching the controller port on GameCube, to making friends. Thank you for putting up with me, even when I was at my lowest. **Christian**, thank you for inspiring me not only to be a better scholar, but also a better person. Our conversations have been more insightful than any class I have taken, and none of this would have been possible without your support. **Cvetomir**, thank you for being such a good friend. Your dedication and hard work have been an inspiration to me, and so are our friendship and the beers we have drunk while sitting on random benches in Lausanne. **Fernando**, thank you for being there for me. Also, thank you for the awesome brunches and dinners we have had, and the amazing food you have cooked. **Irina**, thank you for your support through these difficult years. I wish you all the best in New York. **Justin**, thank you for the great time we have had in the office (pre-pandemic), but also for the discussions about experimental designs. Moving to your office has been a real breath of fresh air. **Konstantinos**, thank you for the great moments we have had, in Switzerland and in Greece. And you were right: I made it, but I am all out of patience and saliva. **Lisa**, thank you for the fun times at the winter and summer schools, and for constantly reminding me that we have a faculty council meeting. **Lemmy**, we miss you. **Moritz**, thank you for sharing

my passion for craft beers. Thank you for the amazing time we have had at conferences and thank you for helping me being less awkward in these social contexts. But mostly, thank you for your support when sh\*t really hit the fan. **Morten**, thank you for the fun lunches and coffee breaks we have had, and the bars we have raided. You are part of the people who made my daily life lighter. Takk og skål. **Nada**, thank you for being an awesome office mate. It is always a pleasure to talk to you and I hope we will get the chance to talk more often in the future. **Paola**, thank you for all those fun moments and your support. Eventually, we made it. **Robb**, thank you for being in my ears every day, making my routines a bit more tolerable. **Sirio**, like Harry Potter's uncle but with an "o". You and I have started this journey together, and (almost) finished it together. You have been following my adventures all along, practically like a tv show. Thank you for being an awesome colleague, and an even better friend. **Thomas**, thank you for our stimulating discussions, and your vibrant energy. Your good mood was often contagious and brightened my days in our office. **Tristan**, thank you for being so easy-going and fun to be around. Colleagues and friends like you have been dearly missed since you left. **Tyler**, thank you for your positive attitude and energy on campus, but also for the beers and poker nights we have had. These were always a great way to release some steam, and some money. **Vaibhav**, thank you for these fun moments. Your wit and critical thinking were a delight on every occasion. I would also like to thank some awesome colleagues, with whom I have not had the chance to spend enough time. I hope we can change that in the future, **Ron**, **Shreekanth**, and **Valentina**.

I would also like to thank some of our professors. I have really enjoyed interacting with **John Antonakis**, **Jean-Philippe Bonardi**, **Adrian Bruhin**, **Charles Efferson**, and **Marianne Schmid Mast**, either through regular conversations or more formal workshops. I would also like to thank **Holger Herz**, for our insightful discussions at the winter and summer schools. Finally, I would like to thank some of the people in this faculty whose great work and attitude have made my life easier. I would like to thank **Anina Eggenberger** and **Stéphane Billeter** for their great work at the Labex. I would like to thank **Corinne Delapierre**, **Michèle Claude**, and **Bénédicte Moreira**<sup>1</sup>, without whom part of the faculty would probably collapse. I would like to thank **Marie-Elisabeth Delille** for her help, and for being such a positive person in general.

To all of these people, I dedicate my first and second chapters. I probably owe each of them a beer.

Matthieu

---

<sup>1</sup> Bénédicte, I am very sorry for all these deadlines I have missed.



## **How good can bad leaders be? The opportunity costs of leader selection**

### Abstract

Although a large body of research investigates what characterizes good leaders and how to best select them, internal promotions in organizations often yield underperforming leaders. The leadership literature provides several explanations for why “bad” leaders exist. However, all this work builds on the premise that the choice of an inefficient leader is a failure of the selection process. In this paper, we take a different stance and suggest that—in some cases—organizations might select underperforming leaders because it is efficient to do so. In essence, we argue that selecting a leader within a fixed group of individuals is similar to allocating any other limited resource. Leader selection has an opportunity cost because the leader is no longer available as a follower. We identify cases in which it is optimal not to select the most competent individual as leader. Finally, we propose to use tournament theory to efficiently select leaders and discuss how to set incentives within a group to identify competencies and to select the most appropriate leader accounting for the context in which the group operates.

Keywords: Leader selection, followership, opportunity costs, comparative advantage

*Matthieu Légeret*

*University of Lausanne / Switzerland*

*matthieu.legeret@unil.ch*

*Benjamin Tur*

*University of Lausanne / Switzerland*

*benjamin.tur@unil.ch*

## I. Introduction

Research on leader emergence has mainly been driven by the aspiration to identify good leaders, and, most importantly, decode the behaviours and traits associated with outperforming leaders (Day, Fleenor, Atwater, Sturm, & McKee, 2014). As a consequence, most scholars have focused on the traits and the behaviours, such as personality, intelligence, transformational leadership style that explain why certain individuals emerge as leaders or why some leaders are better than others (e.g., Antonakis, House, & Simonton, 2017; Arvey, Rotundo, Johnson, Zhang, & McGue, 2006; Judge & Piccolo, 2004). Some leadership scholars have even focused on the evolutionary origins of contemporary leadership, such as why tall men are preferred over short men for leadership positions (Van Vugt & Ahuja, 2010). Others – a minority – have chosen a different approach and focus on what characterizes bad leaders, articulating theories of destructive leadership (Einarsen, Aasland, & Skogstad, 2007) or laissez-faire leadership (Judge & Piccolo, 2004). Overall, leadership scholars have provided an abundance of evidence to inform leadership selection and training.

However, most of the previous research has been operating under the implicit assumption that a group always benefits from selecting the best leader. For instance, Schyns and Schilling (2013:138) noted that “research into leadership has often been guided by the quest to find the most effective person or method to lead”. One consequence of this assumption is that the presence of underperforming leaders at key positions is a failure of the selection process as it reduces the fitness of the group. In this paper we present a theoretical approach to leader selection challenging the common assumption that selecting underperforming leaders is a mistake. Building on the logic of comparative advantage (Ricardo, 1817), we explain why selecting underperforming leaders may be beneficial for an organization. We argue that the process of leader selection should consider both leadership and followership abilities, because the appointment of a leader within a group implies that the leader at least reduces follower

activities. More specifically, in situations where the best leaders also happen to be the best followers, it may be optimal for the group to select an underperforming leader in order to avoid losing the best follower.

The remainder of the paper is structured as follows: The next section reviews the existing literature on bad leaders. We then outline the basic features of our leadership model, before extending it to different organizational contexts and discussing how characteristics of the environment affect optimal leader selection. Finally, we illustrate how leader selection processes should be designed to maximize group efficiency, before concluding and discussing the implications of the proposed theory for research and practice.

## II. Leader selection

### Good and bad leaders

A large part of the leadership literature has focused on identifying the traits and behaviours that make a leader efficient. For instance, intelligence (Judge, Colbert, & Ilies, 2004) is a universally endorsed leadership trait (Den Hartog, House, Hanges, Ruiz-Quintanilla, & Dorfman, 1999), relevant for predicting leadership emergence nonetheless it has been shown to have a curvilinear effect on performance and follower satisfaction (Antonakis et al., 2017). Likewise, leadership styles such as transformational (Bass, 1999; Bass & Avolio, 1994) or charismatic leaderships (House & Howell, 1992; Howell & Avolio, 1992). are typically found to be beneficial for the followers and their organization. Although essential, this stream of research has a rather narrow focus in that it only investigates the positive side of the coin. In fact, there is a tendency in the leadership literature to distance leaders from negative behaviours and outcomes. For instance, bad leadership is typically explained through an individual's negative traits, such as arrogance and ambition (McCall and Lombardo, 1983), being authoritarian and narcissistic (House & Howell, 1992; Malcolm Higgs, 2009), unethical

(Howell & Avolio, 1992), or poor practices and a lack of skills (Conger, 1990). While research on bad leaders is similar to the typical research on the positive aspects of leadership, these negative traits and behaviours are often excluded from the definitions of leadership (Schyns & Schilling, 2013). As a consequence, leadership research has distanced itself from bad leadership by focusing on positive aspects and defining leadership through positive behaviours or outcomes (see also Yukl & van Fleet, 1992). Moreover, what is categorized as “bad leadership” encompasses many different behaviours, ranging from destructive (e.g., Einarsen, Aasland, & Skogstad, 2007) to incompetent (see Kellerman, 2004) leadership. Although both destructive and incompetent leaders are inefficient, they can have very different impacts on their followers. More concerning, while research focused on identifying the competencies that make great leaders, several scholars pointed out that there is a mismatch between what research is saying and what is observed in practice. According to a survey conducted by Gallup, 69% of a sample of 1’000 German employees reported having at least one bad manager in their career (*State of the Global Workplace*, 2017). This is further corroborated by Artz, Goodall, and Oswald (2020) who found that about 13% of employees in Europe report having a “bad boss”. While this only reflects the appreciation of employees and does not directly constitute a valid measure of the effectiveness of these leaders, it points to the fact that leaders might not always perform at the expected level and raises the question of why organizations would select underperforming leaders. Yet, these findings are in line with earlier results from Erickson, Shaw, and Agabe (2007) who conducted a survey about the prevalence of bad leadership. Out of 335 respondents, only 24.8% reported having a good leader, 37.6% reported having an average leader, and 37.4% described their leader as bad. In addition, the reason why they reported having a bad leader has little to do with a lack of positive traits nor the presence of negative ones as research suggests, but rather with how efficient they are: respondents reported that their leader was “unable to deal effectively with subordinates” (p. 35). Finally, the

leadership literature describes bad leaders as a mistake, which in principle should not be sustainable. However, the respondents frequently reported that their bad leaders were either rewarded or even promoted.

Several theories have been proposed to explain the emergence of underperforming leaders. In particular, one mechanism suggests that individuals use implicit theories to make their decisions (Foti, Fraser, & Lord, 1982; Lord, Foti, & De Vader, 1984). Individuals share a common set of beliefs or stereotypes about what traits and behaviours leaders should have in a given situation and select the individuals that display those traits and behaviours. For example, in times of instability, such as war, individuals exhibiting masculine traits are preferred to those with more feminine traits (Little, Burriss, Jones, & Roberts, 2007). Although inferential processes assume that individuals perceive some traits and behaviours as prototypical of good leadership, these traits and behaviours often fall short of predicting leader efficiency. In fact, there is no empirical evidence supporting the assumption that selecting a masculine leader instead of a feminine leader in times of war is associated with higher chances of success. Likewise, cues such as attractiveness (Antonakis & Dalgas, 2009), and height or age (Elgar, 2016) have been found to predict leader emergence, but add little for leader effectiveness. Therefore, implicit theories may fall short in selecting efficient leaders because of a mismatch between our evolved leadership psychology and the contemporary practice of leaders, as societies evolved much faster than genes (Wilson, Van Vugt, & O'Gorman, 2008).

Other scholars, like Zehnder, Herz, and Bonardi (2017), have suggested that bad leaders might emerge because of a mismatch between the individuals who have an incentive to lead and those who are good at leading. Considering that leadership is a public good game where leading is costly (e.g. it involves additional risk, time investment, etc.), the authors suggest that bad leaders could emerge because “*personality characteristics that have been identified with good leadership (vision, charisma, confidence, etc.) are not necessarily identical with those*

*that induce people to step up (such as need for power or career concerns)*” (p. 19). This view therefore explains the emergence of bad leaders through a mismatch between the competencies required to be an efficient leader and the motivations individuals have to take on leadership positions.

Overall, these theories of leader emergence have built a comprehensive framework to understand how inefficient leaders might be selected. Yet, most scholarship is built around the key assumption that selecting underperforming leaders is a mistake. While it is likely that poor leadership is sometimes the result of failures in the selection process, we would like to offer a complementary explanation. In particular, we propose to look at leader selection through the lenses of role allocation. Role allocation suggests that different roles are allocated to different individuals to achieve a common goal. Framing leading and following as roles, it appears that when an individual takes over the role of leader, she abandons her role of follower. We therefore argue that leader selection should not only be driven by the evaluation of leadership traits and competencies, but also by the evaluation of followership competencies that will be lost when one individual is selected to act as leader. By framing leader selection as a role allocation, we investigate the opportunity costs a group faces when it selects an individual for a leadership position.

### The opportunity cost of leader selection

Leaders are often defined through their competencies (e.g., Day et al., 2014; Luthans & Avolio, 2003; Shamir, House, & Arthur, 1993), or through the perception of their followers (e.g., Bligh, Kohles, & Meindl, 2004; Meindl, 1995; Meindl, Ehrlich, & Dukerich, 1985). “Leadership” can therefore be defined through its relational aspect, that is, the “process that occurs between a leader and a follower, groups of followers, or institutions” (Antonakis & Day, 2017, p.5). At the core of this definition resides the essence of what constitutes leadership: the interaction between a leader and her followers (Lord & Brown, 2001). In order to model this

relationship, we consider a group composed of  $n$  individuals. Naturally, individuals can vary in terms of competencies, both in leadership and in followership. For example, an individual might be excellent at crafting and sharing a vision but perform poorly when it comes to following orders, while still being better than other followers in one or both of these dimensions. For simplifications purposes, we define the competencies of an individual  $i$  by how much output this individual can produce. Individuals either follow ( $f_i$ ) or lead ( $l_i$ ). The output of the group may then be thought of as an aggregated measure of performance, where the leader fosters the productivity of the followers. The production function of the group can be represented by the sum of the performance of the followers (noted  $j$ ), enhanced by the competencies of their leader (noted  $i$ ):

$$y = l_i \sum_{j \neq i}^n f_j$$

Our definition of group production answers the call for a better integration of followership in the leadership literature (Bligh, 2011; Uhl-Bien, Riggio, Lowe, & Carsten, 2014) by framing group production as an interaction between the outputs of the leader and the output of followers. Our model relies on several assumptions. First, we focus on situations in which there is a necessity for a leader, so that one member must fulfil the leadership position. Second, we assume that leading and following are mutually exclusive roles: the leader cannot perform followers' duties at the same time as she fulfills her leadership duties. However, this assumption is made for simplicity only and will later be relaxed when we discuss the possibility for part-time leading. Finally, we presume that the performance of the leader and her followers is perfectly observed, and that competencies are fully realized. This assumption allows for a clear identification of individual outputs, which will prove useful for selecting the best leader for a given group. We will later also relax this assumption in order to discuss how to select leaders when competencies are hard to observe.

Although traditional leadership research would suggest only looking at the leadership competencies for leader selection, allocating mutually exclusive roles to individuals bears the same opportunity costs as any other limited resource allocation: it is crucial to consider both gained competencies as well as those forgone. To illustrate this point, we can imagine a team of two individuals. One of the individuals has to be selected as a leader, while the other one will be a follower. In order to make an informed decision, we need to consider their respective competencies in both roles: while individual 1 has excellent skills as a follower and as a leader, individual 2 is generally an average performer. These competences are summarized in Table 1.

-----  
Insert Table 1 about here  
-----

In this example, although individual 1 is the most competent leader, her higher competencies are not sufficient to cover the loss of her performance as a follower. In other words, in order to select the best leader for a given group, it is not enough to only evaluate leadership competencies. The optimal allocation of roles therefore depends on the relative advantage of one individual over the other. Indeed, assigning individual 1 to the leadership or followership position does not only depend on her own skills, but also on the skills of individual B and, therefore, how well the group as a whole would perform under the two different compositions<sup>1</sup>. Although employees are often referred to as “human resources”, to the best of

---

<sup>1</sup> This is similar to the notion of comparative advantage (Ricardo, 1817). While initially written to explain why it can be efficiency enhancing if countries specialize in a free-trade world even if some countries are systematically less productive than others. One of the main reasons why the theory of comparative advantage holds is because countries have finite resources. When a country invests resources to produce one type of good it cannot reuse the same resources to produce another type of good. Countries should therefore specialize in products for which they have a comparative advantage. By drawing on Ricardo’s model, we

our knowledge leader selection has only been studied without considering the opportunity costs of removing an individual from his responsibilities as a follower. In this paper, we depart from the common view on leader selection, which argues that the best leader should be selected, and propose a different perspective emphasizing the opportunity costs of becoming a leader (i.e., not being a follower). We therefore propose to shift the perspective of selection from the individual to the group. However, the trade-off between appointing a leader and losing a follower relies on three elements. First, we discuss the importance of the correlation between leadership and followership competencies. Then, we consider the impact of two contextual elements on leader selection and follower performance: importance and time-intensity of leadership.

Are good followers good leaders?

Although it is commonly assumed that followership performance is a good predictor of leadership performance, it is not always true. In fact, it is not unusual for the two roles to require very different sets of skills. For instance, a production plant may use machines as followers, as they can screw parts every 2 seconds, while a human leader is unable to achieve the same performance as the machines but is able to monitor them. In that case, the correlation between the competencies that make a good follower or an efficient leader is null at best, and may even be negative. In other words, being good at following is not predictive of being good at leadership. In fact, the machine will always perform followership tasks while the human will always perform leadership tasks, even when accounting for opportunity costs. Nevertheless, there are less extreme examples. In hospitals, the qualities characterizing excellent doctors are not necessarily the same qualities associated with being a good manager. The correlation between followership and leadership skills in this context will be low. Similarly, top rank

---

argue that organizations too have finite resources and should therefore exploit the logic behind the theory of comparative advantages.

military officers have little in common with field troopers. Ultimately, when the best followers are not the best leaders, opportunity costs become irrelevant as an organization would not select an excellent follower as the leader. Understanding how followership and leadership skills tend to be correlated in a specific context is highly important to assess whether opportunity costs are important to optimize the performance of the group.

Of course, if the correlation between leadership and followership competencies is low, so that the best leader happens to be the worst follower, then the opportunity costs will be irrelevant. In this specific case, selecting the most competent leader is optimal for the group, as selecting any other leader would increase the opportunity costs as their followership competencies would be at least partially lost. However, if leadership and followership competencies are positively correlated, then the opportunity costs become relevant and the group may face a trade-off between the competencies of their leader and the overall production of the group.

#### Contextual importance the leader and part-time leadership

While leaders are commonly acknowledged to be generally important, there are contexts in which they either have less influence on a groups' performance (Weierter, 1997) or must perform different tasks than the ones inherent to their leadership position. In some instances, the leader will have only a marginal effect on group output while in other instances her role may be critical for the group's output. This last point becomes salient if we consider, for example, the role of a boat captain. When the boat is anchored in the harbour, the importance of the leader on her group is relatively low. Most sailors go about their duties independently, performing simple maintenance and supply tasks. On the contrary, when the boat is at sea going through a storm, the importance of the leader for the group is high. While in many situations the captain's competencies are crucial to coordinate the sailors, under some

circumstances the captain's role is not very different from the sailors'. In our model, the contextual importance of the leader ( $\alpha$ ) therefore acts as a moderator of the leader's influence on her followers' performance. However, if in a given context leadership does not matter at all ( $\alpha = 0$ ), the production of the group will simply be the performance of the followers. While the leader acts as a catalyst of the followers' production, we allow her role to vary between not being needed and infinitely enhancing her followers' production ( $0 \leq \alpha < \infty$ ). In addition, we extend the model to allow the leader to spend a proportion ( $1 - \beta$ ) of her time leading, while she spends the rest ( $\beta$ ) performing as a follower. Consequently, the performance of the followers also contains the performance of the leader as a follower for the proportion of her time she does not spend leading ( $\beta$ ). As an individual is constrained to a fixed amount of time, the proportion of time she spends leading cannot exceed 100% nor be below 0% ( $0 \leq \beta \leq 1$ ). The production function of a group where individual  $i$  is leading can, thus, be expressed as follows (see Appendix for the complete steps):

$$y = [1 + \alpha l_i] \left( \beta f_i + \sum_{j \neq i}^n f_j \right)$$

The contextual importance of leadership essentially increases the benefits of selecting a skilled leader: the more important the leadership, the highest the competencies of the leader should be. Conversely, the time required to be spent on followership tasks has the opposite effect, as it increases the costs of losing a follower, and therefore reduces the relevance of leadership skills. To illustrate the main features of the model, we consider four extreme cases for the contextual importance of leadership and the time spent leading.

**Unnecessary, full-time leader ( $\alpha = \beta = 0$ ).** Although a leader is required, her competencies are not used for the production in this context. The production function is impacted as follows:

$$y = \sum_{j \neq i}^n f_j$$

Therefore, appointing a leader only removes a follower from the team, as she will spend all of her time on useless, yet mandatory activities. In conclusion, the choice of individual that will maximize the production function is the one that yield the lowest opportunity cost: the follower with the lowest competencies. However, if leadership and followership competencies are positively correlated, then this will result in selecting a leader with low competencies.

**Important, absent leader ( $\alpha = \beta = 1$ ).** As leadership is contextually important, but does not prevent the leader from performing as a follower, there are no costs to appointing a leader:

$$y = [1 + l_i] \left( f_i + \sum_{j \neq i}^n f_j \right)$$

As a consequence, the most competent leader can be selected without any opportunity costs, irrespective of the correlation between leadership and followership competencies.

**Important, full-time leader ( $\alpha = 1; \beta = 0$ ).** When leadership is contextually important and takes all of the time available, the opportunity costs of appointing a leader are at their highest:

$$y = [1 + l_i] \left( \sum_{j \neq i}^n f_j \right)$$

Although leadership is crucial, the selected individual will no longer be able to perform as a follower. If leadership and followership competencies are positively correlated, then selecting the most competent leader will also yield a large loss of production. However, if the correlation

is negative, then the trade-off is no more as competent leaders are not the most competent followers.

**Unnecessary, absent leader ( $\alpha = 0$ ;  $\beta = 1$ ).** If leadership is contextually useless, and does not take any time, then there are no opportunity costs at all:

$$y = f_i + \sum_{j \neq i}^n f_j$$

In this case, notwithstanding the correlation between leadership and followership competencies, selecting an individual yields no loss of followership performance nor use of leadership competencies, and selection therefore does not matter.

Appointing an underperforming leader can therefore enhance the performance of a group, especially if three conditions are met. First, leadership and followership competencies must be positively correlated, so that a tradeoff arises from the competencies themselves. Second, leadership must be time-consuming, as leading reduces the leader's performance as a follower. Finally, this trade-off is exacerbated when leadership is contextually less important. Although informative, these four cases represent the extremes of the contextual importance and time-intensity of leadership. We therefore further investigate the various combinations of values these two elements can take on, and their impact on the competencies of the selected leader, when leadership and followership competencies are positively correlated. At the extreme, in a specific context the correlation between the competencies to be a leader and the ones required to be a follower could be perfect (Table 2).

-----

Insert Table 2 about here

-----

The trade-off then depends on the combination of the contextual importance of the leadership and the time required to lead. The selected leader for each combination is illustrated by Figure 1.

-----  
Insert Figure 1 about here  
-----

The grey areas represent the individual selected as a leader for a given combination of importance of leadership ( $\alpha$ ) and time spent as a leader ( $1 - \beta$ ). As one could expect, and in line with the literature, in most cases the optimal leader for the group is the one with the highest leadership competencies: individual 1. However, as the time required to lead increases, the competencies of the optimal leader decrease. This decrease of leadership competencies is driven by the opportunity costs: because leadership and followership competencies are negatively correlated, the more time a leader spends leading, the higher the loss of performance as a follower. As illustrated, this is especially true when the contextual importance of leadership is low as well, as the impact of the leader's competencies on the followers' performance is reduced. Overall, Figure 1 shows that, while in many instances leader selection can be thought of as a quest for the most competent individual, selecting underperforming leaders can sometimes be optimal, and therefore not a mistake. Indeed, the latter cases can lead to organizational issues, such as followers being dissatisfied with their leader, as some of them are more competent than she is. However, with the bigger picture in mind, we can see that what really matters for leader selection is not the competencies of one individual, but the performance of the whole group.

## What matters: role allocation

We have so far argued that leader selection should not rely solely on leader competencies. Rather, the optimal allocation of roles depends on the trade-off between what the individuals can bring as leaders and what the group can lose if these individuals stop acting as followers. Moreover, this trade-off is sensitive to the contextual importance of leadership. As illustrated in the previous examples, organizations should be particularly wary of opportunity costs of selecting a leader when the contextual importance of leadership is low to moderate, leadership is time-intensive, and the correlation between leadership and followership competencies is negative. Indeed, selecting the best leader in those contexts without accounting for opportunity costs may hamper rather than maximize group output. The selection of the optimal leader for the group ultimately boils down to one crucial element: the trade-off between selecting a leader and losing that individual as a follower. However, in order to select the best leader, the competencies of the individuals for both roles must be known. While followership can be assessed through the past performance of individuals, it can be difficult to make projections regarding their competences as leaders, especially if they have never been in such a role before. In The next section we will discuss how to select leaders when the competencies of the individuals as followers are known but their leadership competences are unknown. Moreover, we discuss what organizations can do to assign the right individual to the role of a leader while keeping other group members motivated.

## III. How to select leaders

### Getting the right people at the right place

In an ideal world, an organization would know the competencies of each of its individuals for all the roles that needs to be filled. With this information in hands, the organization can achieve the optimal role allocation by taking into account their opportunity

costs. However, it is rather uncommon for organizations to have prior knowledge of how well its members perform in roles that they have never held. Although the leadership literature has identified some of the relevant factors predicting effective leadership (e.g., Amundsen & Martinsen, 2014; Antonakis, 2014; Judge, Bono, Ilies, & Gerhardt, 2002; Podsakoff, MacKenzie, Moorman, & Fetter, 1990), we argue that in practice many organizations know how their individuals perform as followers, due to observing their performances in the past, but mainly speculate when it comes to leadership competencies. Indeed, we acknowledge that accurately measuring leadership skills prior to becoming a leader may be difficult, therefore we suggest that organizations can resort to use followership competencies as a proxy for leadership competencies. As such, individuals with high followership competencies may also have high leadership abilities. Though simple, this rule of thumb is widely used in practice to promote individuals within the organization. In addition, promoting the best followers to leadership roles is also an effective extrinsic motivational tool. The idea of motivating individuals via promotions is at the core of tournament theory (Becker & Huselid, 1992; Knoeber & Thurman, 1994; Lazear & Rosen, 1981). Tournament theory posits that in order to motivate individuals, an organization can offer a promotion to the individual who will be the most performant follower. The reward that comes with the promotion provides an incentive for individuals to try harder and get promoted, which efficiently allows to select the top performer for a higher rank (assuming that the best follower is also the best leader).

Though tournament theory is well established in both management and economics (see Connelly, Tihanyi, Crook, & Gangloff, 2014), we suggest that in the case of leader selection, it may lead to motivational issues when accounting for opportunity costs. Recall that for tournament theory to work, individuals must have an incentive to show the best of themselves and that being promoted to a leadership role is the incentive. Yet, we have so far made the case that, in some instances, organizations may be better off keeping top performers in followership

positions because of the opportunity costs arising from losing their best followers. Instead, by selecting underperforming individuals for leadership positions, organizations remove the incentive to do as best as one can in a followership position. Consequently, followers might not be motivated by the tournament, as performing well would lower their chances of being selected for a leadership role. Even worse, these individuals may have an incentive to downplay their competencies, in order to have a chance to be selected as the leader. By associating tournament theory to a selection process which accounts for opportunity costs, we pose the following puzzle: how can organizations ensure having the right people in the right place when there are no incentives to be there?

One way to tackle this issue would be to focus on the individuals more likely to lose motivation: the best followers. In order to achieve efficient selection, the organization needs to provide these top performers with an incentive to keep their performance high, as they are too valuable to be selected as leaders. In other words, these followers need to be better off remaining followers than they would, should they be promoted. While the classical approach to tournament theory assumes that individuals are motivated by a promotion to a leadership role, we suggest that organizations should devise incentives schemes which offset the importance of hierarchical promotions. Under such mechanisms, individuals should have an incentive to show the best of themselves in followership roles and be less attracted to leadership roles. Some organizations have already implemented similar incentives schemes. For example, in the banking sector, non-managing vice-president and president titles are commonly used to reward top-performing sales representatives alongside bonuses and salary increases. These non-managing titles serve as incentives for individuals to perform as best as they can in sales positions but are not attached to any leadership responsibilities. By doing so, these organizations created an incentive to perform well in followership roles and avoid (or at least postpone) the promotion of top performers sales representatives to leadership roles. In parallel,

individuals whose performance as follower is not high enough to be indispensable can enter a more traditional tournament for the leadership position. Consequently, the most performant followers are motivated by the rewards inherent to their follower's position, while less performant ones are motivated by the promotion to the leader's role.

What should organizations consider when selecting a leader?

The correlation between leader and follower competencies has an important impact for the essence of our model: the opportunity cost of leading. In a group, if the best individuals with the best followership skills also happen to have the best leadership skills, then the opportunity cost of selecting the most competent leader will be high. As good followers are also good leaders and bad followers are bad leaders, selecting a good leader removes a good follower from the group. Assuming the contextual importance of leadership is low, and leading is time-intensive, it can be optimal for the group to select an individual who is less competent than the most competent individual in the group in terms of leadership. Of course, the opposite is true when leadership and followership competencies are negatively correlated, as selecting the most competent leader will remove a follower with low followership competencies. In such a group, one can afford to select the most competent leader, as the opportunity cost of selecting this individual as a leader is low. In summary, our model suggests that organizations should consider opportunity costs when selecting leaders and that these opportunity costs will have varying importance depending on the relationship between followership and leadership competencies, but also on the contextual importance of leadership, how time-intensive leading is.

## IV. General discussion

### Theoretical implications

Our paper makes several theoretical contributions to the leadership literature. First, our model sheds light on an aspect of leader selection often overlooked: the opportunity costs. As we have argued, selecting an individual as a leader can prevent her from performing as a follower. These opportunity costs imply that under certain circumstances, it can be optimal for a group to select a leader who is not the most competent member of the group. The importance of opportunity costs suggests that scholars should either control for opportunity costs when examining the leader selection process or formally make the assumption that opportunity costs are absent from the context under scrutiny. Our model also raises the question of whether existing theories of leader selection remain unchanged when controlling for opportunity costs. Specifically, by using tournament theory, we have switched from framing followership and leadership as two hierarchical levels to two different roles. Tournament theory suggests that as some followers may be too valuable to be appointed as leaders, it is important to reward them accordingly in order to avoid any negative impact on their motivation. The change of paradigm which assumes that being a leadership position is a better condition than being in a followership position is challenged. In fact, the importance of opportunity costs suggests that under certain circumstances, organizations should grant top performers in followership positions with better incentives than if they were to be promoted in leadership positions. By reframing leading and following as two different roles, we hope to contribute to the call for a better integration of leadership and followership theories. Finally, we have emphasized the importance of the environment in discussing leader selection. By environment, we mainly refer first to the relationship between followership and leadership competencies, and second to the contextual importance and time-intensity of leadership. Although easy to overlook, these three environmental factors impact the selection process, and thus, the competencies of the selected

leader. As with opportunity costs, we suggest that scholars should be particularly weary when theorizing and conducting empirical research on leader selection as they may often overlook the contextual importance of leadership (i.e., is leadership important in the focal environment) and the relationship between followership and leadership competencies.

Finally, it is worth noting that our conclusions do not depend on the functional form of the production function. Although our production function may only be applicable to certain contexts, the trade-off between the selection of a leader and the loss of a follower remains relevant, as individuals are a limited resource. Although Ricardo's (1817) notion of comparative advantage initially described the decision of countries to specialize in the production and trading of certain goods, the importance of opportunity costs is widely acknowledged in various domains, ranging from individual effort provision decisions (Kurzban, Duckworth, Kable, & Myers, 2013) to public investment decisions (Marglin, 1963). Eventually, opportunity costs are relevant whenever one needs to decide on how to allocate limited resources.

### Practical implications

Moreover, this model also has several practical implications. First, related to the last theoretical point, we argue that practitioners should focus more on followership competencies and the correlation between the latter and leadership competencies in their own context. We expect this correlation to vary across jobs as well as fields. In that regard, we recommend that future research considers examining followership measures. While the characteristics of followers more responsive to a leader matter, it is also crucial to understand what kind of tasks followers are responsible for in a given environment and how to measure their performance for these tasks. Accurately measuring follower performance is not only essential to evaluate the opportunity costs of selecting a leader, but they also provide information about the relationship between leadership and followership competencies. This knowledge can then be used to design

a more appropriate selection process. Similarly, testing for leadership competencies should not be limited to existing leaders as it is often the case within organizations. Instead, measuring leadership competencies should be done on a more systematic basis to have a clear understanding of the strengths and weaknesses of each group members. By generating a complete mapping of competences in followership and leadership roles, organizations can ensure that they account for opportunity costs in the leadership selection process. Second, we propose that it is time to challenge the idea that leaders are more important than followers and as a consequence, they should be rewarded more. As argued, while it is important to remunerate leadership enough to have an efficient tournament, excellent followers must be better off in their own role than in a leadership position if an organization wants to thrive. As previously illustrated in this paper, some organizations have already developed financial and non-financial incentives to reward excellent followers, such as sales representatives. These sales representatives benefit from a network of clients and a trust level that cannot be easily replaced by other individuals should they be appointed to leadership positions. Therefore, it is crucial for organizations to keep their key performers motivated using a different incentives scheme than by simply promoting them hierarchically. This is key to using tournaments as an efficient selection method.

### Recommendations for future research

As discussed in our theoretical implications, we call for more research accounting for opportunity costs. It is recommended that management Scholars investigate the specific circumstances in which opportunity costs matter the most and whether organizations already weight opportunity costs in the selection process. Furthermore, scholars may want to investigate how organizations can implement new leader selection processes accounting for opportunity costs. While we have discussed the specific case of tournaments, it is easy to imagine that selecting individuals based on opportunity costs may trigger concerns in terms of

perceived fairness. For example, top performing followers who are not promoted to leadership positions because of opportunity costs might become demotivated and suffer performance losses. A reduced perception of the procedural fairness can lower employee motivation, impacting the whole selection process and its outcome. It is also possible that the selection process we outlined leads to lower performance appraisals for the leader, despite being optimal for the group. As some followers would be more competent leaders, they may perceive the leader as incompetent. For example, past research has shown that the evaluation of a superior decreases for higher levels of education of the evaluator (Artz et al., 2020). New selection processes should account for motivation and transparency concerns as they depart from the classical view that the best follower should be appointed to the leadership position regardless of opportunity costs. Finally, we hope that leadership scholars will consider using mathematical models to develop theories of leadership as these models require making clear theoretical assumptions and as such, may reveal unexplored avenues for future research. While Using mathematical models to theorize is common in economics, this practice is less prevalent in management. Drawing from economic theories, in this paper we have applied a simple yet, well-developed idea to leadership selection: groups outperform when role allocating takes into account opportunity costs. Thus, we answer the existing call for a better integration of economics to the field of leadership (Zehnder et al., 2017) and argue that scholars can develop existing and new theories of leadership by applying methods common practice in economics.

## Conclusion

Our main goal was to change the preconceived idea that selecting the individuals with the best leadership skills is systematically the best strategy for a group. As argued, organizations should operationalize leader selection as a role allocation decision and consider the opportunity costs of promoting an individual to a leadership position. Through different cases, we have showed that selecting leaders who are not the most competent is not necessarily

a symptom of a failure in the selection process. On the contrary, we provided arguments as to why it is sometimes efficient for the whole group to select underperforming leaders and keep the most competent followers in those roles. Surprisingly as the share of human capital became increasingly important in advanced economies, leadership studies remained relatively agnostic to the importance of opportunity costs in the leadership selection process. Omitting opportunity costs in leadership selection and more generally in the management of human resources is even more surprising considering that opportunity costs have been used to allocate capital and financial resources within organizations. At a time where scholars call for more integration of followers in leadership research (Bligh, 2011; Uhl-Bien et al., 2014) as well as building bridges between different research fields, we believe that following this avenue is both timely and relevant.

## V. References

- Amundsen, S., & Martinsen, Ø. L. (2014). Empowering leadership: Construct clarification, conceptualization, and validation of a new scale. *The Leadership Quarterly*, 25(3), 487-511.
- Antonakis, J., House, R. J., & Simonton, D. K. (2017). Can super smart leaders suffer from too much of a good thing? The curvilinear effect of intelligence on perceived leadership behavior. *Journal of Applied Psychology*, 102(7), 1003-1021.
- Antonakis, J., & Dalgas, O. (2009). Predicting Elections: Child's Play! *Science*, 323(5918), 1183.
- Antonakis, J., & Day, D. V. (2017). Leadership: Past, Present, and Future. In J. Antonakis & D. V. Day (Eds.), *The Nature of Leadership* (3rd Revised edition ed.). Thousand Oaks, United States: SAGE Publications Inc.
- Artz, B., Goodall, A. H., & Oswald, A. J. (2020). How Common Are Bad Bosses? *Industrial Relations: a journal of economy and society*, 59(1), 3-39.
- Arvey, R. D., Rotundo, M., Johnson, W., Zhang, Z., & McGue, M. (2006). The determinants of leadership role occupancy: Genetic and personality factors. *The Leadership Quarterly*, 17(1), 1-20.
- Bass, B. M. (1999). Two decades of research and development in transformational leadership. *European journal of work and organizational psychology*, 8(1), 9-32.
- Bass, B. M., & Avolio, B. J. (1994). *Improving organizational effectiveness through transformational leadership*: Sage.
- Becker, B. E., & Huselid, M. A. (1992). The incentive effects of tournament compensation systems. *Administrative Science Quarterly*, 336-350.
- Bligh, M. C. (2011). Followership and follower-centered approaches. *The Sage handbook of leadership*, 425-436.

- Bligh, M. C., Kohles, J. C., & Meindl, J. R. (2004). Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11th terrorist attacks. *The Leadership Quarterly*, *15*(2), 211-239.
- Connelly, B. L., Tihanyi, L., Crook, T. R., & Gangloff, K. A. (2014). Tournament theory: Thirty years of contests and competitions. *Journal of Management*, *40*(1), 16-47.
- Day, D. V., Fleenor, J. W., Atwater, L. E., Sturm, R. E., & McKee, R. A. (2014). Advances in leader and leadership development: A review of 25 years of research and theory. *The Leadership Quarterly*, *25*(1), 63-82.  
doi:<http://dx.doi.org/10.1016/j.leaqua.2013.11.004>
- Den Hartog, D. N., House, R. J., Hanges, P. J., Ruiz-Quintanilla, S. A., & Dorfman, P. W. (1999). Culture specific and cross-culturally generalizable implicit leadership theories: Are attributes of charismatic/transformational leadership universally endorsed? *The Leadership Quarterly*, *10*(2), 219-256.
- Einarsen, S., Aasland, M. S., & Skogstad, A. (2007). Destructive leadership behaviour: A definition and conceptual model. *The Leadership Quarterly*, *18*(3), 207-216.  
doi:<http://dx.doi.org/10.1016/j.leaqua.2007.03.002>
- Elgar, M. A. (2016). Leader selection and leadership outcomes: Height and age in a sporting model. *The Leadership Quarterly*, *27*(4), 588-601.
- Erickson, A., Shaw, J. B., & Agabe, Z. (2007). An empirical investigation of the antecedents, behaviors, and outcomes of bad leadership. *Journal of Leadership Studies*, *1*(3), 26-43.
- Foti, R. J., Fraser, S. L., & Lord, R. G. (1982). Effects of leadership labels and prototypes on perceptions of political leaders. *Journal of Applied Psychology*, *67*(3), 326-333.
- House, R. J., & Howell, J. M. (1992). Personality and charismatic leadership. *The Leadership Quarterly*, *3*(2), 81-108.

- Howell, J. M., & Avolio, B. J. (1992). The ethics of charismatic leadership: submission or liberation? *Academy of Management Perspectives*, 6(2), 43-54.
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87, 765-780.
- Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, 89, 542-552.
- Judge, T. A., & Piccolo, R. F. (2004). Transformational and transactional leadership: A meta-analytic test of their relative validity. *Journal of Applied Psychology*, 89(5), 755-768.
- Knoeber, C. R., & Thurman, W. N. (1994). Testing the theory of tournaments: An empirical analysis of broiler production. *Journal of Labor Economics*, 12(2), 155-179.
- Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and brain sciences*, 36(6), 661-679.
- Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5), 841-864.
- Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, 28(1), 18-27.
- Lord, R. G., & Brown, D. J. (2001). Leadership, values, and subordinate self-concepts. *The Leadership Quarterly*, 12(2), 133-152.
- Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A Test of Leadership Categorization Theory: Internal Structure, Information Processing, and Leadership Perceptions. *Organizational Behavior and Human Performance*, 34, 343-378.
- Luthans, F., & Avolio, B. J. (2003). Authentic leadership: A positive developmental approach. In K. S. Cameron, J. E. Dutton, & R. E. Quinn (Eds.), *Positive organizational scholarship* (pp. 241 - 261). San Francisco: Barrett-Koehler.

- Marglin, S. A. (1963). The opportunity costs of public investment. *The Quarterly Journal of Economics* 77(2), 274-289.
- Meindl, J. R. (1995). The romance of leadership as a follower-centric theory: A social constructionist approach. *The Leadership Quarterly*, 6(3), 329-341.
- Meindl, J. R., Ehrlich, S. B., & Dukerich, J. M. (1985). The romance of leadership. *Administrative Science Quarterly*, 78-102.
- Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *The Leadership Quarterly*, 1(2), 107-142.
- Ricardo, D. (1817). *The Principles of Political Economy and Taxation*. Reprint. *Londong Dent*.
- Schyns, B., & Schilling, J. (2013). How bad are the effects of bad leaders? A meta-analysis of destructive leadership and its outcomes. *The Leadership Quarterly*, 24(1), 138-158.
- Shamir, B., House, R. J., & Arthur, M. B. (1993). The motivational effects of charismatic leadership: A self-concept based theory. *Organization Science*, 4(4), 577-594.
- State of the Global Workplace*. (2017). Retrieved from New York:
- Uhl-Bien, M., Riggio, R. E., Lowe, K. B., & Carsten, M. K. (2014). Followership theory: A review and research agenda. *The Leadership Quarterly*, 25(1), 83-104.
- Van Vugt, M., & Ahuja, A. (2010). *Selected: Why some people lead, why others follow, and why it matters*: Profile Books.
- Weierter, S. J. (1997). Who wants to play "follow the leader?" A theory of charismatic relationships based on routinized charisma and follower characteristics. *The Leadership Quarterly*, 8(2), 171-193.
- Wilson, D. S., Van Vugt, M., & O'Gorman, R. (2008). Multilevel selection theory and major evolutionary transitions: Implications for psychological science. *Current Directions in Psychological Science*, 17(1), 6-9.

Zehnder, C., Herz, H., & Bonardi, J.-P. (2017). A productive clash of cultures: Injecting economics into leadership research. *The Leadership Quarterly*, 28(1), 65-85.

## VI. Appendix A

*Table 1: Competencies of two individuals in role A and role B.*

Individual	Follower competencies	Leader competencies	Group output <sup>2</sup>
1	7	8	40
2	5	6	42

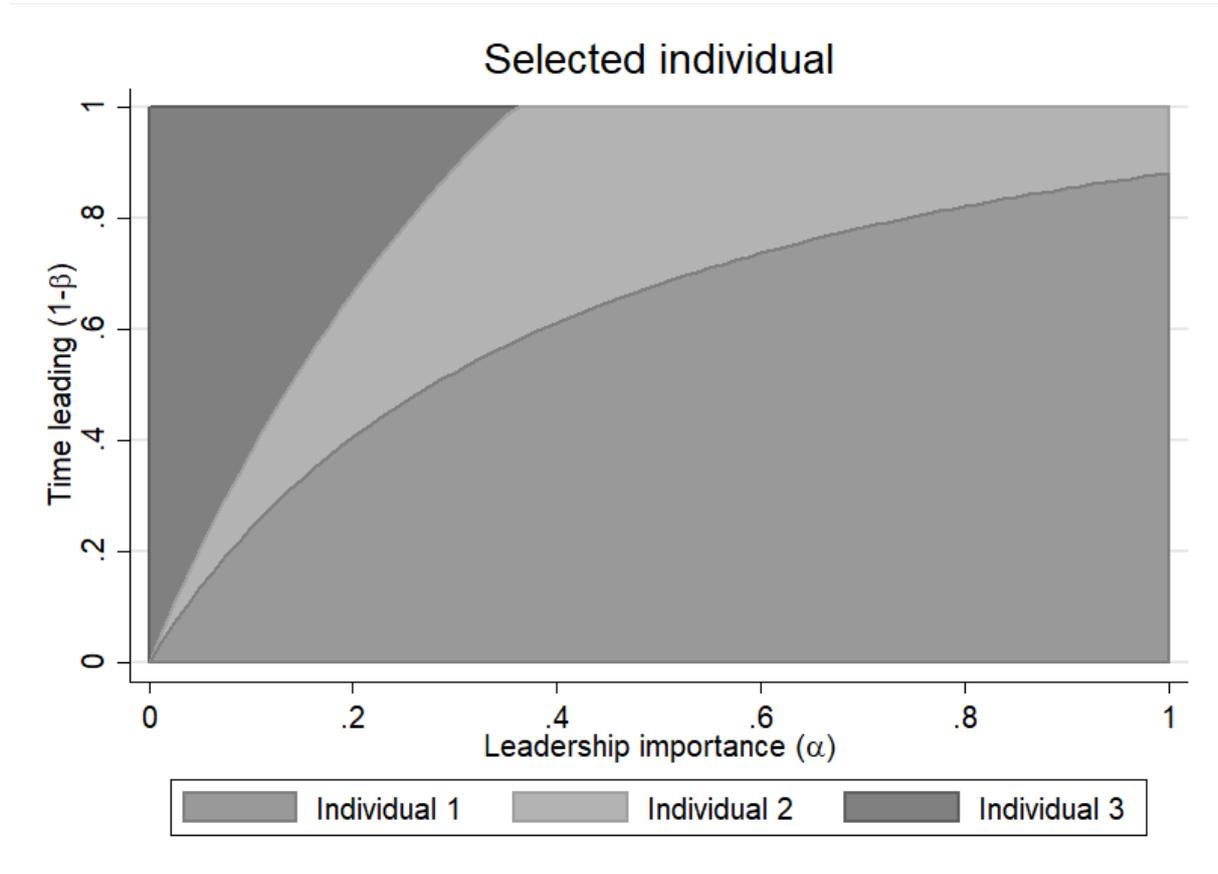
*Table 2: Summary of followership and leadership competencies.*

Individual	Followership output	Leadership output
1	7	1.4
2	5	1
3	3	0.6

---

<sup>2</sup> You can find the calculations of the group outputs hereafter. If individual 1 is the leader, the group's output is  $5 * 8 = 40$ . If individual 2 is the leader, the group's output is  $7 * 6 = 42$ .

Figure 1: Selected leader depending on the combinations of contextual importance and time-intensity of leadership.



## VI. Appendix B

Production function (full model)

Starting from the basic model:

$$y = l_i \sum_{j \neq i}^n f_j$$

We add the contextual importance of leadership ( $\alpha$ ) as an enhancer of followership competencies:

$$y = \alpha l_i \sum_{j \neq i}^n f_j + \sum_{j \neq i}^n f_j$$

We allow individuals to spend a proportion of their time following ( $\beta$ ), while the rest is spent leading ( $1 - \beta$ ):

$$y = \alpha l_i \left( \beta f_i + \sum_{j \neq i}^n f_j \right) + \beta f_i + \sum_{j \neq i}^n f_j$$

Simplifying, we obtain:

$$y = [1 + \alpha l_i] \left( \beta f_i + \sum_{j \neq i}^n f_j \right)$$

## **Boundedly Ethical: How Small Misbehaviors Open the Door for More**

### Abstract

Despite a large body of research on (im)moral behaviors, the psychological mechanisms underlying moral cleansing – the act of compensating for a past misbehavior through a good deed – remain unclear. I propose to explain the activation of cleansing behaviors through the magnitude of the difference between one's moral identity and the past misbehaviors. In order to test my hypotheses, I collected behavioral data in a controlled and incentivized experiment using MTurk (n=615) about charity donations, manipulating the extent of moral deviations through different degrees of temptation. Overall, I find that cleansing behaviors can be explained by how different an individual's behavior is from their moral identity, although the reactions to the treatments are relatively small due to a majority of individuals refusing to provide more efforts to help the charity. I therefore conducted a type analysis, revealing that the difference between moral identity and misbehaviors explains moral cleansing, but also depends on the initial morality of the participants.

Keywords: behavioral ethics, bounded ethicality, donations.

*Matthieu Légeret*

*University of Lausanne / Switzerland*

*matthieu.legeret@unil.ch*

## INTRODUCTION

Individual codes of conduct are very personal and there is wide variety in the extent to which people engage in (im)moral behaviors. An individual's moral identity is defined as a personal set of appropriate behaviors (Aquino & Reed II, 2002; Reynolds & Ceranic, 2007). However, moral identities are not fully static and can be subject to dynamic processes. For instance, if individuals realize that a past behavior is in conflict with their own moral codes, they tend to experience disgust (e.g., Chapman & Anderson, 2013) and may respond with moral cleansing. Whether literal, such as washing one's hands (Zhong & Liljenquist, 2006), or more abstract, such as donating money to a charity (Gneezy, Imas, & Madarász, 2014), moral cleansing aims at compensating for a past immoral behavior (Sachdeva, Iliev, & Medin, 2009). However, there is anecdotal evidence showing that, under some circumstances, individuals do not realize that they are facing an internal moral conflict, or they do not even see the moral aspect of a decision they are involved in, until it is too late.

Gioia (1992) provides an illustrative example for such moral numbness. In 1973, Gioia was working at Ford as recall coordinator. At that time, Ford started receiving reports of the Ford Pinto catching on fire after being involved in accidents at relatively low speed. However, these incidents were not frequent enough to "capture particular attention" (Gioia, 1992:382). In 1978, an investigation by the National Highway Transportation Safety Bureau concluded that the Pinto was faulty, and Ford was heavily criticized for carefully calculating the costs and benefits of recalling the Pinto without considering the moral aspect of the decision (Dowie, 1977; Gatewood & Carroll, 1981). While the comparison of the costs of recalling the cars to the benefits of preventing drivers' death is relevant from a business perspective, many people would agree that purely focusing on this aspect of the problem would not be in line with the behavior that their moral identity prescribes. If Ford's employees were facing internal conflicts, why was moral cleansing not observed?

While the experimental study of immoral behaviors allowed to identify patterns such as moral cleansing, it has often neglected the importance of the perception of internal moral conflicts in triggering these patterns (e.g., see Effron, Cameron, & Monin, 2009; Zhong & Liljenquist, 2006). In fact, moral identity has been shown to be key in understanding immoral behaviors, as individuals often hold distorted beliefs about their own morality (Bazerman & Sezer, 2016; Tenbrunsel & Messick, 2004; Zhang, Fletcher, Gino, & Bazerman, 2015). However, as the literature has neglected the internal conflict between moral identity and behaviors, the mechanisms underlying moral cleansing remain ambiguous (Dolan & Galizzi, 2015).

When studying the dynamics of moral behavior, it is crucial to understand where behavioral deviations from one's own moral identity come from. In the field of behavioral ethics, the framework of bounded ethicality explains morally inconsistent behaviors by assuming that individuals are naïve in that they hold incorrect beliefs regarding their capacities to behave morally. Most individuals overestimate how moral they are as well as their resistance to immoral temptation, which prevents them from seeing how certain situations can lead to immoral behaviors (Chugh & Bazerman, 2007; Chugh, Bazerman, & Banaji, 2005). Bounded ethicality, hence, assumes that individuals make systematic mistakes and fail to see that these mistakes make them deviate from their own moral principles (Banaji, Bazerman, & Chugh, 2003; Banaji & Bhaskar, 2000; Bazerman & Gino, 2012; Chugh et al., 2005).

Past experimental research mainly focused on interventions preventing misbehaviors (e.g., Desai & Kouchaki, 2017; Mayer, Aquino, Greenbaum, & Kuenzi, 2012), but the psychological mechanisms underlying misbehaviors and reactions to misbehaviors have been neglected. In this paper, I extend the framework of bounded ethicality and hypothesize that the magnitude of the deviation from the moral code is decisive for the individuals' dynamic

reaction. I argue that both moral cleansing and repetition of misbehaviors can emerge depending on the magnitude of prior deviations from moral behavior.

I concentrate on contexts where an individual can make the decision to prioritize themselves over a third party. I develop a theoretical model predicting that the behavioral response to previous misbehavior depends on the magnitude of the deviation from the individual's own moral code. The intuition behind the model is the following: If individuals engage in behavior that stands in strong contrast to their moral identity, they are likely to consciously recognize the moral conflict and may therefore engage in moral cleansing to compensate their wrongdoing. Smaller deviations from their initial moral code, in contrast, may remain undetected and may therefore corrupt their moral identity. Consequently, small misbehaviors may be more dangerous than large ones, because individuals may not react to the former so that their moral identity depreciates over time.

This paper makes several contributions to the literature. First, I extend the theoretical literature on bounded ethicality by proposing that reactions to own misbehaviors depend on the magnitude of the deviation from one's own moral code. Second, I propose an experiment that allows us to directly test the central predictions of my model. This setting allows us to elicit behavioral measures at various stages: before, during, and after misbehaviors occurred. Measuring behaviors before a misbehavior allows us to capture the baseline of moral behaviors, which can then be compared to the misbehavior to gauge the extent to which the individual deviated from their moral self. Finally, I advance research on moral cleansing by using continuous behavioral measures. With a few exceptions (e.g., Zhong & Liljenquist, 2006), most of the literature on moral cleansing has been using binary measures, operationalizing cleansing as a probability to cleanse (e.g., Effron et al., 2009). I therefore propose to identify the different degrees to which individuals cleanse through continuous measures and provide evidence for the mechanisms behind moral cleansing.

## THEORETICAL BACKGROUND

In behavioral approaches to ethics, the past two decades have seen the rise of the framework of bounded ethicality (Banaji & Bhaskar, 2000; Bazerman & Gino, 2012; Bazerman & Tenbrunsel, 2011; Chugh & Bazerman, 2007; Chugh et al., 2005). Bounded ethicality refers to the “psychological processes that lead people to engage in ethically questionable behaviors that are inconsistent with their own preferred ethics” (Bazerman & Sezer, 2016, p. 99). The framework of bounded ethicality can be described as a transposition of bounded rationality (Simon, 1955, 1956) to the domain of morality. Echoing the seminal work of Herbert Simon, individuals are assumed to only be boundedly aware of their own limitations as well as the moral aspect of certain situations<sup>1</sup>. 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 center of its explanations, it says little about the psychological mechanisms underlying misbehaviors nor the reaction to misbehaviors.

Part of the psychological literature on (im)moral behaviors has focused on individuals’ reactions to their own misbehaviors. Indeed, for most individuals such misbehaviors come as

---

<sup>1</sup> 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).

a paradox between one's moral identity and behavior and have to be rationalized one way or another (Tenbrunsel & Messick, 2004). There are various ways through which individuals deal with identity conflicts ranging from psychological to behavioral. 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 misbehaviors (moral cleansing<sup>2</sup>). Moral cleansing is used to describe the behavior of an individual who behaves morally after having behaved immorally, in order to compensate for the previous behavior (Sachdeva et al., 2009; Stone, Aronson, Crain, Winslow, & Fried, 1994; Zhong & Liljenquist, 2006). In line with bounded ethicality, evidence shows that the higher the moral self, the more likely the cleansing (Tetlock, Kristel, Elson, Green, & Lerner, 2000). 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 behavior has not yet been fully understood (Dolan & Galizzi, 2015).

In this paper I investigate how individuals react to misbehaviors of different magnitudes, that is, departures from their own identity to different degrees. Drawing on the work of Simon on bounded rationality (1955, 1956), I propose to distinguish between salient and inconspicuous misbehaviors. 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 misbehavior as they were instructed to punish another individual with potentially harmful and even fatal electric shocks.

---

<sup>2</sup> 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 of compensation for past immoral behaviors.

As a reaction, several participants verbally expressed their discomfort and even lied to the experimenter to avoid hurting others. However, if a misbehavior 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), and resembles the automatic decision-making process described by Gioia (1992) in the Ford Pinto example. I therefore argue that one's moral identity serves as a reference point. If an individual believes they are highly moral and misbehaves, the discrepancy between their identity beliefs and actual behavior 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 behaviors. On the contrary, if the misbehavior is inconspicuous, then this behavior 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 behavior, although the latter will no longer be considered as such due to the identity being adapted. My psychological mechanism can be described through the metaphor of a frog entering a cooking pot full of water. If the water is already boiling, the frog will react and try to escape. However, if the temperature of the water is slowly increasing, then the frog will not realize the danger before it is too late. Consequently, I hypothesize that salient moral dissonance has a positive impact on the likelihood of future moral behaviors. This behavioral pattern – a misbehavior triggering a better behavior – is a behavioral expression of moral cleansing.

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

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

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

In a nutshell, individuals display different behavioral patterns depending on whether their moral dissonance is salient or not (see Figure 1).

-----  
Insert Figure 1 about here  
-----

Furthermore, at the core of bounded ethicality resides the beliefs about one's moral identity (Banaji et al., 2003; Chugh et al., 2005; Messick & Bazerman, 1996; Zhang et al., 2015). For instance, Chugh et al. (2005) argued that individuals who believe they are more moral than they actually are may be blinded by these beliefs, and are consequently more likely to misbehave when tempted. In the same vein, past research has made the argument that individuals overestimate their morality, compared with others' (Banaji et al., 2003; Tenbrunsel, 1998). However, as I discriminate between salient and inconspicuous misbehaviors, I need to distinguish between the beliefs about moral identity and the beliefs about sensitivity to temptations. An individual who overestimates their moral identity is more likely to face salient misbehaviors, as their reference point is high. Likewise, underestimating one's sensitivity to temptation leads to salient misbehaviors, as one's reaction to temptations is unexpectedly high. Despite identity and sensitivity beliefs being central to the framework of bounded ethicality, there have been very little empirical investigations of the difference between the two. I therefore propose to compare these two types of beliefs about oneself to i) the same beliefs about others, and ii) the actual behaviors. As past research suggests, I hypothesize that individuals overestimate their moral identity and underestimate their sensitivity, compared to

both others' identity and sensitivity, and compared to their actual moral behavior and sensitivity.

*Hypothesis 3a: Individuals overestimate their moral identity compared to both their perception of others' moral identity and their own actual behavior.*

*Hypothesis 3b: Individuals underestimate their sensitivity to temptation compared to both their perception of others' sensitivity to temptation and their own actual sensitivity.*

To summarize, misbehaviors can lead to either moral cleansing or subsequent misbehaviors. I propose that the saliency of the prior misbehavior is what drives the two opposing behavioral patterns. Moreover, I posit that individuals believe they are generally more moral than others: they overestimate their moral identity and underestimate their sensitivity to temptations.

## **EXPERIMENTAL SETTING**

I conducted an online experiment in order to test my hypotheses. In a first step, I collected the behavioral data for Hypotheses 1 and 2, using the effort task described by Abeler, Falk, Goette, and Huffman (2011). Participants were paid a fixed wage of 6 USD, for an average completion time of 50 minutes (7.12 USD per hour). One week later, I collected additional data through a questionnaire in order to control for demographics, as well as altruism and time preferences. The latter were measured using two items from the preference survey modules (Falk, Becker, Dohmen, Huffman, & Sunde, 2016). Participants who answered the follow-up questionnaire received 0.5 USD and took on average 1.85 minutes to complete it (16.2 USD per hour). Finally, I replicated the experiment in a hypothetical setting, asking a new pool of participants what they believe others chose, and what they believe they would have done. This last step allows us to test Hypotheses 3a and 3b regarding the importance of beliefs about the moral self and sensitivity to temptations. In this hypothetical setting, participants

were paid 0.5 USD for participating, and could earn an additional bonus based on their answers. On average, participants took 6.2 minutes to answer the questionnaire and were paid 0.62 USD (6 USD per hour).

**Sample and design.** I collected data from 615 participants on Amazon Mechanical Turk, in three batches in consecutive days, only allowing US citizens with prior experience on MTurk and an approval rate of at least 90% to participate in order to maintain the relevance of the charity. I ran three experimental conditions, two of which have 200 participants, while the third one has 215. A total of 925 workers started the experiment, but only 615 completed it. Among the 310 workers who dropped out, only 16 dropped out after being exposed to a treatment. Moreover, these dropouts were not concentrated on a particular treatment: 5 were in the first treatment, 7 were in the second treatment, and 4 were in the third treatment. These data indicate that there was no treatment-specific attrition in the experiment. In terms of demographics, I have a predominantly male sample (59.76%), with age ranging between 18 and 70 years old (median between 31 and 35 years old). The sample is predominantly white (76.68%), with a median level of education equal or equivalent to a bachelor's degree (52.64%).

**Procedure.** The experiment was programmed using oTree (Chen, Schonger, & Wickens, 2016), allowing to display experiments on web browsers and therefore facilitating data collection on online platforms such as MTurk. The experimental design consisted of three parts intended to capture each stage of the moral updating. In each part, I asked participants to count how many ones are present in a table randomly filled with ones and zeroes. As argued by Abeler et al. (2011), participants do not need any prior knowledge to complete this effort task. Moreover, the performance is easy to measure and incurs an effort cost on the participants due to the repetitive nature of the task. Participants had to choose between two options: a small and a large table (see Figure 1). In order to prevent learning effects from impacting the experiment,

participants counted four small and four large tables in a training phase, helping them familiarize with the task.

Past the training phase, counting the ones in the large table donates 10 cents to a charity, while counting the ones in the small table removes 1 cent from the overall donation. I purposefully chose a relevant charity given the context of the study: Americares. At the time of data collection, Americares had issued a call for donations to help the direct and indirect victims of the COVID-19 pandemic. Participants were informed that, should they finish the experiment with a negative amount for the donation, this amount would be subtracted from the donations of other workers. One part containing 10 tables, participants made a total of 30 decisions with consequences ranging from an increase in the overall donation of 3 USD and a decrease of 30 cents. The tables in each of the three rounds are in fact the same, displayed in a random order to maintain the difficulty constant across rounds.

In the first and third parts, all participants were exposed to the same tables: small tables contain 6 lines and large tables contain 9 lines. Having the same parameters in the first and last parts allowed us to measure the impact of exposures to moral deviations. In the second part, I introduced the treatments by manipulating the number of lines in the small tables while keeping the consequences for the donation constant. The purpose was to reduce the efforts necessary to complete the small table to varying degrees, and consequently attract participants to reduce the donation to the charity. Participants were informed about the summary of their decisions and the consequences for the charity for the part they just finished, to prevent them from seeing the consequences of their decisions before a given part is over. A summary of the overall payoffs was displayed at the very end of the experiment.

Finally, participants had the possibility to make a donation out of their wage at the end of the experiment. As my experimental task is rather demanding, I allowed for ex-post

donations to control for participants who would prefer to finish the experiment as fast as possible and compensate with a donation at the end. This final donation allows to distinguish between participants who want to maximize their hourly wage at the expense of the charity from participants who want to minimize their effort but still care about the charity. While both would only choose low effort, the latter would make a donation at the very end of the experiment.

-----  
Insert Figure 2 about here  
-----

**Treatments.** At the beginning of the second part, participants were randomly assigned to one of my three conditions: control, low temptation, and high temptation. The two treatments manipulating temptation aim at tempting participants to choose the small table more often than they did in the first part of the experiment, in order to trigger moral deviations of different magnitudes. Manipulating the magnitude of the moral deviations through temptation serves two purposes. First, it is an operationalization of the saliency of moral deviations. Deviating in favor of a table that is similar to the usual one is less salient than deviating in favor of a table that is much shorter. The difference between the size of the tables chosen in the first part and the size of the tables chosen in the second part is therefore used as a proxy for saliency. Second, tempting participants instead of imposing a shorter table maintains their free will, and prevents them from disengaging from the consequences of their decisions (e.g., Bandura, 1990, 1999, 2002).

In the *control condition*, the respective sizes of the small and large tables remained the same as in the first and third rounds. It therefore serves as a baseline and captures any potential time trend of the effort choices (see Figure 3).

-----  
Insert Figure 3 about here  
-----

In the *low temptation treatment*, the number of lines in the small table was reduced from 6 to 5. While small, this change still reduces the effort needed to count the small table and renders it slightly more attractive (see Figure 4).

-----  
Insert Figure 4 about here  
-----

In the *high temptation treatment*, the number of lines in the small tables was further reduced to 3. This condition significantly reduces the time needed to count the small table, increasing its attractiveness (see Figure 5).

-----  
Insert Figure 5 about here  
-----

**Measures.** I operationalized moral behaviors as the proportion of large tables counted in each of the three parts of the experiment. In a demographic questionnaire, I elicited measures for altruism and time preferences by using two items described in Falk et al. (2016). Altruism is measured by telling participants to imagine they just won \$1000 and asking how much of this amount they would donate to a charity. Time preferences are measured by asking participants whether they prefer a certain amount now or a higher amount in 12 months. The question is repeated 5 times with varying amounts in order to refine the participant's preference. While

hypothetical, both the measures have been found to be correlated with behavioral measures with a “high explanatory power” (Falk et al., 2016:16). Moreover, I controlled for household income, employment, and age category.

## RESULTS

**Aggregated results.** In order to evaluate potential cleansing effects, I compare the differences in terms of decisions made by the participants across conditions. In other words, I am interested in knowing whether participants in the treatments chose different mixes of tables than the ones in the control group i) from the first to the second part, and ii) from the second to the third part. Overall, I therefore expect similar choices in the first part, decreases of different magnitudes in the second part for the low and high temptations treatments, and an increase in the third part only for the high temptation treatment (indicating moral cleansing).

Figure 6 provides an overview of participants choices in all three parts of our three treatments. The overall pattern of the data looks as expected. In part 1 where there is not yet a difference between treatments, the proportion of high effort is very similar in all three conditions. Thereafter the two temptation treatments show the expected patterns. In the low temptation condition, I observe an intermediate drop in the proportion of high effort when the participants face a slightly stronger incentive to provide low effort in part 2. This effect is not reversed in part 3 when incentives return to the initial level. In the high temptation condition where the incentive to provide low effort is stronger in part 2, I see a more pronounced drop in the provision of high effort. This effect is partially undone in part 3 when incentives are set back to their initial level. In the control condition there is only a small decrease in effort provision in part 2 and thereafter the effort level remains roughly constant.

-----

Insert Figure 6 about here

-----

I test the statistical significance of these observed patterns with a regression analysis reported in Table 1. I estimated a full model with the immediate and delayed impacts of the two treatments, that is, the interaction between the treatments and the different parts of the experiment (immediate: Low\*Part2 and High\*Part2; delayed: Low\*Part3 and High\*Part3). Moreover, I controlled for the single effects of the parts (Part2 and Part3), being allocated to a treatment (Low and High), as well as several demographics (Sex and Age) and control variables (Altruism and Patience).

The regression confirms that randomization worked so that there is no statistically significant difference in the baseline proportion of high effort provided in part 1 across conditions (control: 0.39, low temptation: 0.37, high temptation: 0.39).<sup>3</sup> The small decrease in the second part of the control condition (the proportion of high effort is 0.35) is not statistically significant ( $p = 0.274$ ). In the low temptation condition, in contrast, the proportion of high efforts in the low temptation treatment significantly decreased from part 1 (0.37) to part 2 (0.30,  $p = 0.038$ ), yet, as expected, there were no significant changes from part 2 to part 3 (0.29,  $p = 0.863$ ). This pattern is in line with Hypothesis 2 that an inconspicuous moral dissonance is likely to lead to persistent subsequent misbehavior. In the high temptation condition, I observe a large and significant decrease in the proportion of high effort from part 1 (0.39) to part 2 (0.23,  $p = 0.000$ ). However, different than in the low temptation condition, the proportion of high effort recovers at least partially in part 3. There is a significant increase from part 2 (0.23) to part 3

---

<sup>3</sup> control condition vs. the low temptation treatment ( $p = 0.657$ ), control condition vs. high temptation treatment ( $p = 0.967$ ), and low vs. high temptations treatments ( $p = 0.686$ ).

(0.33,  $p = 0.008$ ). This pattern provides support for Hypothesis one which predicts that a salient moral dissonance is likely to lead to moral cleansing.

However, although the within treatment analysis reported so far is supportive of Hypotheses 1 and 2, a clean test of my predictions requires a comparison of the between part differences across conditions. Such a diff-in-diff analysis established whether the patterns observed in the temptation treatments are statistically distinguishable from the time trend observed in the control condition.

The analysis reveals that participants in the low temptation treatment did not reduce their efforts from part 1 to part 2 significantly more than participants in the control condition ( $p=0.283$ ). This finding indicates that the magnitude of the additional misbehavior triggered by the low temptation was too small to be statistically distinguishable from the pure time trend in the control condition. Regarding the change between parts 2 and 3 the analysis confirms that participants in the low temptation treatment did not change their choices of efforts significantly more than participants in the control condition between part 2 and part 3 ( $p=0.941$ ). Finally, the choices of efforts in the low temptation treatment for part 2 and part 3 combined were also not significantly lower than the ones in the control condition ( $p=0.119$ ). The low temptation treatment did not, overall, significantly increase the proportion of misbehaviors.

Participants in the high temptation treatment, in contrast, reduced their efforts significantly more than participants in the low temptation treatment ( $p=0.004$ ) and participants in the control condition ( $p=0.000$ ). Moreover, participants in the high temptation treatment also increased their efforts from part 2 to part 3 by 0.09 (to 0.33), which is significantly more than participants in the low temptation treatment ( $p=0.000$ ) and participants in the control condition ( $p=0.000$ ). This final proportion in the high temptation condition is not significantly different from the

ones in the control condition (0.34,  $p = 0.770$ ) nor the low temptation treatment (0.29,  $p = 0.400$ ).

To summarize, individuals who were exposed to a large temptation displayed a morally inconsistent behavioral pattern, by misbehaving in the second part but not in the third part. This behavioral pattern is consistent with moral cleansing, thus confirming hypothesis 1. Moreover, this behavioral pattern remains after introducing controls for sex, age, patience and altruism (see Table 1). While the behavioral pattern of individuals in the low temptation treatment is consistent with Hypothesis 2, the evidence in support of the hypothesis is weak, because the low temptation treatment did not yield results that were significantly different from the control condition. This lack of significant differences across treatments may indicate that the size effects are relatively low, or that there is heterogeneity in the individual reactions to the manipulations. 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, I conducted further analyses to identify types in order to identify different reactions to the magnitudes of temptation. This analysis allows to refine the investigation of the heterogeneity of moral identities.

-----  
Insert Table 1 about here  
-----

**Type identification.** As the aggregated results may hide different behavioral pattern, and there is evidence for heterogeneity of reactions to moral deviations, I apply a finite mixture regression in order to identify different types of individuals within the sample, using the method described in Bruhin, Fehr-Duda, and Epper (2010). Finite mixture models (FMM) are typically used to investigate latent heterogeneity, that is, they assume that an unobserved variable affects

individual reactions in a systematic way. FMM essentially assign individuals to different classes based on this unobserved individual characteristic and allows to measure the respective effects of our treatments depending on the different types of individuals (e.g., see Bruhin et al., 2015). However, in order to identify types an FMM requires assumptions regarding i) the number of types in the sample, ii) the variables that are type-specific and iii) the variables that are constant across types. While it is possible to try different numbers of types and compare the resulting models, it is necessary to specify ex-ante which variables from Table 1 will vary across types, while the other variables remain common.

I henceforth assume that individuals may differ in terms of initial moral behavior, reactions to temptation, and compensation after being tempted. I expect individual heterogeneity in terms of reactions to the treatments for two reasons. First, individuals who do not provide any – or very little – high effort in the first part of the experiment cannot be pushed to provide less, due to the experimental design. Second, in line with Tetlock et al. (2000) I expect individuals with a very high base morality to have a higher awareness of their moral deviations and therefore to be less prone to misbehaviors. On the contrary, I assume that the effects of being randomly attributed to one condition or another in the first part and being in the second or third parts are constant across types.

In order to select a model, I looked at four criteria: the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the normalized entropy criterion (NEC), and the integrated completed likelihood criterion (ICL). While AIC and BIC are measures of goodness of fit (although not without shortcomings, see Biernacki, Celeux, & Govaert, 2000; Celeux & Soromenho, 1996), NEC and ICL are indicators of how well the FMM can distinguish between types. I therefore ran finite mixture models with the number of types ( $K$ ) ranging from two to five and compared them based on the four indicators. Without surprise, both measures of fit improve as the number of types increases (see Table 2). Likewise, the

identification criteria improve with the number of types, with the ICL decreasing and the NEC equal to zero, indicating a high quality of classification (Biernacki, Celeux, & Govaert, 1999). For this reason, I decided to discard the model with only one or two types. Moreover, the FMM attributed individuals to types with a posterior probability that is either close to zero or one for both  $K = 3$  and  $K = 4$ , while for  $K = 5$  some individuals remain between two types (see Figure 7). Although the model with  $K = 4$  yields a better fit, it may not add much to the analysis to distinguish between four types of individuals compared to three. In fact, a model with three types allows to distinguish between low, moderate, and high initial moral behaviors, while a model with four types would constrain one of the three categories to be separated in two. This can also lead to a higher number of misattributions to types for the model with  $K = 4$  than  $K = 3$ , as shown in Figure 7. Consequently, I retained the FMM with three types for the final analysis of type-based reactions.

-----

Insert Table 2 about here

-----

-----

Insert Figure 7 about here

-----

In the selected FMM with  $K=3$ , Type 1 individuals constitute 57.3% of the sample, Type 2 individuals amount to 22.9% of the sample, and Type 3 individuals represent 19.7% of the sample (Table 3).

-----

Insert Table 3 about here

-----  
-----  
Insert Figure 8 about here  
-----

The different behavioral patterns per condition identified by the finite mixture model are summarized in Figure 8. The differences between types, both in term of base effort provision and in terms reactions to the experimental treatments are summarized in Table 4.

**Type 1.** Type 1 individuals display a low initial moral identity, as they rarely provide high efforts. They are therefore expected not to react to temptations as they already are close to the lowest morality allowed by the experimental design. Overall, they are not expected to show any increase or decrease throughout the experimental parts. Type 1 individuals provided an average initial proportion of high effort equal to 0.0548 across conditions. Although Type 1 participants in the low temptation treatment reduced their efforts significantly more than participants in the control condition ( $p=0.022$ ), the effect sizes are very small. For instance, participants in the control condition reduced their efforts by 0.04 from part 1 to part 2. Likewise, Type 1 participants in the high temptation treatment did not decrease their efforts significantly more than participants in the low temptation treatment ( $p=0.437$ ) and participants in the control condition ( $p=0.135$ ). Moreover, Type 1 participants in the low temptation treatment did not increase their efforts from part 2 to part 3 significantly more than participants in the control condition ( $p=0.591$ ). Although Type 1 participants in the high temptation treatment increased their efforts significantly more in the last part than participants in the low temptation treatment ( $p=0.016$ ) and participants in the control condition ( $p=0.028$ ), the

increase is only of 0.04 and therefore small. In conclusion, Type 1 participants initially display extremely low efforts, which essentially leaves no room for relevant reactions to the treatments.

**Type 2.** Type 2 individuals initially provide almost only high efforts, displaying a really high moral identity. These individuals are therefore expected to be sensitive to temptation only if the temptation is quite strong. I expect Type 2 individuals to be prone to moral cleansing behaviors, because they will mostly experience salient moral dissonances. Type 2 individuals provided an initial proportion of high efforts of 0.95, averaged across conditions. Type 2 participants in the low temptation treatment did not reduce their efforts from part 1 to part 2 significantly more than participants in the control condition ( $p=0.847$ ). However, Type 2 participants reduced their efforts significantly more than both participants in the low temptation treatment ( $p=0.002$ ) and participants in the control condition ( $p=0.001$ ). Moreover, although there were no significantly different changes of choices of efforts from part 2 to part 3 between Type 2 participants in the low temptation treatment and participants in the control group ( $p=0.755$ ), participants in the high temptation treatment increased their proportion of high efforts by 0.20. This increase is significantly higher than the changes observed in the low temptation treatment ( $p=0.001$ ) and in the control condition ( $p=0.001$ ). However, 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 ( $p = 0.798$ ). As for type 1, type 2 individuals were not affected by the low temptation treatment and thus only provide support for Hypothesis 1.

**Type 3.** Type 3 individuals display a mixed moral identity. They are expected to react to temptation more than Type 2 individuals as they do not have a strong initial moral identity. Type 3 individuals are therefore expected to drive the results described in Hypotheses 1 and 2. In addition, they fail to cleanse in the low temptation treatment, despite their misbehaviors. Type 3 individuals provided an initial proportion of high efforts amounting to 0.69 across

conditions. Contrary to the previous two types, Type 3 participants in the low temptation treatment decreased their efforts significantly more than participants in the control condition ( $p=0.006$ ) from part 1 to part 2. Similarly, Type 3 participants in the high temptation treatment also decreased their efforts significantly more than participants in the control condition ( $p=0.000$ ). However, the decrease of 0.14 observed in the low temptation treatment was not significantly different from the decrease of 0.55 in the high temptation treatment ( $p=0.197$ ). Despite the lack of significant difference between the two treatment effects, two different reactions were observed between part 2 and part 3. While Type 3 participants in the low temptation treatment did not change their proportion of high efforts significantly differently than participants in the control condition ( $p=0.783$ ), participants in the high temptation treatment increased their proportion of high efforts by 0.12. This increase is significantly higher than the changes observed for participants in the low temptation treatment ( $p=0.047$ ) and participants in the control condition ( $p=0.021$ ). However, the final proportion of high efforts in the high temptation treatment, summing to 0.26, 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, I find support for both Hypothesis 1 and Hypothesis 2 within type 3 individuals.

-----  
Insert Table 4 about here  
-----

To summarize, I found support for Hypothesis 1 in Type 2 and Type 3 individuals, as both showed behavioral patterns that are in line with moral cleansing. Moreover, Hypothesis 2 is supported for Type 3 individuals, as they failed to cleanse after a smaller decrease of morality. I can therefore conclude that moral cleansing seems to be at least partially triggered by the

magnitude of a prior moral deviation, but also that moral deviations depend on the initial morality of individuals. However, as the majority of the sample is composed of Type 1 individuals whose initial morality is low, the average effects are weakened when looking at the aggregated results. The FMM therefore allowed to capture heterogeneity of i) initial levels of efforts, and ii) treatment reactions.

**Distortion of beliefs.** Part of the literature on moral behaviors suggests that one explanation for misbehaving relies in distorted beliefs about one's morality (Chugh & Bazerman, 2007; Chugh, Bazerman, & Banaji, 2005). More specifically, distorted beliefs may prevent individuals from seeing their own moral transgressions, which in turn prohibits cleansing mechanisms from being triggered. To investigate the accuracy of individual beliefs, I replicated the experiment in a hypothetical setting (n=402) on MTurk. Participants were randomly allocated to one of the three conditions of the initial study, before being presented the same instructions as the initial study. However, they were asked to predict how many small tables the participants in the initial study had chosen. Each part of the initial experiment was presented in the same order, to recreate similar conditions. For each part, participants were given 10 ranges to choose from and instructed them that choosing the right range would grant them a bonus of USD 0.5 per correct prediction (in addition to a fixed wage of USD 0.5). In a second step, participants were asked to indicate, for each of the three parts of the initial study, how many small tables they would have personally counted. While the first step is an incentivized task aimed at assessing their beliefs about other individuals' behaviors, the second one only assesses their own intentions and is therefore not incentivized (besides the fixed wage they receive for the completion of the task).

-----

Insert Figure 9 about here

-----

I operationalized the beliefs about one's moral identity by aggregating the hypothetical decisions across conditions for the first part of the experiment. The beliefs about others were operationalized in the same manner. Participants reported they would count 7.26 large tables out of 10, while they believe others would count 5.62 large tables. In line with past research (e.g., Banaji et al., 2003; Chugh et al., 2005; Tenbrunsel, 1998) and Hypothesis 3a, participants overestimated their moral identity compared with their beliefs about others ( $p = 0.0000$ ). Moreover, the average number of large tables counted in the first part of the initial study (3.85) was significantly smaller than both the participants' beliefs about themselves ( $p = 0.0000$ ) and others ( $p = 0.0000$ ).

-----

Insert Figure 10 about here

-----

Sensitivity beliefs were operationalized by calculating the relative difference between the beliefs about the small tables counted in the first and the second parts of the study, that is, the estimated reactions to the treatments divided by the estimated behaviors in the first part. Participants in the control condition did not report any significant relative change between the first and the second parts of the study for themselves ( $p = 0.1322$ ) nor others ( $p = 0.3141$ ). While the two beliefs are not significantly different from one another ( $p = 0.9089$ ), both the beliefs about themselves and the beliefs about others are significantly smaller than the actual relative change of behavior that occurred in the initial study ( $p = 0.0066$  and  $p = 0.0054$  respectively). Similarly, participants in the low temptation condition reported no significant sensitivity for themselves ( $p = 0.5442$ ), whereas they believe others are sensitive to low temptations ( $p = 0.0074$ ). However, participants underestimated both their own sensitivity ( $p$

= 0.0031) and the sensitivity of others ( $p = 0.0170$ ) compared to the sensitivity from the initial study. Participants in the high temptation reported significant sensitivities for themselves ( $p = 0.003$ ) as well as others ( $p = 0.0011$ ), although both are significantly underestimated compared to the actual behavioral reactions to the high temptation treatment ( $p = 0.0001$  and  $p = 0.0002$  respectively). However, there were no significant differences between the sensitivities reported for themselves and others ( $p = 0.6874$ ), contradicting Hypothesis 3b. Finally, participants reported no significant long-lasting effects of the treatments on the third part of the experiment, contradicting the behavioral data.

## **GENERAL DISCUSSION**

In this paper, I investigated the impact of moral deviations of different magnitudes on the likelihood of subsequent misbehaviors. The present experimental evidence supports my assumption that large moral deviations increase the likelihood of moral cleansing. However, my analysis of the individual heterogeneity shows that both the sensitivity to temptations and the extent to which individuals cleansed depended on their initial level of effort provision. My results are pioneer in that I explain different behavioral patterns through a concept previously overlooked: the magnitude of moral deviations. In line with my expectations, large moral deviations lead to lower likelihoods of subsequent misbehaviors, while small moral deviations increase the likelihoods thereof, at least in part of my sample. Theoretically, I explain this mechanism by arguing that large deviations are salient and trigger a reaction, while small deviations remain inconspicuous and may be taken as a signal from the moral self. Finally, my results support the framework of bounded ethicality, in that I found that individuals systematically overestimate their moral identity, both compared to their beliefs about others' moral identity and compared to their own behavior. However, my analysis of the beliefs about sensitivity to temptations showed that individuals only underestimate their sensitivity compared to their actual behaviors, but do not believe they are less sensitive than others.

## **Theoretical and Practical Implications**

The present work makes several important contributions to the behavioral ethics literature. First, this paper advances research on moral behaviors by providing a behavioral measure of cleansing behaviors. While past research has often neglected the moral identity of individuals in their study of cleansing effects, I have focused on capturing every stages of misbehaviors: the past, present, and future. Moreover, my analysis of types shows that individuals' reactions to temptations, that is whether they deviate from their identity under tempting pressures, depends on their identity itself. While very moral individuals resisted low temptations, individuals with a lower initial morality succumbed to them. Moreover, I have found that individuals with a high moral identity cleanse fully once the temptations are removed, so as to recover from temptation to regain their initial moral behavior. On the other hand, individuals who were less moral had a decreasing morality over time. These individuals consequently do not recover their initial moral behavior, but only cleanse partially. This is especially important when considering whether some individuals are more prone to corruption than others, as tempting situations are consequently more of a problem for individuals who do not have strong moral identities.

Second, I depart from past literature by including a measure of morality prior to the introduction of my treatments. As I showed that (for some individuals) cleansing behaviors are triggered by large departure from one's moral identity, it is crucial to capture this initial level of morality, which serves as a baseline. This baseline can then be compared to i) the behavior once a treatment is introduced, and ii) the final behavior after the treatment is removed.

Third, I investigated individual beliefs not only about moral identity (i.e., behavior), but also about sensitivity to temptation (i.e., reaction to the treatments). While the literature often assumes that individuals overestimate their moral identity and underestimate their

sensitivity to temptation, both compared to other and to reality, I found that this is only partially true for the beliefs about one's sensitivity to temptation. Although individuals did underestimate their sensitivity compared to the actual behaviors, they did not believe being less sensitive than others. These results emphasize the importance to distinguish between beliefs about moral identities and beliefs about sensitivities to temptations and calls for further exploration of the latter.

### **Limitations and future directions**

My research is not without shortcomings. While I find general support for Hypothesis 1 and partial support for Hypothesis 2, only 42.6% of my sample are affected. Indeed, the majority of the sample chose to provide low efforts and show little concern for the charity. I speculate this may be due to the pool of participants, as workers on Amazon Mechanical Turk primarily aim at maximizing their hourly wage<sup>4</sup>. Nevertheless, my study is rather conservative in that I tried to capture morality phenomena in a context that is not favorable to moral behaviors.

While I focused on the effect of moral deviations on subsequent misbehaviors, I did not include moral licensing. Although I acknowledge its importance in understanding the mechanisms behind cleansing and licensing behaviors, I focused on reactions to misbehaviors.

---

<sup>4</sup> Several participants mentioned this in the comment they left at the end of the experiment. Here is a selection of these comments:

“I need to protect any money I earn at this point, I am poor”

“MTurk used to offset my income, but since COVID-19, it has temporarily become my only income since I'm still awaiting unemployment benefits to kick in.”

“i decided not to donate because i need the funds”

Likewise, several participants doubted the veracity of the impact on the charity:

“I am never sure in studies like these whether donations are real. I prefer to make mine knowing what's actually happening.”

“I just wanted to finish the task quickly. I don't think you're actually going to decrease your donation (if you're going to make one at all). This was a psychological experiment to see if people would help or hurt a donation. (This is my guess.)”

However, I hope to see more research including continuous behavioral measures also in the investigation of moral licensing. However, I expect licensing effects to be of smaller magnitudes, as individuals tend to display stronger reactions to events with negative valences, such as misbehaviors, than positive ones (Kahneman & Tversky, 1979).

Finally, my investigation of individual reactions to misbehaviors could open the door to different types of interventions aiming at reducing misbehaviors. Such interventions could, for example, focus on improving the moral identity of individuals, so as to reduce their sensitivity to low temptations and trigger full cleansing after being exposed to high temptations. Moreover, I hope to see future research investigating how the awareness of temptations could help reduce the sensitivity to these temptations.

## **Conclusion**

This work offers a promising beginning for the study of moral behaviors. My results demonstrate how the likelihood of subsequent misbehaviors is affected by the magnitude of prior misbehaviors, providing evidence for the framework of bounded ethicality. While small misbehaviors increase the likelihood of subsequent misbehaviors for individuals with moderate moral identities, large misbehaviors decrease leading to moral cleansing for most individuals.

## 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.
- Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423.
- Ariely, D., & Norton, M. I. (2008). How actions create—not just reveal—preferences. *Trends in Cognitive Sciences*, *12*(1), 13-16.
- Banaji, M. R., Bazerman, M. H., & Chugh, D. (2003). How {Un} ethical Are. *Harvard Business Review*.
- Banaji, M. R., & Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. *Memory, brain, and belief*, 139-175.
- Bandura, A. (1990). Mechanisms of moral disengagement in terrorism. *Origins of terrorism: Psychologies, ideologies, states of mind*, 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.
- Biernacki, C., Celeux, G., & Govaert, G. (1999). An improvement of the NEC criterion for assessing the number of clusters in a mixture model. *Pattern Recognition Letters*, 20(3), 267-272.
- Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7), 719-725.
- Bruhin, A., Fehr-Duda, H., & Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78(4), 1375-1412.
- Bruhin, A., Goette, L., Roethlisberger, A., Markovic, A., Buchli, R., & Frey, B. M. (2015). Call of duty: the effects of phone calls on blood donor motivation. *Transfusion*, 55(11), 2645-2652.
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2), 195-212.
- 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.

- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88-97.
- 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.
- Desai, S. D., & Kouchaki, M. (2017). Moral symbols: A necklace of garlic against unethical requests. *Academy of Management Journal*, 60(1), 7-28.
- 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.
- Dowie, M. (1977). How Ford put two million firetraps on wheels. *Business and Society Review*, 23(3), 46-55.
- Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of Experimental Social Psychology*, 45(3), 590-593.
- Falk, A., Becker, A., Dohmen, T. J., Huffman, D., & Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Working Paper*.
- 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.
- Gatewood, E., & Carroll, A. B. (1981). The anatomy of corporate social response: The Rely, Firestone 500, and Pinto cases. *Business Horizons*, 24(5), 9-16.

- Gioia, D. A. (1992). Pinto fires and personal ethics: A script analysis of missed opportunities. *Journal of Business Ethics, 11*(5), 379-389.
- Gneezy, U., Imas, A., & Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science, 60*(11), 2645-2658.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 263-291*.
- Klotz, A. C., & Bolino, M. C. (2013). Citizenship and counterproductive work behavior: A moral licensing view. *Academy of Management Review, 38*(2), 292-306.
- Mayer, D. M., Aquino, K., Greenbaum, R. L., & Kuenzi, M. (2012). Who displays ethical leadership, and why does it matter? An examination of antecedents and consequences of ethical leadership. *Academy of Management Journal, 55*(1), 151-171.
- Messick, D. M., & Bazerman, M. H. (1996). Ethical leadership and the psychology of decision making. *MIT Sloan Management Review, 37*(2), 9.
- 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.
- 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.
- Reynolds, S. J., & Ceranic, T. L. (2007). The effects of moral judgment and moral identity on moral behavior: an empirical examination of the moral individual. *Journal of Applied Psychology, 92*(6), 1610.
- 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.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571-610.
- 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.
- Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, 17(2), 223-236.
- 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.
- Tost, L. P. (2011). An integrative model of legitimacy judgments. *Academy of Management Review*, 36(4), 686-710.
- Zhang, T., Fletcher, P. O., Gino, F., & Bazerman, M. H. (2015). Reducing bounded ethicality: How to help individuals notice and avoid unethical behavior. *Organizational Dynamics*.
- Zhong, C.-B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, 313(5792), 1451-1452.



## Appendix A

FIGURE 1. Theoretical Model for the Link between Prior and Subsequent Misbehaviors, Moderated by the Saliency of the Moral Dissonance

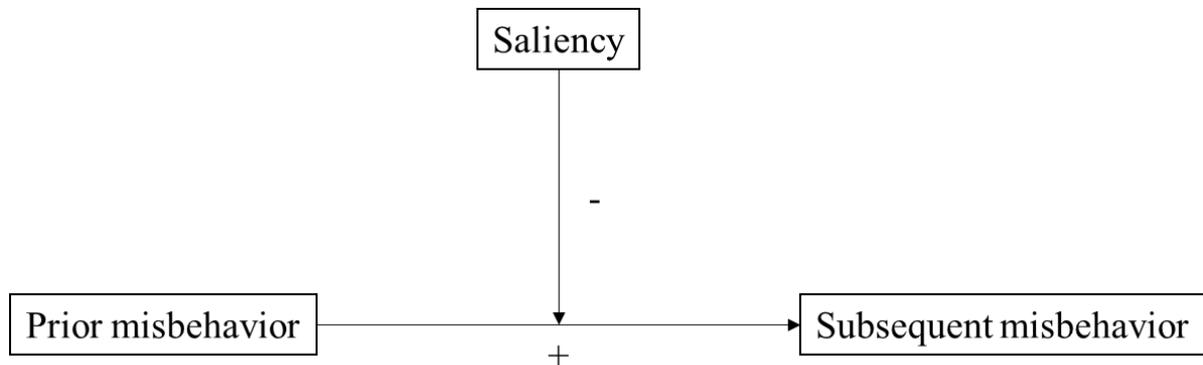


FIGURE 2. Choice between the small and large tables as displayed on the participants' screen in the first and third parts.

### 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	0	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	1	0	0	0	1	0
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

FIGURE 3. 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														Table B: increase of the donation by \$0.1													
1	0	1	0	0	0	0	0	1	0	0	1	1	1	0	1	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	0	1	1	0	1	1				
0	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1				
1	0	0	0	1	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1	0	1	1	0				
1	1	0	1	1	1	0	0	1	0	0	1	1	1	1	0	1	1	1	0	1	1	1	0				
1	0	0	1	0	1	0	1	0	1	1	0	1	1	0	1	0	0	1	1	1	0	0	0				
1	0	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	0	1	1	1	1	1	1				
0	1	1	0	0	1	1	1	0	1	1	0	1	1	0	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	1	1	0	1	1	0	1				
1	0	0	0	1	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1	0	1	1	0				
1	1	0	1	1	1	0	0	1	0	0	1	1	1	0	1	1	0	1	1	1	1	1	0				
1	0	0	1	0	1	0	1	0	1	1	0	1	1	0	1	0	0	1	1	0	1	0	0				
1	1	0	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	0	1				
0	1	1	0	0	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	0	1				

How many ones are in the table you chose?

Next

FIGURE 4. Choice between the small and large tables as displayed on the participants' screen in the second part (low temptation condition).

## Part 2/3, decision 1/10

Choose a table.

Table A: decrease of the donation by \$0.01														Table B: increase of the donation by \$0.1													
1	0	1	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	1	1	1	0	1	0				
0	1	1	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	0	1	1	1	1	1				
0	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1				
1	0	0	0	1	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1	0	1	1	0				
1	1	0	1	1	1	0	0	1	0	0	1	1	1	1	0	1	1	0	1	1	1	1	0				
1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1	0	1	1	0	1	0	0				
1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	1	1	1	0				
0	1	1	0	0	1	1	1	0	1	1	0	1	1	0	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	1	1	0	1	1	0	1				
1	0	0	0	1	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1	0	1	1	0				
1	1	0	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	1	1				
0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	1	1	0	0	0	1				
0	1	1	0	0	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	0	1				

How many ones are in the table you chose?

Next

FIGURE 5. Choice between the small and large tables as displayed on the participants' screen in the second part (high temptation 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

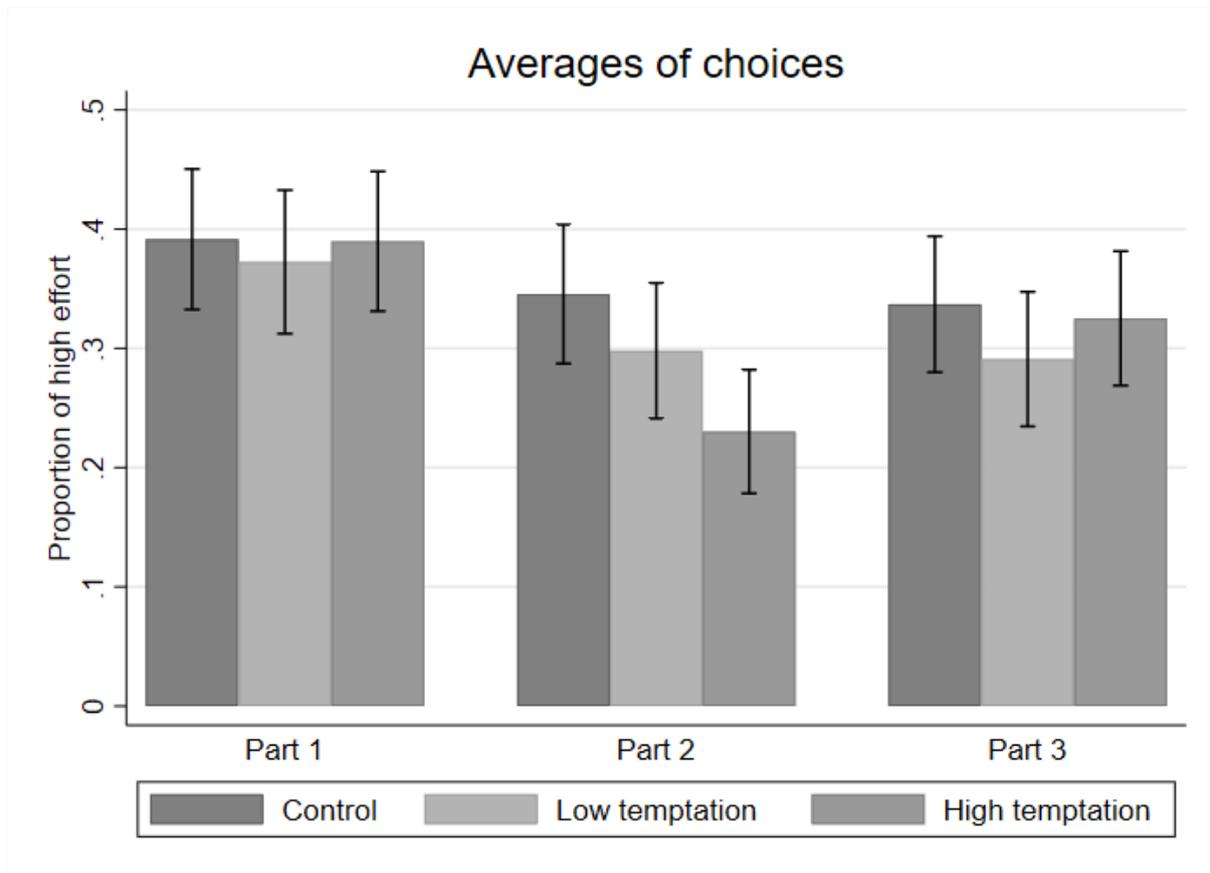
**Table B: increase of the donation by \$0.1**

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?

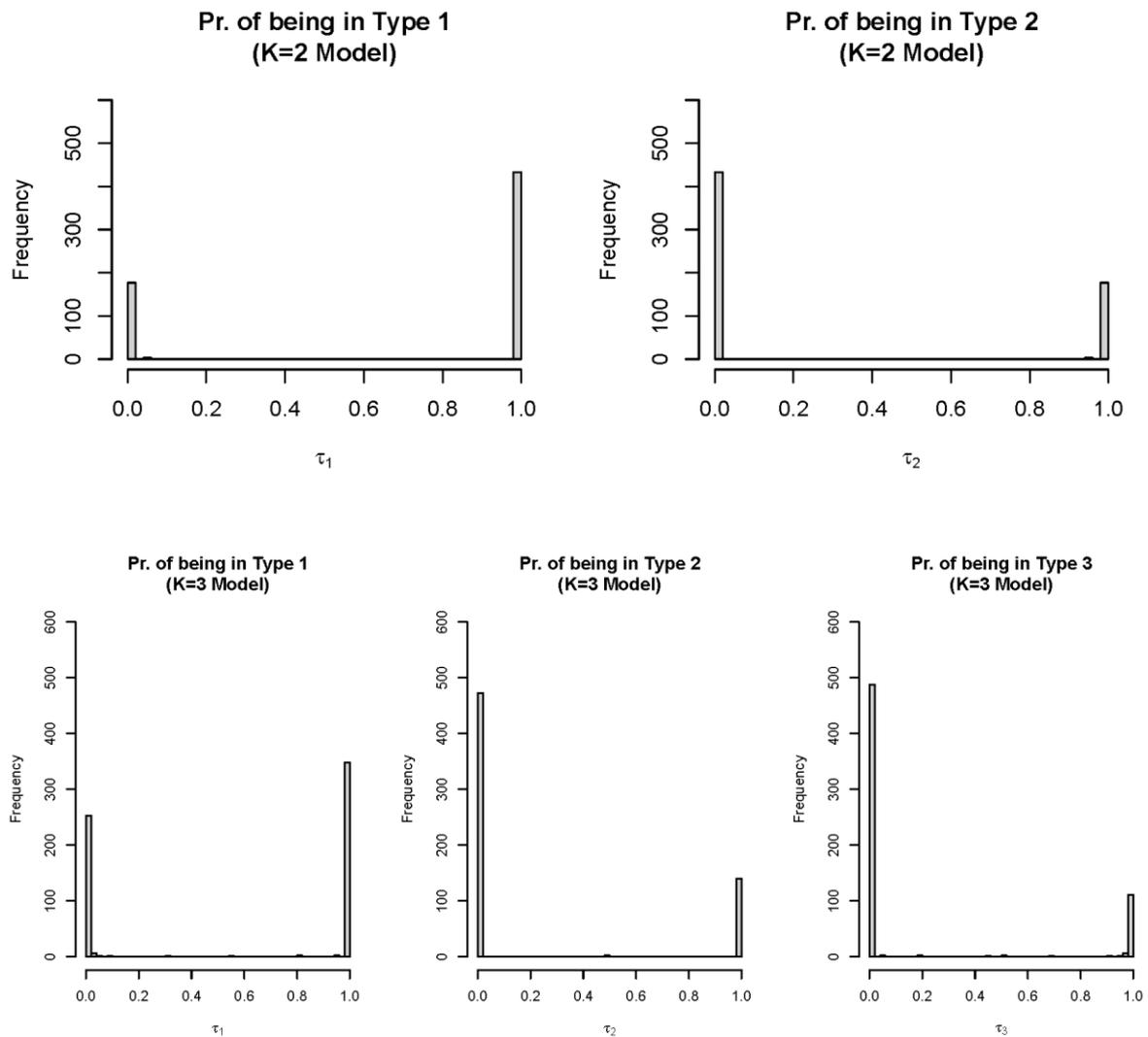
Next

FIGURE 6. Choices of high efforts per condition



Notes: Average proportion of high effort (large tables) counted by the participants across parts, per treatment. The error bars represent plus/minus one standard deviation.

FIGURE 7. Distributions of the posterior probabilities of belonging to types, based on the results of the Finite Mixture Models



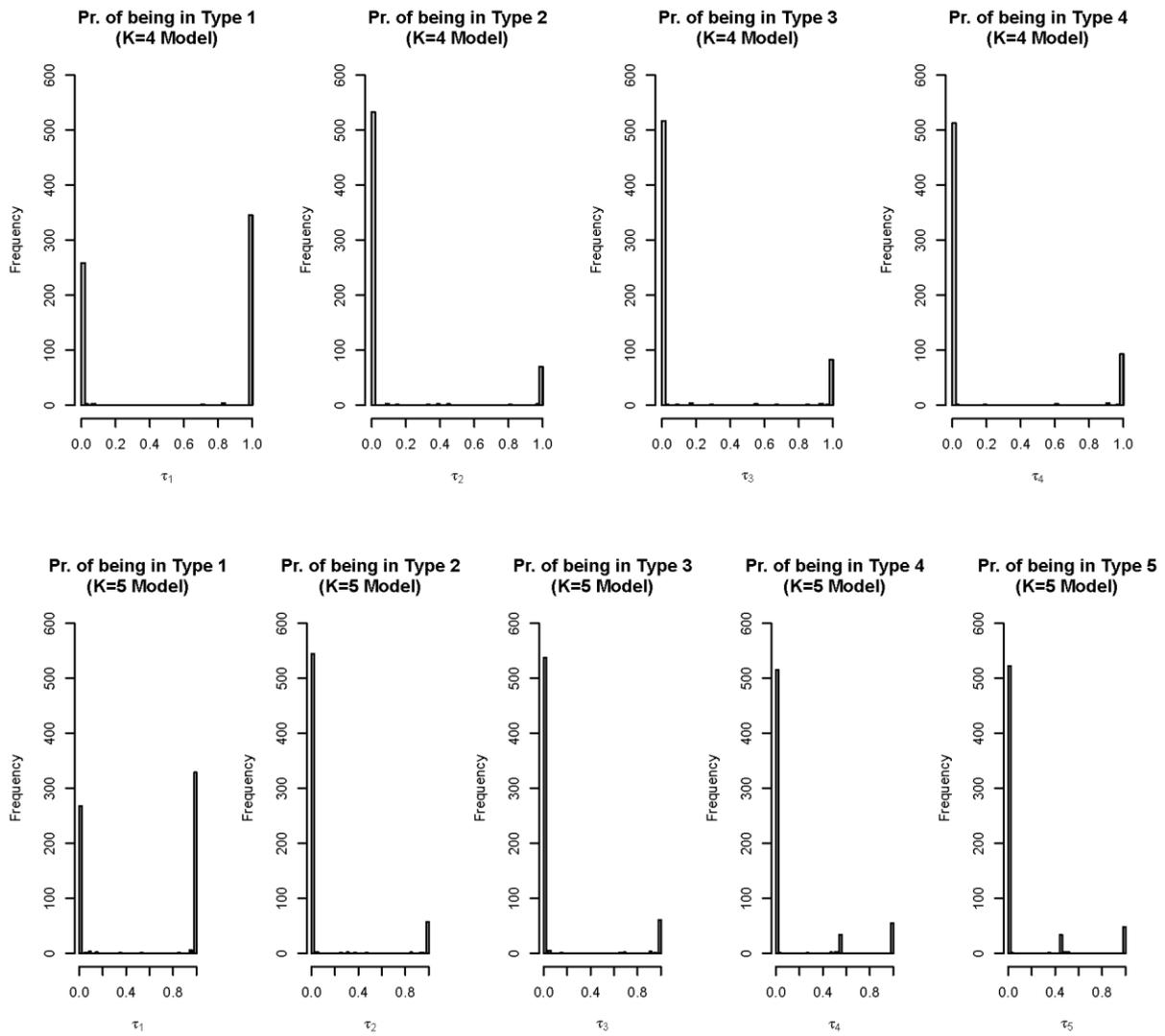
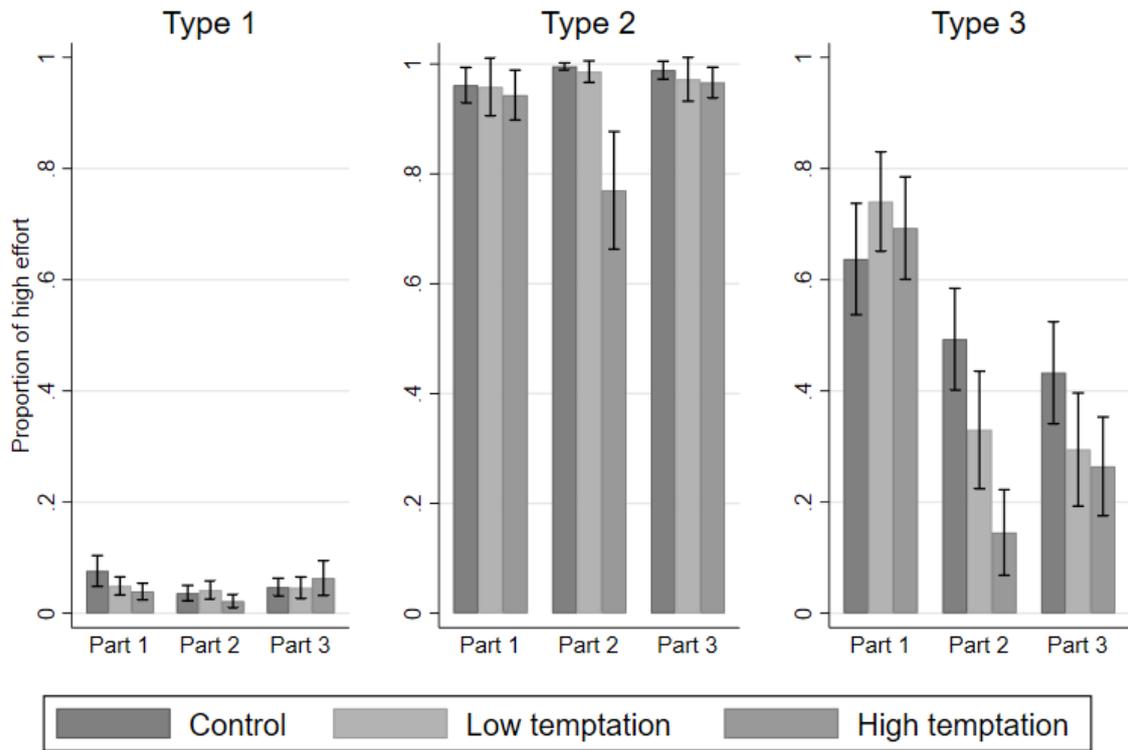


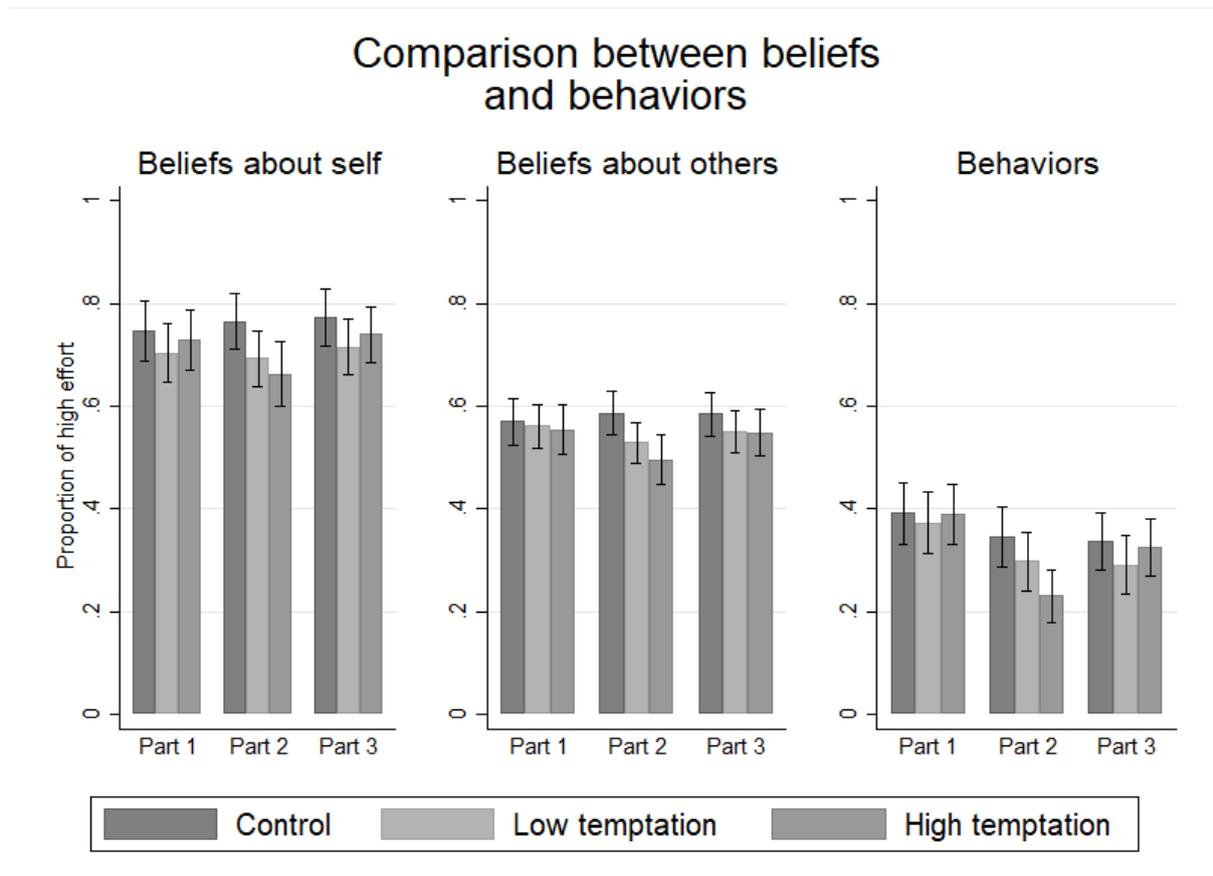
FIGURE 8. Choices of high efforts of Type 1 individuals per condition (averaged by part)

## Behavioral patterns identified



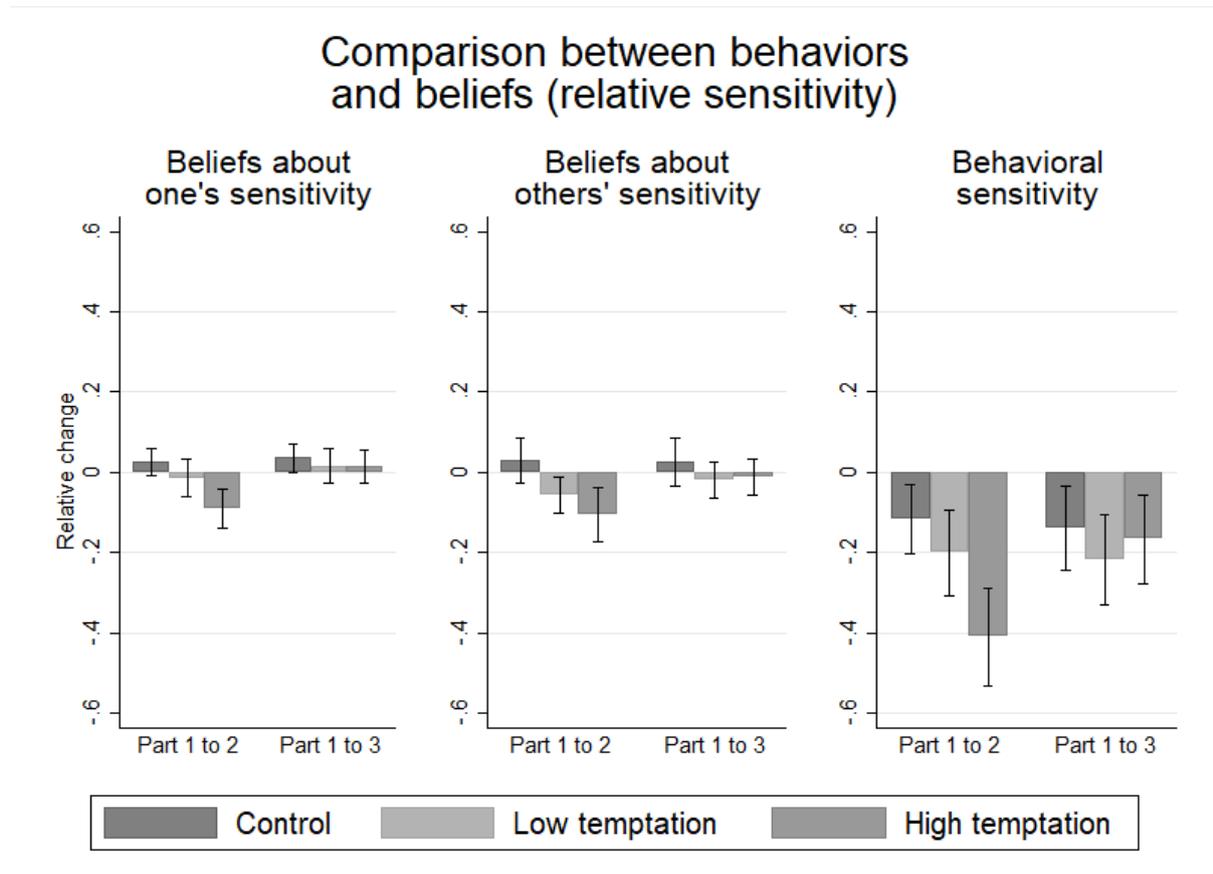
Notes: Average proportion of high effort (large tables) counted by the participants across parts, per treatment, for each of the three identified types of individuals. The error bars represent plus/minus one standard deviation.

FIGURE 9. Comparison between actual choices of efforts and beliefs about self and others



Notes: On the left, the bars represent the average proportions of high effort (large tables) that participants reported they would count across parts, per treatment. In the middle, the bars represent the average proportions of high effort that participants reported others would count across parts, per treatment. On the right, the bars represent the average proportions of high efforts that participants actually counted in the behavioral experiment, across parts, per treatment. For all, the error bars represent plus/minus one standard deviation.

FIGURE 10. Comparison between actual treatment effects (immediate and delayed) on efforts and beliefs about self and others



Notes: On the left, the bars represent the average changes in the proportions of high effort provided (large tables counted) that participants reported they would count. In the middle, the bars represent the average proportions of high effort that participants reported others would count across parts, per treatment. On the right, the bars represent the average proportions of high efforts that participants actually counted in the behavioral experiment, across parts, per treatment. For all, the changes from part 1 to part 2 represent the treatment effect, that is, the reaction to temptations. The changes from part 1 to part 3 represent the potential long-lasting effects of the treatments. The error bars represent plus/minus one standard deviation.

## Appendix B

TABLE 1. Mean immediate and delayed treatment effects of the choices of high efforts

	Model 1	Model 2	Model 3	Model 4
--	---------	---------	---------	---------

Constant	0.392 ***	0.283 ***	0.222 ***	0.227 ***
	(0.010)	(0.026)	(0.027)	(0.027)
Low	-0.019	-0.010	-0.019	-0.019
	(0.015)	(0.018)	(0.018)	(0.018)
High	-0.002	0.006	-0.0003	-0.001
	(0.015)	(0.018)	(0.017)	(0.017)
Part2	-0.046 **	-0.048 **	-0.048 **	-0.048 **
	(0.015)	(0.017)	(0.017)	(0.017)
Low*Part2	-0.029	-0.028	-0.028	-0.028
	(0.021)	(0.025)	(0.025)	(0.025)
High*Part2	-0.114 ***	-0.109 ***	-0.109 ***	-0.109 ***
	(0.021)	(0.025)	(0.025)	(0.025)
Part3	-0.054 ***	-0.041 *	-0.041 *	-0.041 *
	(0.015)	(0.017)	(0.017)	(0.017)
Low*Part3	-0.027	-0.027	-0.027	-0.027
	(0.021)	(0.025)	(0.025)	(0.025)
High*Part3	-0.010	0.0003	0.0003	0.0003
	(0.021)	(0.025)	(0.025)	(0.025)
Sex		-0.08 ***	-0.0784 ***	-0.077 ***

		(0.009)	(0.009)	(0.009)
Age		0.026 ***	0.029 ***	0.029 ***
		(0.002)	(0.002)	(0.002)
Altruism			0.0003 ***	0.0003 ***
			(0.00002)	(0.00002)
Patience				-0.001
				(0.0004)
N	18450	12450	12450	12450
R2	0.011	0.033	0.047	0.047

---

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

TABLE 2. Model selection criteria

	K = 1	K = 2	K = 3	K = 4	K = 5
AIC	24348.61	10202.35	6744.55	5069.95	3832.57
BIC	24426.83	10327.52	6916.65	5288.99	4098.54
NEC	1.00	0.00	0.00	0.00	0.00
ICL	24426.83	10330.58	6928.84	5311.08	4175.65

TABLE 3. Proportions of Type 1 (low morality), Type 2 (high morality), and Type 3 (moderate morality) identified in the sample by the Finite Mixture Model

Type 1	Type 2	Type 3
--------	--------	--------

Proportion	0.573	0.229	0.197
SE	(0.020)	(0.017)	(0.016)

TABLE 4. Mean immediate and delayed treatment effects of the choices of high efforts for the three types identified by the Finite Mixture Model.

	Type 1	Type 2	Type 3
Type-specific			
Constant	0.071*** (0.015)	0.988*** (0.016)	0.615*** (0.026)
Low*Part2	0.012 (0.02)	0.04 (0.025)	-0.242*** (0.068)
High*Part2	0.011 (0.02)	-0.157** (0.057)	-0.408*** (0.056)
Low*Part3	-0.024 (0.025)	0.034 (0.033)	-0.267*** (0.068)
High*Part3	0.062* (0.028)	0.048† (0.028)	-0.284*** (0.067)
Common coefficients			
Low	0.004 (0.019)		
High	-0.015 (0.019)		
Part2	-0.046** (0.017)		

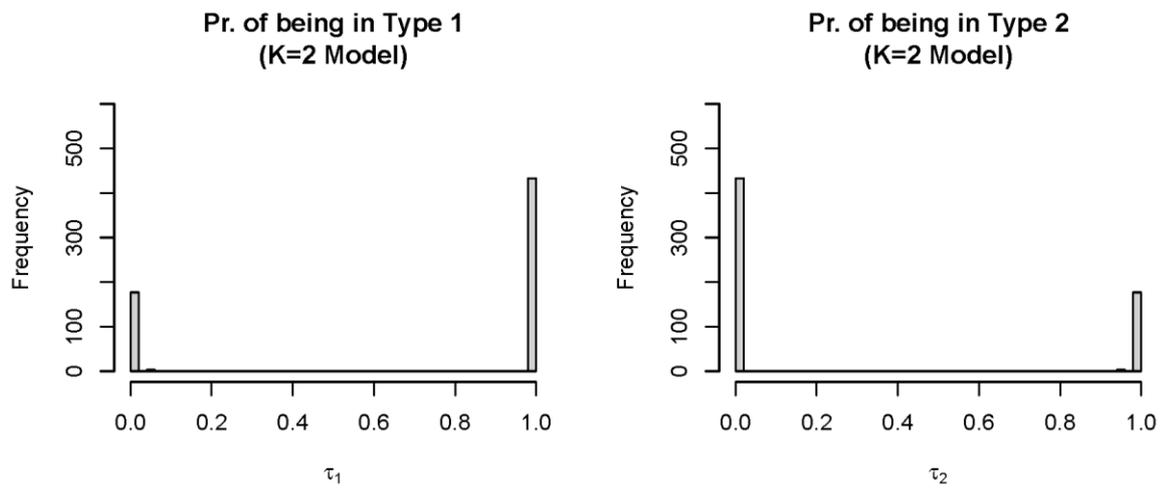
Part3	-0.054**
	(0.021)
Sigma	0.281
	(0.007)

---

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

### Appendix C

Additional resource: Distributions of the posterior probabilities of belonging to types, based on the results of the Finite Mixture Models for two types of individuals



## **Organizational Changes as Micro-level Processes: Old Phenomena, New Typology**

### **Abstract**

Organizational changes are typically categorized as either divergent or convergent. While divergent changes are characterized by an organizational reorientation, convergent changes build on the existing organizational form. Yet, several scholars have reported opposing dynamics for the same type of change, leading to a certain confusion regarding the current typology of organizational changes. In this paper, I suggest that the distinction between divergent and convergent changes is therefore not sufficient to categorize organizational changes. Building on the emerging literature on microfoundations, I describe organizational changes through the learning strategies of change recipients: the organizational members. Eventually, I draw from the interplay between an organization and its members to propose a different typology of organizational changes. This new typology can then serve future to identify patterns of changes, with distinct dynamics.

Keywords: Organizational change, typology, learning strategies, dynamics

*Matthieu Légeret*  
*University of Lausanne / Switzerland*  
*matthieu.legeret@unil.ch*

The first months of 2020 were heavily marked by the outbreak of the COVID-19 pandemic. As the coronavirus spread through the world, so did a set of new sanitary measures, often imposed by governmental regulations. Likewise, the needs of society at large evolved with the pandemic. These sudden changes in the environment of organizations not only had an impact on organizational processes, but also on organizational members. For instance, many organizations eventually resorted to remote work, and many employees discovered the benefits and costs of working from home. Inevitably, throughout its life, an organization will face changes that are significant enough to challenge its very core (Nadler & Tushman, 1989). These radical changes can be rapid and discontinuous (Romanelli & Tushman, 1994), or slow and incremental (Amis, Slack, & Hinings, 2004). Whether abrupt or incremental, successful radical changes not only impact organizations but also the individuals within the organizations, through a change of appropriateness of practices. For example, Greenwood, Suddaby, and Hinings (2002) focused on how radical change for accounting firms modified the role and practices of accountants within their firm. Traditionally, the literature distinguishes such radical or divergent changes, which are characterized by the reorientation of the organization (e.g., Greenwood & Hinings, 1996), from convergent changes. However, the denomination of “radical” or “divergent” (opposed to “convergent”) appears to be insufficient by itself, in that opposing dynamics have been observed for a single type of change. Moreover, although many scholars have focused on the perspective of individuals facing changes (e.g., Bartunek, 1984; Bartunek & Moch, 1987; Dunham, Grube, Gardner, Cummings, & Pierce, 1989; Lau & Woodman, 1995; Oreg, Bartunek, Lee, & Do, 2018), the interplay between change as an organizational phenomenon and the individual processes underlying it is often missing.

In this paper, I challenge the current typology of organizational change, and propose to distinguish changes based on (i) the organizational environment and on (ii) the decision strategies of the individuals within the organization. In particular, I consider how changes in

an organization's environment can require organizational members to adapt their learning strategies, but also how different strategies can be a source of change. A better understanding of individual processes not only helps comprehend why different dynamics can arise at the organizational level, but also how to influence those dynamics by acting on the individual strategies. Because I emphasize the interplay between an organization and its members, my paper follows recent interests in multi-level approaches (e.g., Chandler & Hwang, 2015; Powell & Colyvas, 2008; Powell & Rerup, 2017).

The main goal of this paper is to offer an alternative classification of organizational changes that yields consistent dynamics for a given type of change. Specifically, I distinguish between two types of changes – evolution and revolution – for which learning strategies are in adequacy with the organizational environment and two types changes – maladaptation and waste – for which the learning strategies are not aligned with the environment. In this respect, the present work makes three important contributions. First, a more precise typology of organizational change offers the opportunity to see new patterns emerge and can therefore strengthen the understanding of the macro-level phenomenon. Second, focusing on the microfoundations of organizational change—by integrating the underlying cognitive mechanisms of the organizational members in the face of change—provides a better understanding of the organizational phenomena itself (Schilke, 2018). Third, as I consider the interplay between the macro and the micro levels, I open the black box of individual processes (Powell and Rerup, 2017) and add psychological plausibility and therefore realism to the debate.

## **ORGANIZATIONAL CHANGES**

### **Typology and Dynamics of Change**

Change can take on many forms. Accordingly, it has been defined as “an empirical observation of difference in form, quality, or state over time in an organizational strategy, a program, a product, or the overall organization” (Van de Ven & Poole, 1995: 512). Whether a change of practice (e.g., Greenwood et al., 2002; Kostova & Roth, 2002; Townley, 2002) or a change of interpretations (e.g., Bartunek & Moch, 1987; Isabella, 1990; Lau & Woodman, 1995), change is a dynamic process that impacts the individuals within the organization in different ways and triggers different attitudes (e.g., see Lau & Woodman, 1995). Organizational changes are commonly separated in two categories: convergent and divergent changes (see Greenwood & Hinings, 2006, for an overview of the current state of the literature). Convergent changes are changes that ameliorate the current form of an organization. For example, a convergent change could be the implementation of practices that will increase the performance of individuals, while maintaining the previous goals and performance measures. Conversely, divergent changes are changes that alter the organizational form itself. Such radical changes are characterized by an organizational reorientation (Greenwood & Hinings, 1996; Greenwood & Hinings, 2006; Greenwood et al., 2002; Nadler & Tushman, 1989; Romanelli & Tushman, 1994; Tushman & Romanelli, 1985). Although most scholars agree on the distinction between convergent and divergent changes, different dynamics of change have been theorized and observed. On the one hand, Tushman and Romanelli (1985; see also Miller & Friesen, 1980, and Romanelli & Tushman, 1994) have argued that convergent changes take place incrementally over time to improve the current organizational form, while divergent changes are disruptive, rapid and discontinuous, and fundamentally alter the current state of an organization. On the other hand, Greenwood and Hinings (1996, 2006; see also Amis et al., 2004; Greenwood et al., 2002) have argued that divergent changes can also happen in an

incremental manner and take a long time to be implemented. Although these two streams of research seem to oppose each other, they both provide evidence supporting their theoretical accounts on the dynamics of organizational change. Moreover, both sides agree on one key element: the importance of the perspective of the individuals facing changes. For instance, while Anderson and Tushman (1990; see also Tushman & Rosenkopf, 1996; Virany, Tushman, & Romanelli, 1992) emphasize the inability of current executives to undergo radical changes, Greenwood and Hinings (1996, 2006) argue that groups of individuals who are unsatisfied with the status quo may be favourable to changes.

In my view, the current typology of organizational change, specifically, the distinction between convergent and divergent changes, is insufficient for two reasons. First, it focusses on consequences of the change itself, yet pays little attention to the source of the change. Second, and relatedly, it does not consider the underlying processes, that is, the perspective of organizational members. To address these shortcomings, I suggest grounding the typology of organizational change on characteristics of the organizations' environment and the individual processes that take place within the organizations. Consequently, I build on the common ground between these two streams of research by switching the scope of analysis from the impact of change on the organization to the interaction between the learning strategies and the organizational environment. Several scholars have already emphasized the importance of studying the mechanisms or dynamics of change to better understand the change processes (Dacin, Goodstein, & Scott, 2002; Greenwood et al., 2002; Hatch, 1993; Isabella, 1990; Tsoukas, 1989; Van de Ven & Poole, 1995). Some have focused on the perspective of the individuals facing organizational changes (Bartunek & Moch, 1987; Isabella, 1990; Lau & Woodman, 1995), while others have linked different levels to one another (Greenwood & Hinings, 1996), by considering the impact of groups on responses to change. Despite these efforts, approaches linking the micro-level processes, such as cognitive perspectives, to the

changes at the macro-level remain scarce. In line with this view, Powell and Rerup (2017) called for more research on the microfoundations of organizations, in an attempt to open the black box that is often used to represent individual processes.

## **MICROFOUNDATIONS OF ORGANIZATIONAL CHANGES**

Although the dynamics of change constitute a puzzle at the organizational level, it is often useful to look at how individual practices are impacted, that is, how organizational phenomena can affect the daily life of individuals within the organizations (e.g., Cardinale, 2018; Harmon, Green Jr, & Goodnight, 2015; Tracey, 2016). This approach, often called microfoundations, aims at inhabiting the macro-level events with a cognitive perspective (Schilke, 2018) or, more generally, individual viewpoints (Powell & Rerup, 2017). Arguably, looking at the microfoundations of a phenomenon allows for multi-level explanations and may hence offer a more accurate picture of the bilateral dynamics that animate the relationship between an organization and its individuals (Chandler & Hwang, 2015; Powell & Colyvas, 2008).

### **Individual and Social Learning Strategies**

To be accurately described, individual processes need to be contextualized. In his description of bounded rationality, Simon (1997) emphasized the importance of both the natural cognitive limitations of individuals and the environments for which their processes have been built. Boundedly rational individuals are assumed to hold limited, yet realistic capacities, that allow them to rely on simplified decision-making processes such as heuristics (see Gigerenzer & Gaissmaier, 2011; Gigerenzer & Selten, 2001; Gigerenzer & Todd, 1999; Luan, Reb, and Gigerenzer, 2019). Heuristics are decision strategies that rely on a limited amount of information to make fast, yet efficient decisions (Gigerenzer & Gaissmaier, 2011). Close to

Simon's work, a key aspect of heuristics is that they are adapted to their environment: they are ecologically rational (Todd, Gigerenzer, & the ABC Research Group, 2012).

Within organizations, some behaviors or processes are more appropriate than others, or more efficient. However, in many instances these behaviors do not emerge automatically, out of instinct, but instead have to be learnt. For example, a new organizational member may have to learn the rites and routines already in use in the organization, in order to fit their new environment. However, learning processes can take on different forms, typically described as individual and social learnings. Individual learning, or asocial learning, relies on trial-and-error, and requires engaging in costly searches for information regarding one's environment (Boyd & Richerson, 1985, 1995; Giraldeau, Valone, & Templeton, 2002; Rogers, 1988). Social learning, on the other hand, implies learning from other individuals, typically by copying their behaviors. Social learning therefore is less costly than individual learning and can be more adaptive (Henrich & Gil-White, 2001; Boyd & Richerson, 1985, 1988; Henrich & Boyd, 1998). Classic examples of social learning include imitation of random individuals (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981), imitation of successful individuals (Boyd & Richerson, 1985, 2002; Schlag, 1998, 1999; Henrich & Gil-White, 2001; Hertwig & Hoffrage, 2013), or a mix of different imitation strategies (Laland, 2004; McElreath et al., 2008; Whiten, Horner, Litchfield, & Marshall-Pescini, 2004). Arguably, imitation strategies fall into the category of heuristics, such as *imitate-the-successful*, or *imitate-the-majority* (see Hertwig & Hoffrage, 2013), as they are more frugal in terms of searching costs than first-hand experiences<sup>1</sup>.

The premise of this work is to consider organizational members as decision-makers. As such, individuals must be in adequacy with their environment in order to be more efficient: the

---

<sup>1</sup> However, it is worth noting that heuristics are not limited to routinized processes, as they can also be used for innovation (Kheirandish & Mousavi, 2018).

other organizational members, and, ultimately, the organization itself. Likewise, an organization must adapt to fit its environment. Coming back to the example of the pandemic, most—if not all—organizations were impacted such that they had to adapt to their new environment. For simplification purposes, I will adopt the perspective of the organizational members, and refer to both the organization and its environment as the members’ environment. Individuals hold a repertoire of practices that are deemed appropriate, or efficient, within their environment. Following the principle of ecological rationality, this set of practices should be adapted to their environment. Should the environment change, the organizational members may have to adapt, and in so doing, learn by resorting to individual or social learning strategies. A change in the environment, either at the organizational level or the institutional one, can change the set of practices deemed appropriate inside the organization. As a consequence, organizational learners may have to learn new routines or practices in order to remain efficient. However, the way they learn what new practices have a higher fit with the organization will impact the dynamics of the organizational change. For instance, if an organization were composed only of individual learners, then the process of changing practices by engaging in trial-and-error for each individual would be very costly, either in time or resource, and therefore inefficient. This interplay between learning strategies and their environment, the strategies and the appropriate practices, as well as these practices and change dynamics is summarized in Figure 1.

-----

Insert Figure 1 about here

-----

*Learning strategies as a Response to Change*

When a change in the environment affects the fitness of individuals' practices, for example by reducing their efficiency or appropriateness, then the practices may need to be adapted or dropped entirely to be replaced by new ones. In both instances, organizational members need to learn, either individually to innovate, or socially by copying another organizational member. If an organization is composed of a high proportion of individual learners, then the processes of (i) finding more adapted practices and (ii) harmonization of practices will be slow and costly (Boyd & Richerson 1985; Henrich & Boyd 1998). In contrast, social learning is more efficient in terms of costs, provided that individuals hold reliable cues as to whom they should copy (Boyd & Richerson 1985; Schlag, 1998, 1999; Henrich & Gil-White 2001). Moreover, if a change requires a higher cost of searching for alternatives, then social learning becomes more adaptive (Henrich & Gil-White, 2001).

#### *Learning Strategies as a Source of Change*

It is also possible for organizational practices to evolve over time without any relevant change taking place in the organization's environment. For example, practices can improve through time due to experience. This process, often referred to as first-order change in the organizational change literature (Bartunek & Moch, 1987), aims at deepening a pre-existing practice. However, this perfection of practices can only occur under two conditions (Galef, 1995; Laland, 2004). First, some of the organizational members must be individual learners, as they keep on exploring on their own and provide a source of innovation. Second, their practices need to be copied by social learners in order to propagate through the organization.

#### *The Ecological Rationality of Learning Strategies*

Just like practices and organizations fit their environment, the decision to learn individually or socially depends on what other organizational members decide to do. At the extreme, if an organization were entirely composed of social learners, then it would reach a

stable state where practices no longer evolve due to a lack of innovation (Laland, 2004). Indeed, there needs to be a balance between individual and social learners so that both strategies are adaptive (Barnard & Sibly, 1981; Rogers, 1988). Moreover, learning strategies must be adapted to the characteristics of the organization's environment. If an organization's environment is stable, then a low proportion of individual learners is sufficient to innovate before being copied to propagate the new practices through the organization. However, if the environment is constantly changing, then social learners are no longer adaptive, as they will always be late on the latest changes (McElreath, et al., 2008).

So far, I have advocated that two factors are important for the categorization of organizational changes: an organization's environment characteristics and the learning strategies of its members. The types of organizational changes resulting from these factors are summarized in Table 1 and explained below<sup>2</sup>.

**Revolution.** In stable environments, social learning is adaptive as it leads to lower costs (Boyd & Richerson, 2002; Henrich & Gill-White, 2001), as long as there is a source of innovation such as a small proportion of influential individual learners to copy (Galef, 1995; Laland, 2004). If there is a reliable source of innovation, then social learners can imitate the source and new practices can spread throughout the organization.

**Maladaptation.** When an organization's environment is prone to frequent changes, social learning is far less adaptive (McElreath et al., 2008). Because of the repeated changes, social learners are constantly late on more recent changes. At the extreme, if changes occur at every period, then social learners will always be late – and therefore less efficient than individual learners – as they rely on outdated information.

---

<sup>2</sup> Note that in order to compare the different cases, I operate under the assumption that learning strategies are exogenous and fixed.

**Waste.** In environments where changes are scarce, individual learning is costly (Boyd & Richerson, 1985; Henrich & Boyd, 1998; Henrich & Gil-White, 2001). However, while individual learning is less adaptive in such environments, they are a necessary source of innovation for social learners to copy from (Galef, 1995; Laland, 2004). Individual learners therefore generate discontinuous evolutions as they keep on sampling their environment and passively transmit new practices.

**Evolution.** Individual learning is more adaptive if an organization's environment is frequently changing (McElreath et al., 2008). As practices change frequently, sampling the environment allows individual learners to constantly innovate and adapt faster than if they were imitating past behaviors.

-----  
Insert Table 1 about here  
-----

## **DISCUSSION**

The aim of this paper has been to suggest a novel typology of organizational changes, by opening the black box often used to represent change recipients. On this account, I have proposed a multi-level description (summarized in Figure 1) accounting for the dynamics of learning strategies and practices, as well as their interaction with the organizational environment. Moreover, I have built on past research on individual social learnings, in an attempt to increase the psychological plausibility of the processes I described. Finally, I have proposed to distinguish between four types of changes: revolution, late revolution, discrete evolution, and continuous evolution. While this paper is at an early stage, it should serve as a basis for a further identification of organizational change patterns. I will continue by outlining

the main contributions and limitations of this work and finish by suggesting potential avenues for future research.

### **Theoretical Contributions**

First, I have emphasized the perspective of the organizational members in the context of organizational change. In so doing, I have described individuals as decision makers, who can choose between individual and social learning strategies. The multi-level dynamics are described in an attempt to contribute to the microfoundations of organizational change and open, at least partially, one of its black boxes (Lau & Woodman, 1995; Powell & Rerup, 2017). Not only have I included the decision strategies of the organizational members, but also the link between strategies and different environments, as well as their fitness. The resulting types of organizational change provide a basis for the identification of (i) responses to change and (ii) sources of change.

Second, the present framework aims at providing a more accurate picture of the individual processes underlying changes at the organizational level. It hence builds on evidence on organizational change and suggests a different typology of changes than the one identified by the literature. While a considerable amount of research has given attention to individuals facing changes (e.g., Bartunek & Moch, 1987; Dunham et al., 1989; Lau & Woodman, 1995; Oreg et al., 2018), the consideration of cognitive processes and relationships between organizational phenomena and individual processes remains scarce. In that regard, I have looked into the microfoundations of organizational change, that is, the processes unfolding when organizations and their members interact.

### **Limitations and Directions for Future Research**

I acknowledge several limitations of the proposed framework. First and foremost, for simplification purposes I have not described the various evolutionary paths that can unfold

from different types of social learning strategies, nor their comparison with individual learning strategies (see Boyd & Richerson 1985; McElreath, et al., 2008). Moreover, in order to establish a new typology, I have assumed that learning strategies are given exogenously and cannot change. Yet, the fitness of a strategy also decreases as the proportion of organizational members who rely on it increases (Laland, 2004). However, I intend to cover these in future work, and encourage future research to focus on these comparisons in order to further explore the microfoundations of organizational phenomena and open additional black boxes.

Second, I have left an important concept from the literature on organizational changes out of the debate: the attitudes toward change. While I acknowledge the importance of attitudes, I purposefully focused on decision-making strategies and environmental characteristics. However, it would be possible to extend this work to include change attitudes. For instance, Oreg et al. (2018) has argued that the source of change can impact the individual mechanisms of attitude generation. It is likely that individual learners hold positive attitudes toward changes, as they seek to innovate. On the contrary, one could make the argument that social learning is a strategy that should be adopted by organizational members who do not want to bear the costs of trial-and-error as they do not enjoy innovating.

Finally, although part of the literature is focusing on how to external forces can influence change processes (e.g., Greenwood, Suddaby, and Hinings, 2002), I have only considered organizational members as independent decision-makers. Furthermore, I have purposefully ignored any hierarchical notions that could take place within the organizations. However, I expect future research to investigate the role of hierarchy in learning. For instance, one could expect hierarchically higher organizational members to be more influential, and therefore to serve as role models to copy from. Yet, the existence of strong innovative figures can reduce the fitness of individual learning strategies for the other members (Barnard & Sibly, 1981; Rogers, 1988) and foster social learning.

## **Implications for Practice**

This work has two main implications for practice. First, it points at the importance for an organization and its members to know their environment. Indeed, social learning is more adaptive in stable environments than in continuously changing ones. Second, it emphasizes the importance of having both social and individual learners in order to be adaptive: while social learners reduce the costs of the change process, they require a source of new ideas to copy from.

## **Conclusions**

Although very tentative, this work draws on psychology and biology to open the black box of individual processes in the context of organizational change. By encompassing the interplay between an organization and its members, and how this interplay impacts organizational change, the suggested typology of organizational change should serve as a basis to identify patterns of changes and help better understanding the phenomenon of organizational change itself.

## REFERENCES

- Amis, J., Slack, T., & Hinings, C. R. (2004). The pace, sequence, and linearity of radical change. *Academy of Management Journal*, 47(1), 15-39.
- Anderson, P., & Tushman, M. L. (1990). Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative Science Quarterly*, 604-633.
- Barnard, C. J., & Sibly, R. M. (1981). Producers and scroungers: a general model and its application to captive flocks of house sparrows. *Animal behaviour*, 29(2), 543-550.
- Bartunek, J. M. (1984). Changing interpretive schemes and organizational restructuring: The example of a religious order. *Administrative Science Quarterly*, 355-372.
- Bartunek, J. M., & Moch, M. K. (1987). First-order, second-order, and third-order change and organization development interventions: A cognitive approach. *The Journal of Applied Behavioral Science*, 23(4), 483-500.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. University of Chicago press.
- Boyd, R., & Richerson, P. J. (1988). An evolutionary model of social learning: the effects of spatial and temporal variation. *Social learning: psychological and biological perspectives*, 29-48.
- Boyd, R., & Richerson, P. J. (1995). Why does culture increase human adaptability?. *Ethology and Sociobiology*, 16, 125-125.
- Boyd, R., & Richerson, P. J. (2002). Group beneficial norms can spread rapidly in a structured population. *Journal of Theoretical Biology*, 215(3), 287-296.
- Cardinale, I. (2018). Beyond constraining and enabling: Toward new microfoundations for institutional theory. *Academy of Management Review*, 43(1), 132-155.

- Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton University Press.
- Chandler, D., & Hwang, H. (2015). Learning from learning theory: A model of organizational adoption strategies at the microfoundations of institutional theory. *Journal of Management*, 41(5), 1446-1476.
- Coleman, J. S. (1994). *Foundations of social theory*: Harvard university press.
- Dacin, T. M., Goodstein, J., & Scott, R. W. (2002). Institutional theory and institutional change: Introduction to the special research forum. *Academy of Management Journal*, 45(1), 45-56.
- Dunham, R. B., Grube, J. A., Gardner, D. G., Cummings, L., & Pierce, J. L. (1989). *The development of an attitude toward change instrument*. Paper presented at the Academy of Management annual meeting, Washington, DC.
- Galef Jr, B. G. (1995). Why behaviour patterns that animals learn socially are locally adaptive. *Animal behaviour*, 49(5), 1325-1334.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–482.
- Gigerenzer, G., & Selten, R. (Eds.) (2001). *Bounded rationality: The adaptive toolox*. Cambridge, MA: MIT Press.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. Oxford University Press, USA.
- Giraldeau, L. A., Valone, T. J., & Templeton, J. J. (2002). Potential disadvantages of using socially acquired information. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1427), 1559-1566.

- Greenwood, R., & Hinings, C. R. (1996). Understanding radical organizational change: Bringing together the old and the new institutionalism. *Academy of Management Review*, *21*(4), 1022-1054.
- Greenwood, R., & Hinings, C. R. (2006). Radical organizational change. *The SAGE handbook of organization studies*, *2*, 814-842.
- Greenwood, R., Suddaby, R., & Hinings, C. R. (2002). Theorizing change: The role of professional associations in the transformation of institutionalized fields. *Academy of Management Journal*, *45*(1), 58-80.
- Harmon, D. J., Green Jr, S. E., & Goodnight, G. T. (2015). A model of rhetorical legitimation: The structure of communication and cognition underlying institutional maintenance and change. *Academy of Management Review*, *40*(1), 76-95.
- Hatch, M. J. (1993). The dynamics of organizational culture. *Academy of Management Review*, *18*(4), 657-693.
- Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and human behavior*, *22*(3), 165-196.
- Hertwig, R., & Hoffrage, U. (2013). Simple heuristics: the foundations of adaptive social behavior. In *Simple heuristics in a social world* (pp. 3-36). Oxford University Press.
- Isabella, L. A. (1990). Evolving interpretations as a change unfolds: How managers construe key organizational events. *Academy of Management Journal*, *33*(1), 7-41.
- Kheirandish, R., & Mousavi, S. (2018). Herbert Simon, innovation, and heuristics. *Mind & Society*, *17*(1-2), 97-109.

- Kostova, T., & Roth, K. (2002). Adoption of an organizational practice by subsidiaries of multinational corporations: Institutional and relational effects. *Academy of Management Journal*, 45(1), 215-233.
- Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, 32(1), 4-14.
- Lau, C.-M., & Woodman, R. W. (1995). Understanding organizational change: A schematic perspective. *Academy of Management Journal*, 38(2), 537-554.
- Luan, S., Reb, J., & Gigerenzer, G. (2019). Ecological rationality: Fast-and-frugal heuristics for managerial decision making under uncertainty. *Academy of Management Journal*, 62(6), 1735-1759.
- McElreath, R., Bell, A. V., Efferson, C., Lubell, M., Richerson, P. J., & Waring, T. (2008). Beyond existence and aiming outside the laboratory: estimating frequency-dependent and pay-off-biased social learning strategies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509), 3515-3528.
- Miller, D., & Friesen, P. H. (1980). Momentum and revolution in organizational adaptation. *Academy of Management Journal*, 23(4), 591-614.
- Nadler, D. A., & Tushman, M. L. (1989). Organizational frame bending: Principles for managing reorientation. *Academy of Management Perspectives*, 3(3), 194-204.
- Oreg, S., Bartunek, J. M., Lee, G., & Do, B. (2018). An affect-based model of recipients' responses to organizational change events. *Academy of Management Review*, 43(1), 65-86.
- Powell, W. W., & Colyvas, J. A. (2008). Microfoundations of institutional theory. *The Sage handbook of organizational institutionalism*, 276, 298.

- Powell, W. W., & Rerup, C. (2017). Opening the Black Box: The Microfoundations of Institutions. In R. Greenwood, C. Oliver, T. B. Lawrence, & R. E. Meyer (Eds.), *The SAGE Handbook of Organizational Institutionalism (2nd ed.)* (pp. 311-337). London & Thousand Oaks, CA: Sage.
- Rogers, A. R. (1988). Does biology constrain culture?. *American Anthropologist*, *90*(4), 819-831.
- Romanelli, E., & Tushman, M. L. (1994). Organizational transformation as punctuated equilibrium: An empirical test. *Academy of Management Journal*, *37*(5), 1141-1166.
- Schilke, O. (2018). A micro-institutional inquiry into resistance to environmental pressures. *Academy of Management Journal*, *61*(4), 1431-1466.
- Schlag, K. H. (1998). Why imitate, and if so, how?: A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory*, *78*(1), 130-156.
- Schlag, K. H. (1999). Which one should I imitate?. *Journal of Mathematical Economics*, *31*(4), 493-522.
- Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). MIT press.
- Todd, P. M., Gigerenzer, G., & the ABC Research Group. (Eds.). 2012. *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.
- Townley, B. (2002). The role of competing rationalities in institutional change. *Academy of Management Journal*, *45*(1), 163-179.
- Tracey, P. (2016). Spreading the word: The microfoundations of institutional persuasion and conversion. *Organization Science*, *27*(4), 989-1009.

- Tsoukas, H. (1989). The validity of idiographic research explanations. *Academy of Management Review*, 14(4), 551-561.
- Tushman, M. L., & Romanelli, E. (1985). Organizational evolution: A metamorphosis model of convergence and reorientation. *Research in Organizational Behavior*, 7, 171-222.
- Tushman, M. L., & Rosenkopf, L. (1996). Executive succession, strategic reorientation and performance growth: A longitudinal study in the US cement industry. *Management Science*, 42(7), 939-953.
- Van de Ven, A. H., & Poole, M. S. (1995). Explaining development and change in organizations. *Academy of Management Review*, 20(3), 510-540.
- Virany, B., Tushman, M. L., & Romanelli, E. (1992). Executive succession and organization outcomes in turbulent environments: An organization learning approach. *Organization Science*, 3(1), 72-91.
- Whiten, A., Horner, V., Litchfield, C. A., & Marshall-Pescini, S. (2004). How do apes ape?. *Animal Learning & Behavior*, 32(1), 36-52.

## Appendix

Figure 1: Interplay between individuals' processes and practices, and their environment (adapted from Coleman 1994).

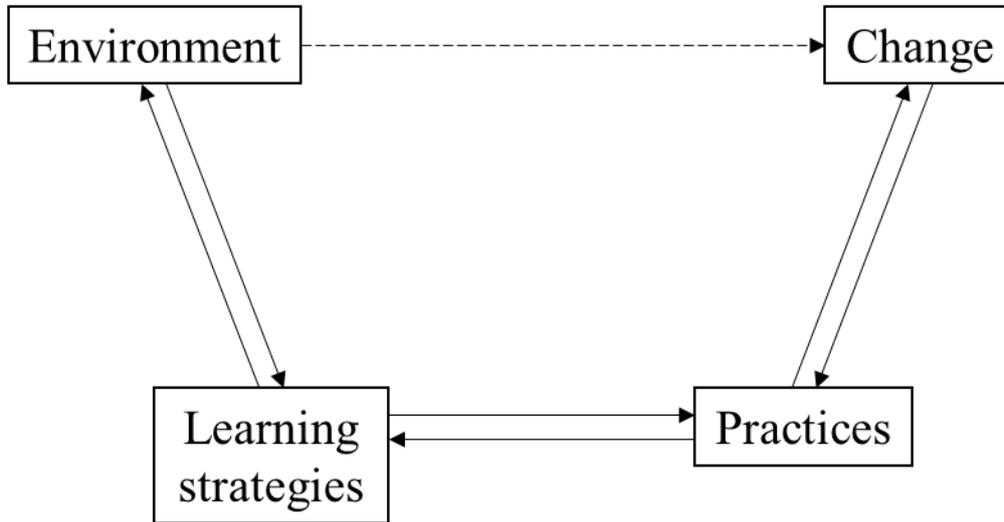


Table 1: Typology of organizational changes, based on the organization's environmental characteristics and the learning strategies of the organizational members.

		Environment	
		Stable	Changing
Learning	Social	Revolution	Maladaptation
	Individual	Waste	Evolution