

Dynamics of Norms in Decision-Making

**A Psychological Analysis Combining Theory, Experiment,
and Social Simulation**

Cumulative Dissertation

submitted in fulfillment of the requirements for the academic degree of

Doktorin der Philosophie (Dr. phil.)

at the Faculty of Human Sciences

at the University of Kassel

by Marlene Clara Lucia Batzke

Kassel, August 2023

Reviewers:

Prof. Dr. Andreas Ernst

Prof. Dr. Mirjam Ebersbach

Dr. Friedrich Krebs

Commission Members:

Prof. Dr. Georg von Wangenheim

Dr. Anita Körner

Date of Submission:

30th August 2023

Date of Defense:

8th December 2023

CONTENTS

ACKNOWLEDGEMENTS.....	5
LIST OF PUBLICATIONS.....	6
SUMMARY.....	7
1. INTRODUCTION.....	9
2. RESEACH QUESTIONS AND GENERAL APPROACH.....	11
2.1 Aim and Research Questions.....	11
2.2 Cooperation Norms in Social Dilemmas.....	12
2.3 General Approach: Theory, Social Simulation, and Experiment.....	14
2.3.1 A Theory on the Dynamics of Norms.....	15
2.3.2 An Agent-Based Model on Norm Dynamics.....	16
2.3.3 An Experimental Approach to Norm Dynamics.....	20
3. NORMS AND NORM DYNAMICS IN THEORY.....	22
3.1 Norm Definitions and Taxonomies.....	22
3.2 Theoretical Work on Norms.....	24
3.3 Paper I: The Effects of Norms on Environmental Behavior.....	26
3.4 Norm Terminology of the Present Work.....	28
4. NORM DYNAMICS IN AGENT-BASED MODELING.....	29
4.1 Norm Research in Social Simulation.....	29
4.2 Norm Internalization in Agent-Based Modeling.....	31
4.3 Paper II: Conditions and Effects of Norm Internalization.....	33
4.4 DINO Model 1.1: Results from a Revised Model.....	35
4.4.1 Revision from DINO Model 1.0 to 1.1.....	36
4.4.2 Results from DINO Model 1.1.....	37
5. NORMS AND NORM DYNAMICS IN EMPIRICAL RESEARCH.....	40
5.1 Norms in Psychological Studies.....	40

5.2	Experiments on the Dynamics of Norms.....	42
5.3	Paper III: Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation.....	43
6.	PREVIEW ON COMPARING NORM DYNAMICS IN AGENT-BASED MODELING AND EXPERIMENT	44
7.	GENERAL DISCUSSION.....	46
7.1	Addressing the Research Questions.....	46
7.1.1	How do Different Types of Norms Change?.....	47
7.1.2	How are Norms Internalized and What are the Effects of Internalization?.....	49
7.2	Limitations and Future Research.....	50
7.2.1	External Validity.....	50
7.2.2	Generalizability.....	52
7.2.3	Internal Validity.....	53
7.3	Outlook and Conclusion.....	54
	REFERENCES.....	56
	APPENDIX.....	69
	Appendix A – Paper I: The Effects of Norms on Environmental Behavior.....	69
	Appendix B – Paper II: Conditions and Effects of Norm Internalization.....	114
	Appendix C – Results from DINO Model 1.1.....	146
	Appendix D – Paper III: Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation.....	148
	Appendix E – Working Paper: An Experimental Attempt at Validating an Agent-Based Model on Decision-Making, Social Norm Change, and Norm Internalization.....	198
	Erklärung zum Eigenanteil.....	212
	Eidesstattliche Versicherung und Erklärung.....	214

ACKNOWLEDGEMENTS

I am grateful to several groups of people, without whom this thesis would not have been possible. My special thanks go to my supervisor Andreas Ernst for his continuous support, his tireless willingness to engage in critical dialogue, for always taking the time when needed, and for introducing me to the world of social simulation. His scientific way of thinking, psychological expertise, and worldview have shaped this work – and me. I am grateful to Mirjam Ebersbach and Anna Helfers for their expert suggestions on experimental design, planning, and execution. I thank Georg von Wangenheim and Fabian Mankat for inspiring discussions and our joint theorizing on what norms are and how they work, which laid the foundation to this thesis. To Georg von Wangenheim I am also grateful for spotting a logical flaw in the model. I am thankful to Friedrich Krebs for his assistance in the model revision. I thank all those involved in the *ZumWert* project for the interdisciplinary enriching collaboration and for co-authoring a publication. I am grateful to the University of Kassel for the funding of the project. I am thankful to the student assistants at the CESR, Paula Rosendahl and Martin Lühr, for assisting at diverse stages of the research process. I thank my colleagues at the CESR for the friendly atmosphere at the institute and for the good time beyond. Finally, my warmest thanks go to my family and friends for being with me all the way.

LIST OF PUBLICATIONS

Large parts of this thesis have already been published, are in press, or are submitted for publication. In the following, it is indicated which parts of this thesis are based on which publications.

Section 3.3 is based on:

Dannenberg, A., Gutsche, G., Batzke, M. C. L., Christens, S., Engler, D., Mankat, F., Möller, S., Weingärtner, E., Ernst, A., Lumkowsky, M., von Wangenheim, G., Hornung, G., & Ziegler, A. (in press). The effects of norms on environmental behavior. *Review of Environmental Economics and Policy*.

Section 4.3 is based on:

Batzke, M. C. L., & Ernst, A. (2023b). *Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation* [Manuscript submitted for publication]. Center for Environmental Systems Research, University of Kassel.

Section 5.3 is based on:

Batzke, M. C. L., & Ernst, A. (2023c). Conditions and Effects of Norm Internalization. *Journal of Artificial Societies and Social Simulation*, 26(1), 1–31.
<https://doi.org/10.18564/jasss.5003>

SUMMARY

The present thesis presents an attempt at studying the dynamics of norms in decision-making. Norms are behavioral rules, governing individuals' behavior and organizing social life. While social norms emerge in social interaction and influence decision-making, individuals also develop their own personal norms. The concept of norm internalization describes the process of how personal norms develop and change. While the existence, role, and effects of norms on decision-making have been well studied, there is a lack of understanding how norms change. The aim of the present work was twofold: (1) addressing differences between change processes of different types of norms and (2) focusing more particularly on how norms are internalized. To study norm dynamics, a multi-method approach was chosen, combining theory, social simulation, and experimental research. This methodological triad allowed taking a dynamic perspective on norms. While these three elements complement each other, they also individually contributed to addressing the research questions.

In the theoretical contribution, a psychologically grounded conceptual model of norm change and decision-making was presented. It was based on a refined norm taxonomy that disentangles different aspects that define norms and represents them individually. That way, direct effects of each distinct type of norm may be addressed. The conceptual model states theoretically and empirically grounded assumptions about how different types of norms change and influence each other as well as decision-making. Moreover, it includes assumptions about the norm internalization process. A review of empirical literature with respect to the conceptual model showed research gaps concerning the mechanisms underlying norms, including norm internalization, and their dynamic interactions.

The behavioral and social consequences of the theoretical assumptions over time were investigated via social simulation, more specifically agent-based modeling. Large parts of the conceptual model were implemented into an agent-based model, which demanded extending the conceptual model, making it computationally explicit, and applying it to a situational context. The agent-based model entails a dynamic theory of decision-making, including different types of norms as well as non-normative behavioral motivators. Norm internalization was assumed a slow learning process, occurring slower than the learning of social norms, and influencing decision-making on a higher level than social norms. The agent-based model simulates the behavior of three agents playing a social dilemma game. Simulation experiments

investigating the conditions of norm internalization showed that a cooperative personal norm can be internalized as appropriate with different underlying motivational structures – even those favoring non-cooperation from the start. Simulation experiments addressing the effects of norm internalization further suggested that norm internalization leads to norm compliance and showed possible conditions for social equality and behavioral stability (with norm internalization resembling habits) as well as conditions for inequality and instability.

The assumed differences in the temporal dynamics of personal norm change (i.e., norm internalization) and social norm change were investigated in an experimental study. Participants played a repeated social dilemma game with artificial co-players representing a predominantly cooperative or uncooperative social setting, depending on the experimental condition. This was expected to affect slow learning of personal norms. Additionally, the cooperativeness of the social setting was varied repeatedly *within* conditions, which was expected to result in fast changes in social norms. Participants' personal and social norms were assessed throughout the game. In support of the assumptions, personal norms tended to change more slowly and were affected not just by situational factors (as in the case of social norms) but also personal factors. Yet, an assumed linear change in personal norms was not supported, leaving the question of how norm internalization proceeds open.

The simulated and experimental data of behavior, social norms, and personal norms were compared with each other, aiming for a deeper understanding of norm change processes. Participants' and agents' social norm changes showed similarities. Regarding personal norm change, the agent-based model was improved by implementing a new mechanism: a negativity bias, making internalization of the appropriateness of cooperation more difficult compared to defectivity. Hence, the comparison led to suggest that a negativity bias is a possible mechanism in norm internalization in the social dilemma context.

There are several limitations to the present work, relating to the external and internal validity and generalizability of the results. Yet, the present work represents one step towards a psychological approach to studying norm dynamics in decision-making. It provides various starting points for future research and argues for the importance of an interdisciplinary view on norms.

1. INTRODUCTION

Norms are informal rules and principles informing everyday behavior and all aspects of social life. They are an essential aspect in coordinating human co-existence. They are ubiquitous, manifold, differ among people, and are potentially contradictory, even within a single person. Norms exist on different levels. Social norms describe what many people consider appropriate or normal behavior, emerging in social interactions. While individuals adhere to and learn social norms, they also develop their own personal norms, being what the individual considers appropriate or inappropriate behavior. Most individuals perceive and want to perceive themselves as independent, self-consistent actors and as ‘doing the right thing’ – whatever that means individually. Norm internalization is a key concept behind these strivings, describing how an individual develops and changes its personal norms. Social as well as personal norms guide behavioral decisions and play an important role in maintaining social order and cooperation.

Norms are subject to change, and they have the potential to generate large-scale behavior change. Being an important part of societal transformation processes, they are vital for facing the pressing challenges of our time, such as climate and environmental destruction. So far, the dynamics of norms, concerning the questions of how and when they change and stabilize, are little understood. Particularly the process of personal norm change, also referred to as norm internalization, remains one of the great riddles of norm research. Norm internalization has been shown important for norm compliance, maintenance, and long-term behavior change. Theorists have also ascribed a crucial role to norm internalization in social change and societal transformation. Yet, the questions of how and when norms are internalized remain open.

Norms have a long tradition of being studied and theorized about by scientists from a wide range of disciplines. At its heart, thinking about norms calls for an interdisciplinary perspective, as they are the product of behavioral decisions, social interactions, and the societal context. Based on this coupling of individual behavior and the social context, norms can be understood as a complex system. Hence, they are an emergent phenomenon to the complex interplay of socially embedded individuals, while at the same time influencing individual level decision-making. This dynamic perspective on norms comprises a temporal dimension of change over time and a social dimension of mutual influence due to their social embeddedness.

To address questions that follow from a dynamic perspective on norms, in the present dissertation, an interdisciplinary and multi-method approach was applied. First, a theoretical approach was taken, describing norm change on different levels. Second, a social simulation method was applied that allows investigating norm dynamics directly: agent-based modeling. Third, an experimental approach to studying norm change was deployed. While the basis for this approach was an interdisciplinary perspective on norms, the theoretical background strongly builds on psychological literature.

The aim of studying norm development, maintenance, and change, is gaining a deeper understanding of behavioral and social change. Norms per se are neither ‘good’ nor ‘bad’. They may facilitate extraordinary acts of large-scale cooperation, such as compliance with social distancing during the COVID-19 pandemic. At the same time, there are many societies governed by social norms that are harmful to their members, for instance, discriminating people of a specific gender, sexuality, or skin color, or promoting lifestyles harmful to the environment and climate. Understanding of the dynamics of norms may support bringing about desirable change.

The present dissertation represents an attempt at studying norm dynamics via combining theory, social simulation, and experiment. Due to this methodological triad, the following (second) chapter introduces the general approach against the background of the research questions. The following three chapters then address: theoretical work on norms and norm dynamics (Chapter 3), norm dynamics in social simulation (Chapter 4), and empirical research on norms and norm dynamics (Chapter 5). As contributions were made to each of these fields, Chapters 3 to 5 begin with a presentation of the current state of research and are followed by a summary of the original research that was conducted within the scope of the present dissertation. Chapter 6 presents a preview of ongoing work on comparing norm dynamics in simulation and experiment. Finally, Chapter 7 discusses the insights gained from the present work regarding the research questions, discusses limitations, provides an outlook, and concludes.

2. RESEACH QUESTIONS AND GENERAL APPROACH

The following chapter presents (1) the aim and research questions of the present dissertation, (2) the context in which norms were studied, namely cooperation norms in social dilemmas, and (3) the general approach of combining theory, social simulation, and experiment.

2.1 Aim and Research Questions

All human societies are governed by informal or formal, spontaneous or deliberated norms. There is a myriad of norms ruling all aspects of social life. They are considered the “grammar” (Bicchieri, 2006), “glue” (Gelfand, 2018), or “wheels” (Conte & Dellarocas, 2001) of society. Part of societal adaptation to environmental and societal pressures is that norms constantly change – existing ones are adjusted, and new ones emerge (Conte et al., 2014). The existence, structure, and role of norms has long been theorized about by philosophers such as David Hume, Thomas Hobbes, and Immanuel Kant. Their immense significance for guiding individuals’ decisions and social dynamics has substantive evidence. The next frontier in norm research is a better understanding of how they change. Only by understanding the underlying processes of norm change, we can begin to figure “where, when, and how social norms can be a solution to solve large-scale problems, but also to recognize their limits” (Andrighetto & Vriens, 2022, p. 1). Norms represent an important factor in the in-between of societal level and individual level transformation. A crucial concept at that intersection is the one of norm internalization, describing how these levels interact (Hoffman, 2000; Neumann, 2014; Vygotsky, 1930/1981). This makes norm internalization not only relevant for individual level decision-making but also for societal transformation processes. To foster change towards a socially and environmentally just future, it needs a better understanding of the mechanisms of norm development, change, and internalization.

The present dissertation aimed at studying norm dynamics – norm development and change over time resulting from interactions between individual behavior and the social context. More specifically, the aim of the present work was twofold. First, potential differences in the dynamics of different types of norms were to be investigated, answering to the question: How do different types of norms change? This incorporates the questions of how, when, and why each type of norm changes as well as the question of potential differences between norm

processes. Second, the process of norm internalization, meaning the development and change of personal norms, was given a particular focus, assuming that it would differ from other types of norm changes. As the process itself is yet little understood, the central questions regarding norm internalization that the present work aimed at addressing are: How are norms internalized and what are the effects of internalization?

To pursue these questions, a multi-method approach was chosen, combining theory, social simulation, and experiment. While these three elements complement each other, they also individually contributed to addressing the research questions. In the theory, assumptions about how a specific type of norm changes within the individual and how different types of norms affect each other and the individual decision-making process were presented. Moreover, theoretical assumptions about how norms are internalized and how this affects decision-making were postulated. Social simulation (more specifically: agent-based modeling) allowed investigating the effects of norm change, such as the effects of norm internalization on behavioral and social dynamics, as well as the conditions for norm change. In the experiment, the theoretical assumptions about differences in the dynamics of different types of norms were empirically tested. It allowed detecting associated factors in norm change and examining the behavioral importance of different types of norms.

By choosing this multi-method approach, the present dissertation also aimed for a methodological goal of taking a step towards bridging the gap between psychology and agent-based modeling. From the psychological perspective, agent-based modeling holds the potential of addressing new questions concerning change and dynamic interactions, advancing theory development, and investigating the interrelations of psychological phenomena and the social context (see Section 2.3.2). From the perspective of agent-based modeling, psychology offers great depth in providing concepts to understand the human mind. This is much needed as psychological realism is often missing in existing agent-based models. The present work therefore aimed at exploring the synergies of combing theoretical and experimental psychological research with agent-based modeling.

2.2 Cooperation Norms in Social Dilemmas

Norms arguably influence decisions in most situations, and it can be assumed that in most social situations, norms may emerge. The context that was chosen for studying norms in the present work is conflict situations. Conflict situations, especially social dilemmas, are defined as a mixture of purely cooperative situations and purely competitive situations (Ernst, 1997). Hence,

they are mixed motive situations, where both cooperative and competitive motivations play a role, directing towards conflicting behavioral decisions. They therefore pose a particular challenge for the individual as there is no clear right or wrong choice (Axelrod, 1984), depending very much on the context and its history. Social dilemmas are ubiquitous in our everyday lives, ranging from the next holiday trip to living in a shared apartment to the usage of a common good. Climate change, environmental destruction, extinction of the species – many of the most pressing challenges that humankind faces in the 21st century put the individual in a social dilemma (Dawes, 1980). All that makes the study of norm dynamics particularly interesting and relevant in the context of a social dilemma, since norms pose a possible solution to solving the individual's conflict (Ostrom, 1990, 1999; Thøgersen, 2008).

In game theory (von Neumann & Morgenstern, 1944), dilemmas are described in form of strategic games. Real-world dilemmas are simplified, standardized, and abstracted into a game, which is a system of numerical representations of behavioral consequences (so-called payoffs) depending on the choices of the players (Liebrand et al., 1992). Dilemma games allow formally describing and analyzing the structure of a dilemma. Moreover, they serve as an experimental design to study human behavior in dilemmas (Ernst, 1997). The social dilemma is characterized by two properties (Dawes, 1980). Each individual is better off by making a socially defecting (i.e., non-cooperative) choice (e.g., not recycling, watering the lawn in summer or finishing off the milk from the collective refrigerator without refilling the stock) than a socially cooperative choice, irrespective of what the others in society do. However, if all cooperate rather than defect, everyone in society profits (Axelrod, 1984).

The most famous social dilemma is the *prisoner's dilemma game* (Luce & Raiffa, 1957). Therein, the scenario is described as follows. Two offenders have jointly committed a felony. They are caught by the district attorney and put into two separate cells for questioning, having no chance of communicating with each other. The district attorney cannot prove their guilt, so she offers each the chance to confess. If only one of them confesses, that person will be released from prison (*temptation*), while the other receives the maximum prison term (*sucker's payoff*). If both confess, both prisoners receive a medium-length term (*punishment*). If neither confesses, both receive a short term (*reward*). Hence, each player may cooperate with the other player (i.e., not confess) or defect (i.e., confess). Each player is tempted to defect by achieving the best individual outcome, when being the only defector. Mutual cooperation is rewarded with the best collective outcome. Yet, being the only cooperator is the worst payoff. This describes the core of the conflict that individuals face in a prisoner's dilemma (cf., Liebrand et al., 1992).

In research, games are often played repeatedly, consisting of multiple rounds. The iterated prisoner's dilemma with only two players is the simplest social dilemma. Yet, the two-player scenario is not representative of real-world social dilemmas in general due to a couple of limitations that are unique to the two-person scenario (Dawes, 1980). Negative consequences of defection are scaled down to the one other player, rather than diffused over a group. Also, a player in iterated games can use its own behavior to directly discipline the other player. As soon as there are more than two players involved, these limitations do not apply anymore, the bigger the group, the lesser.

As a first step towards a more representative social dilemma, in the present work, an iterated 3-person prisoner's dilemma game was selected for studying norms via agent-based modeling as well as experimental research. Hence, it is norms of cooperation and defection in social dilemmas that were investigated in the present work. Norms play a vital role in motivating cooperative behavior in social dilemmas. They motivate cooperation in small groups (Bicchieri et al., 2023) and can create tipping points for large-scale transformations (Nyborg et al., 2016). Cooperation is essential for solving many societal problems, such as environmental and climate protection. However, individual self-interest often conflicts with the collective interest, making it challenging to achieve cooperation (Hardin, 1968). Norms of cooperation can act as a powerful tool for motivating individuals to act in the interest of the collective. By studying these norms, we can better understand how to promote cooperative behavior and prevent social and societal crises resulting from selfish actions.

2.3 General Approach: Theory, Social Simulation, and Experiment

Addressing the interplay of individual level and social level norm dynamics represents an innovative perspective in psychological norm research. Yet, it also bears several challenges. Psychology provides little theories on how norms change, let alone theoretical superstructures combining theory on norm change relating to individuals' decision-making processes. The dynamics of norm internalization are particularly challenging, lacking not only specific theoretical assumptions but also empirical research addressing this internal learning process.

To address the intended research questions, an interdisciplinary and multi-method approach was chosen, combining theory, social simulation, and experiment. In the following, it is further elaborated on how theory, social simulation, and experiment contribute individually and how they interact, complementing each other.

2.3.1 A Theory on the Dynamics of Norms

Only theory may answer to the questions of *how* norms influence decision-making, *how* change in norms occurs, and *how* norms are internalized. Theorizing about norm change and norm internalization demands for a dynamic perspective on norms. While many influential psychological norm and/or decision-making theories describe how norms affect behavior (e.g., the *theory of planned behavior*, Fishbein & Ajzen, 1981; see Section 3.2), they explain static snapshots in behavioral patterns (Eberlen et al., 2017; Smaldino et al., 2015). Yet, elements to the concept of norms are that (1) they are constantly changing and (2) they emerge at the interplay of individuals' behaviors and social dynamics (Neumann, 2014). Thus, to develop a theory of norms that describes norm change and norm internalization as well as individual behavior, it seems fruitful to include the temporal as well as the social dimension.

The temporal dimension relates to including feedback processes of how behavior in turn affects norms or other (preceding) variables that influence norms. These feedback loops describe how behavior affects psychological variables that influence the decision-making and resulting behavior at the next point in time via adaptation and learning processes (Jager & Ernst, 2017). Including these in a theory involves an interdisciplinary perspective as an individual's behavior may also have consequences on the social and physical environment (Schlüter et al., 2017). One person's behavior may influence another person's norms and behavior, which leads to the relevance of incorporating the social dimension in a theory on norm dynamics. An individual's perception of a certain norm can be the result of social interactions, observing a neighbor's behavior or collective behavioral patterns, being confronted with societal rules, and so forth (Nyborg, 2018). Hence, norms may emerge on different levels within society, in the individual's mind and as social phenomena (Edmonds, 2014). To understand norm dynamics, it is important to consider both levels and their interaction (Conte et al., 2014). As a result, a theory on norm dynamics could not only be able to explain norm change, norm internalization, and behavior change in individuals, but social and societal change as well. Moreover, it may provide hints to leverage points to foster such a change.

We need theories to know which data to collect (Epstein, 2008). A theory serves "as an anchoring tool for the research process" (Eberlen et al., 2017, p. 151). It is the foundation for developing and testing experimental hypotheses in different experimental situations. Theory serves as a reference frame to compare experimental results against. It allows generalizing observed phenomena and applying them to experimental situations that are different than the studied ones (Klein, 2014). Simulating a theory allows testing the consequences of the

assumptions over time, which may lead to develop new hypotheses, improve theory, and so forth (see the following Section 2.3.2).

To address the research questions (see Section 2.1), a theoretical contribution in form of a conceptual model on norm change and decision-making was made. It is presented in Chapter 3. It combines existing assumptions on how different types of norms change over time grounded in a decision-making framework and introduces assumptions on the interdependence of different types of norms as well as how norms are internalized.

2.3.2 An Agent-Based Model on Norm Dynamics

Norms are the product of behavioral decisions and social interactions. This coupling of individual behavior and the social context is characteristic for complex social systems and often leads to nonlinear dynamics (Gilbert & Troitzsch, 1999; Jackson et al., 2017, Wilensky & Rand, 2015). It contributes to the stability of norms, which sometimes change painfully slowly over a long time. However, once change is happening, it may also cause self-amplifying processes (Rogers, 2003). The more people may already have solar panels on their roofs, the more pressure might others feel, to get solar panels as well. These cascade effects can lead to a tipping-point (Nyborg et al., 2016), which introduces a new ‘normal’: most people have solar panels on their roofs.

The human mind, powerful as it is, struggles with thinking in these complex systems (Resnick, 1994). People give processes in time insufficient consideration, have trouble in dealing with exponential developments, and tend to focus on a main effect, discounting side effects (Dörner, 1980). Moreover, people have difficulties thinking on different levels, such as the individual and social level, falsely attributing properties from one level to the other (Resnick, 1996; Resnick & Wilensky, 1998). Based on the interaction of individual elements, qualitatively new properties may arise at the aggregate system level, which is encapsulated in the concept of emergence, being the central characteristic of complex systems (Holland, 2000; Wilensky & Rand, 2015).

Hence, norms are an emergent phenomenon to the complex interplay of socially embedded individuals (Edmonds, 2014). While theoretical assumptions build the foundation for thinking about how norms change and are internalized, they are unfit to investigate dynamics like the conditions of *when* and *why* norms change and are internalized across time and their effects on the behavioral and social dynamics. Similarly, many statistical methods are not well suited for answering these questions, focusing on statistical regularities between

variables rather than the underlying processes that account for an observed phenomenon (Smith & Conrey, 2007). For investigating the relationship between microlevel individual decision-making as well as norm internalization processes and macrolevel social norm emergence, a method capable of studying complex systems and its nonlinear dynamics is advantageous (Jackson et al., 2017). This is where agent-based modeling comes into play.

What is Agent-Based Modeling?

Agent-based modeling is a type of computer simulation. It is used within the field of social simulation, wherein a computer program is written to study some aspect of a social system (Gilbert & Troitzsch, 1999). Social simulation in turn is part of the larger field of the computational social sciences (Gilbert, 2008). In agent-based modeling, a social system is represented by its individual elements (Edmonds, 2014). Many social phenomena are emergent to the interactions of individuals (Gilbert, 2008), such as crowding, overuse of a resource, or social norms. The ‘agents’ in agent-based modeling are autonomous entities (Wilensky & Rand, 2015), which may refer to people, households, institutions, nations, viruses, and so forth. Agents possess behavioral rules, may interact with each other and their physical or biological environment, and may differ among each other in their goals, strategies, or characteristics (Railsback & Grimm, 2019). They may adapt, learn, store items in memory, and change their behavior or their behavioral strategies over time, depending on their interactions with other agents and their environment (Gilbert, 2008). Hence, the loop is closed between individuals’ decision-making, their behavior, the behavioral consequences, and learning/adaptation processes influencing the next decision (Jager & Ernst, 2017). That way, behavioral dynamics over time are explicitly represented and can be investigated.

The idea of modeling is common in most social sciences: “One creates some kind of simplified representation of ‘social reality’ that serves to express as clearly as possible the way in which one believes that reality operates” (Gilbert, 2008, p. 2). That equally applies for statistical modeling or a verbal theoretical description (Bossel, 1994). An agent-based model is a computer program; hence, it may simulate dynamics over time. Apart from agent-based modeling, there are other types of computer simulations, which can be used to study social systems, for instance, microsimulation, system dynamics, and cellular automata (Gilbert & Troitzsch, 1999). Each have their strengths and limitations, while a thorough review goes beyond the scope of the present work. Characteristic for agent-based modeling is the explicit representation of heterogeneous agents with some elements of cognition, such as individual

learning or reasoning processes (Edmonds, 2014). Modeling takes place on the individual level, being a ‘bottom up’ approach not demanding knowledge about the aggregated level (Railsback & Grimm, 2019). Hence, agent-based modeling allows for a truly psychological perspective. As the system level dynamics emerge from agent interactions, the complex interplay between levels can be investigated (Lorenz et al., 2021). Many other modeling and simulation types focus on one level of organization (Smaldino et al., 2015). Therefore, agent-based modeling seems particularly well suited for studying the relation between psychological processes and social phenomena, such as the individual and social level dynamics of norm change (Conte et al., 2014). Moreover, agent-based modeling is unlimited in its ability to incorporate agent heterogeneity (Smaldino, 2020). Interindividual psychological differences are arguably highly relevant for norm change and internalization (Vygotsky, 2004). For these reasons, agent-based modeling was applied in the present work.

Why Simulate a Norm Theory?

A verbal psychological theory on the dynamics of norms may describe how norm change in individuals proceeds. A simulated theory may show how the theoretical assumptions about psychological norm processes in individuals play out over time and lead to social dynamics of change. By generating simulated behavior based on psychological theory, the agent-based model can describe the assumed underlying processes and mechanisms that may cause norm or behavioral change to emerge. This ‘generative’ explanation of a theoretical agent-based model therefore differs from those of a verbal psychological theory and statistical analysis as aggregation of experimental data (Epstein, 2008; Wilensky & Resnick, 1999). Experimental data may explain static relations between variables and via statistical analyses such as multiple regression statistical regularity can be shown (Smith & Conrey, 2007). Yet, by artificially producing a norm or behavior change, agent-based modeling may provide a deeper understanding of a phenomenon (Smaldino et al., 2015) and should represent an additional value to psychological (norm) research (Conte et al., 2014).

That way, agent-based modeling serves as a tool for theory development (Jager, 2017). As argued above, people struggle with conceptualizing and thinking in complex systems (Dörner, 1980; Resnick, 1994). The agent-based model functions as an “object to think with” (Railsback & Grimm, 2019, p.10) or “protheses for the imagination” (originally by Murray Gell-Mann; cited by Smaldino et al., 2015, p. 303), rigorously testing the consequences of the theoretical assumptions, exceeding the limitations of the experimenter’s mind. The simulated

theory may then serve as a “virtual laboratory for experimenting with theory” (Lorenz et al., 2021, p. 626; see also Dowling, 1999; Troitzsch, 2017). It allows conducting experiments on artificial populations and exploring how macrolevel phenomena relate to microlevel interactions (Jager & Ernst, 2017; Lorenz et al., 2021). It allows addressing research questions concerning norm dynamics, such as: *when* and *why* are norms internalized across time, depending on which psychological or social conditions? What are the effects of norm internalization on the behavioral and social dynamics? Simulation experiments are thereby not bound to the limitations of the real world (Wilensky & Rand, 2015), like ethical concerns, the temporal rate of change, or feasibility, such as investigating the role of norm internalization by comparing people with a norm internalization process to those without. Simulation outcomes can then be compared to theory and empirical data, which may lead to adapting theoretical assumptions, specifying boundary conditions, or developing new assumptions. That way, agent-based modeling contributes to developing, specifying, testing, and improving psychological theories (Jager & Ernst, 2017; Smaldino, 2020).

Simulation results can serve as hypotheses for further empirical investigation. One might be concerned that the agent-based model would only produce the outcomes that are predicted by the theory and hard-wired into it. Yet, psychological theories tend to describe individual-level behavior and predictions about the group level are not easily deducible. In fact, outcomes are often very different to what was expected (Flache & Macy, 2005). The potential for unexpected outcomes to arise despite full knowledge of the initial conditions is considered one of the strengths of agent-based modeling (Epstein, 1999). Simulation results may lead to new research questions. “It’s the new questions [...] that produce huge advances, and models can help us discover them” (Epstein, 2008, Section 1.15). Hence, agent-based modeling can stimulate empirical research, representing a bridge between theory and empirical research (Waldherr & Wettstein, 2019).

Simulating a dynamic norm theory also comes with challenges. Verbally formulated psychological theories rarely have the level of specification that is necessary for developing an agent-based model (Schlüter et al., 2017). To be simulated, any theory needs to be made explicit in computational terms, thus translated into behavioral rules of individual agents (Ernst, 2010). While that represents a challenge for model developers, it can be an asset to psychological theory development (Jager, 2017). It forces scientists to be explicit and not rely on the vagueness of words (Gilbert, 2008; Smaldino, 2020).

Validation of a psychological agent-based model poses another difficulty. Suitable empirical data of the psychological microlevel processes may not be sufficiently available to allow systematic validation (Gilbert, 2008). Existing psychological data often differ from what is needed for model validation (Jager & Ernst, 2017) or even can be assessed (Smith & Conrey, 2007). Another validation technique is comparing the simulation results with macrolevel empirical data. Thereby, statistical properties of the simulation outputs are matched with real-world social statistics (Moss & Edmonds, 2005). Moreover, the agent-based model can be compared with theory. Theory is used to derive assumptions about relationships between variables, which are then compared with simulation results under different parameter settings (Gilbert, 2008). Nevertheless, the complexity of simulation results makes validation a challenging task.

To avoid overly complex models, agent-based modeling simplifies and abstracts real-world systems along the purpose of the model (Railsback & Grimm, 2019). Hence, agent-based modeling demands for a selection of those aspects, which are assumed to matter most (Smaldino, 2017). Regarding a norm-based theory of decision-making that poses the question: Which are the most important factors influencing decision-making apart from norms? Representing a simplified version of reality, illuminating the core dynamics of certain aspects of a complex system is considered a model's strength, rather than error (Smaldino, 2020), since a model as complex as the real world would be just as confusing. As Georg Box famously put it: "All models are wrong, but some are useful." So far, agent-based modeling has gained little attention in psychology, while many have argued for and demonstrated its potential for psychology (Eberlen et al., 2017; Jackson et al., 2017; Jager & Ernst, 2017; Lorenz et al., 2021; Nowak et al., 1990; Smaldino, 2017; Smith & Conrey, 2007).

Based on a theory on the dynamics of norms, an agent-based model was developed that allows addressing the questions of the behavioral and social dynamics of norm change. It is presented in Chapter 4. Development of the agent-based model demanded for making the theory explicit in computational terms and applying it to a situational context.

2.3.3 An Experimental Approach to Norm Dynamics

An experimental approach to studying the dynamics of norms completes the methodological triad of the present dissertation. Empirical research has shown that norms promote behavioral change (Goldstein et al., 2008; Keizer et al., 2008; Schultz et al., 2007, 2008). At the same time, they have been shown to stabilize behavior (Szekely et al., 2021). To better understand the

conditions of norm maintenance and change and the potential as well as limitations of norms in contributing to desirable social change, it is important to investigate how they work (Andrighetto & Vriens, 2022).

For investigating norm dynamics empirically, processes (i.e., time-series data) need to be assessed. This represents a route rarely taken in experimental psychological norm research, which tends to focus on the behavioral effects induced by norm interventions (Cialdini et al., 1990; Nolan et al., 2008). Studying norm-behavior effects allows drawing conclusions about the effectiveness of an exact intervention in a specific context. Yet, it does not address the underlying processes of norm change (Andrighetto & Vriens, 2022). The questions regarding how an intervention works and why it was or was not effective therefore remain open. As a result, there is little knowledge about how, when, and why norms change and how they influence behavioral decisions.

An empirical study allows testing theoretical assumptions and drawing confirmatory conclusions (Eberlen et al., 2017). Only by collecting empirical evidence, theoretical assumptions as well as simulation results regarding norm dynamics can be validated. While empirical data are always ‘true’, they do not tell us how they came about or how likely their occurrence was. Agent-based modeling may bridge that gap, generating behavioral patterns by playing through possible paths (Waldherr & Wettstein, 2019). The generated patterns can then be compared to the observed empirical phenomena and thereby offer hypotheses about causal explanations for the underlying mechanisms (Bruch & Atwell, 2015; Jager & Ernst, 2017; Smaldino, 2017). That way, the underlying theoretical assumptions can be supported (or not), and a deeper understanding of empirical data be achieved.

Moreover, empirical results can be used to inform, calibrate, and adapt model assumptions. While the experimental approach is limited by practicability, such as the number of observations, agent-based models are not subject to these limitations. They can be scaled up across time, people, and interactions (Andrighetto & Vriens, 2022), leading to new hypotheses to be tested empirically. That way, in a recursive process, theory, simulation, and data all can be improved (Lorenz et al., 2021).

For empirically testing theoretical assumptions and simulation results, an experimental design to studying norm change was developed and a first experiment conducted. To increase transferability of agent-based model and study, both were designed to be tightly interwoven. The study is presented in Chapter 5. It addresses the theoretical assumption that different types

of norms differ in their temporal dynamics. The experimental data were additionally used to be compared to simulated data from the agent-based model, which is presented in Chapter 6.

3. NORMS AND NORM DYNAMICS IN THEORY

Norms have been of great interest to researchers from various scientific disciplines, including psychology, philosophy, sociology, economics, political sciences, and law. The fields of economics and psychology have grown closer over the past years. Their norm research can be seen as building on each other and complementary. Therefore, the following chapter includes research from both disciplines, striving for an interdisciplinary understanding of norms, their dynamics, and effects. It starts with presenting norm definitions and taxonomies (Section 3.1). Next, theoretical work on norms is outlined (Section 3.2). Subsequently, the research article that is part of the present dissertation is summarized, presenting a norm taxonomy, a conceptual model on the interrelatedness of different types of norms as well as decision-making, and an overview of existing empirical norm research (Section 3.3). Lastly, the norm terminology used in the present work is presented (Section 3.4).

3.1 Norm Definitions and Taxonomies

Norms have been defined as shared beliefs about how to behave enforced by sanctions and rewards (e.g., Schwartz & Howard, 1982) or shared rules of conduct that are sustained by approval and disapproval (Elster, 1989). The fact that norms are considered socially shared and socially enforced is emphasized in the construct of social norms. Another aspect often related to social norms is that they are subjective or perceived, meaning that they are properties of individuals (Ajzen, 1991; Bicchieri, 2006; Schwartz & Howard, 1982). Some authors have added the idea of a statistical regularity to the definition of social norms, describing them as “a predominant behavioral pattern within a group, supported by a shared understanding of acceptable actions and sustained through social interactions within that group” (Nyborg et al., 2016, p. 42). Similarly, Bicchieri and Dimant (2019, p. 447), stated that:

A social norm is a rule of behavior such that individuals prefer to conform to it on condition that they believe that (a) most people in their reference network conform to it

(empirical expectation) and (b) that most people in their reference network believe they ought to conform to it (normative expectation).

This definition states two prerequisites for a behavioral rule to be a social norm: the empirical and normative expectation. This limits the phenomena regarded as social norms. For instance, if only the normative expectation is met as in a shared moral or religious norm, Bicchieri and Dimant would not call that a social norm. What all the presented definitions of social norms have in common is that they define social norms by their social enforcement mechanisms. Hence, they are by definition complied with due to social pressure (either real or imagined).

The empirical and normative expectations (cf. Bicchieri, 2006; Bicchieri & Dimant, 2019) relate to two aspects or types of social norms that have often been differentiated (Cialdini et al., 1990; Deutsch & Gerard, 1955; Thøgersen, 2006). Cialdini and colleagues (1990) introduced them as injunctive and descriptive norms in their *focus theory of normative conduct*. The descriptive norm refers to what is considered normal in the sense of an empirical regularity, meaning what most others do (corresponding to the empirical expectation). The injunctive norm describes what is considered normatively appropriate, meaning what should be done (corresponding to the normative expectation). In contrast to the above given definitions by Nyborg (2018) and Bicchieri and Dimant (2019), Cialdini et al. (1990) considered the injunctive and descriptive norms as different types of social norms, being “conceptually and motivationally distinct” (p. 1015). This was supported by Jacobson et al. (2011), showing that the different norm types are associated with different goals. While descriptive norms motivate action by providing evidence of which behavior is effective and adaptive, injunctive norms motivate through the expectation of gaining or maintaining social approval.

Apart from social norms, it has been assumed that individuals internalize norms, developing their own personal norms (Ajzen, 1991; Schwartz & Howard, 1982; Thøgersen, 2006). The process of how personal norms develop is called internalization (Hoffman, 2000; Kohlberg, 1984; Ryan & Deci, 2017). Personal norms have most famously been defined by Schwartz (1977) as the self-expectation to show a behavior and associated with the feeling of moral obligation as well as negative emotions such as „guilt, self-deprecation, loss of self-esteem, or other negative self-evaluations” when violated (p. 231; see also Schwartz & Howard, 1981, 1982). Yet, definitions in personal norm research vary from feelings of personal obligation (Bamberg & Schmidt, 2003), to feelings of guilt (de Groot & Steg, 2009), or behavioral habits (Schahn & Bertsch, 2003). They have also been referred to as personal normative beliefs, referring to an individuals’ normative beliefs regarding appropriate behavior

(Andrighetto et al., 2015; Bicchieri & Xiao, 2009; Szekely et al., 2021), or moral norms/rules, relating personal norms to morality (Bicchieri & Dimant, 2019; Nyborg, 2018; Thøgersen, 1999).

Like social norms, personal norms have often been defined by their enforcement mechanisms, which are assumed to be more internal. For instance, Bicchieri and Dimant (2019) defined a moral rule as not depending on social conditionality. It is rather complied with because one believes it is the right thing to do. This idea was elaborated by Thøgersen (2006) in his extended norm taxonomy, defining a type of norm by its underlying motivation. While personal norms are considered more internal to the individual than social norms, Thøgersen again distinguished different types of personal norms. They may either be superficially internalized (i.e., an introjected norm), finding expression in experiencing guilt or pride, or fully internalized (i.e., an integrated norm), making any enforcement superfluous. Other than Schwartz, Thøgersen thus stated that personal norms may either be internally enforced or not at all.

Reviewing existing definitions and norm taxonomies shows that there are no common definitions of the different types of norms (Conte et al., 2014). This goes along with a large norm vocabulary (see Section 3.4), most of which the reader has been spared here, including conventions, rules, obligations, laws, prescriptions, and proscriptions, contributing to conceptual confusion. A clear and distinct definition of norms and its different types builds the foundation for successful interdisciplinary norm research. This includes disentangling the subject who holds the norm and how a norm is enforced. Only by representing the different components individually, their potential covariation can be shown as well as their independent effects and interactions investigated.

3.2 Theoretical Work on Norms

Norms are considered an important factor in decision-making. Hence, various theories of decision-making and action include norms. One of most influential and best validated psychological theories of decision-making, aiming at explaining and predicting conscious behavioral decisions, is the *theory of planned behavior* (Ajzen, 1991). Therein, subjective norms are considered one of the three components influencing the behavioral intention, resulting in behavior, apart from the attitude and perceived behavioral control. The subjective norm is defined as the perceived expectation of significant others concerning the behavior and

a person's motivation to comply with it (see also Fishbein & Ajzen, 1981) and thus relates to the concept of an injunctive norm (see Section 3.1).

Another decision-making theory, famous for understanding how personal norms influence behavior, is the *norm activation model* (Schwartz, 1977; Schwartz & Howard, 1981, 1982). It was introduced as a decision-making theory for helping behavior and later transferred to a variety of behavioral domains, such as environmental behavior (Bamberg et al., 2007; Davison et al., 2014; de Groot & Steg, 2009; Hunecke et al., 2001). In the norm activation model, personal norms are the only direct determinants of behavior. Yet, the effect of personal norms on behavior may be diverted through various defense mechanisms, neutralizing the feeling of obligation and thus preventing the according action. Personal norms in turn are influenced and may be activated, depending on a number of factors (while the number differs between versions of the theory and its applications). Mostly two factors have been shown to influence personal norms, being the ascription of responsibility and awareness of consequences (Bamberg, 2013; de Groot & Steg, 2009; Steg & de Groot, 2010). They again are assumed to be influenced by different values in the *value-belief-norm theory*, being an extension of the norm activation model (Stern, 2000).

Several authors have combined both decision-making theories (i.e., the theory of planned behavior and the norm activation model), showing that social norms influence personal norms (Onwezen et al., 2013; Shin et al., 2018; Thøgersen, 1999; see also the *stage model of self-regulated behavioral change*, Bamberg, 2013). The relationship between social and personal norms was also subject to theories about norm internalization. *Moral development theories* (Kohlberg, 1964; Piaget, 1932/1965) describe how an individual's moral reasoning develops over time, moving from the imitation of social norms to more mature stages, in which self-chosen moral principles are obtained through "moral 'cognition' (judgement and reasoning)" (Kohlberg, 1978, p. 84). Similarly, in social psychology it is assumed that acquiring norms is based on social cognition and social learning (Howard & Renfrow, 2003; Kelly & Davis, 2018; Theriault et al., 2021). This can be translated into the idea of social norms influencing the development of personal norms (see also the *social cognitive theory*, Bandura, 2001; *moral socialization theory*, Hoffman, 2000; *social learning theory*, Miller & Dollard, 1941; *group norm theory*, Sherif & Sherif, 1953). In the *self-determination theory*, internalization is described as the integration of external values and regulations into the self, enforced through punishments and rewards (Deci & Ryan, 1985; Ryan & Deci, 2000, 2017). The external regulations then become a motivation in itself and thus "self-determined". This

idea was revisited in Thøgersen's (2006) norm taxonomy as presented in the previous section (see Section 3.1).

The presented overview illustrates that psychological and adjacent theoretical works on norms are rarely concerned with change over time. The presented theories tend to focus on (a) the effects of social and personal norms on behavior or (b) the effect of social norms on personal norms. There is a lack of theory stating how decision-making and norms dynamically interact over time (Conte et al., 2014), with theories rather providing snapshots on how a single behavioral decision is made. Moreover, there is little theoretical work on how different types of norms are related (apart from the effect of social norms on personal norms) with respect to a clean division of different types of norms (see Section 3.1). Lastly, there is a lack of a psychologically plausible explicit theory concerning the norm internalization process (Hollander & Wu, 2011; Neumann, 2010b).

3.3 Paper I: The Effects of Norms on Environmental Behavior

The first research paper that is part of the present dissertation titled *The Effects of Norms on Environmental Behavior* was accepted for publication by the *Review of Environmental Economics and Policy* (Dannenberg et al., in press). The accepted manuscript is presented in Appendix A. Therein, a refined norm taxonomy, a conceptual model on how different types of norms influence each other as well as the decision-making process, and an overview of empirical norm research with respect to the conceptual model are presented.

The norm taxonomy differentiates norms along four dimensions with all of them being assumed orthogonal to one another. First, the subject who holds the norm (social vs. personal), second the quality of the norm (descriptive vs. injunctive), third the perspective on the norm (objective vs. perceived, i.e., subjective), and fourth its enforcement mechanism (personally vs. socially vs. legally enforced) defines the type of norm. This comprehensive taxonomy builds upon and integrates existing norm taxonomies, while addressing the problem that existing psychological and economic norm definitions often overlap, focus on specific aspects, and/or entangle different dimensions such as associating the subject of the norm with specific enforcement mechanisms (see Section 3.1). While, for instance, Bicchieri and Dimant (2019) defined a social norm as the concurrence of empirical (i.e., descriptive norm) and normative expectations (i.e., injunctive norm), including the mechanisms that enforce it (see Section 3.1), the present taxonomy represents each of these elements separately. Any enforcement mechanism is disentangled from the actual definition of a norm, while 'social' and 'personal'

refer to the subject who holds a norm, rather than the reason why someone follows a norm. By representing the characteristics of norms individually, a clean division of the different types of norms is achieved. This should allow for better investigating the direct effects of each type of norm as well as addressing questions regarding their interactions.

Based on the presented taxonomy, a conceptual model on the interconnectedness of different types of norms on the individual decision-making level as well as the societal level was developed. Central to the conceptual model are the following four types of norms: *social injunctive norms* ('what most others consider (in)appropriate'), *social descriptive norms* ('what most others usually do'), *personal injunctive norms* ('what I consider (in)appropriate'), and *personal descriptive norms* ('what I usually do'). Norm internalization refers to the process of how personal injunctive norms change. The conceptual model provides assumptions on how these different types of norms develop and change over time including the process of norm internalization grounded in theoretical and empirical research from social, cognitive, developmental, and motivational psychology. This aimed at combining and advancing existing theoretical works on norms (see Section 3.2) as well as dynamic models on norms, which tended to focus on one type of norm rather than the dynamic interplay of different types of norms (see Section 4.1). Moreover, it provides psychologically plausible assumptions on how norms are internalized, which so far lacks a clear understanding (Neumann, 2010b).

In the last part of the article, an overview of empirical research is presented addressing the question of which norm-norm and norm-behavior relations of the conceptual model have already been studied. For the overview, psychological and economic papers were considered that studied norms empirically in relation to environmentally friendly behavior, as the societal transition towards a more sustainable future is a particularly relevant, interesting, and active area of norm research. The literature review also aimed at providing policymakers with knowledge about how to effectively promote pro-environmental behavior. The overview revealed that the effects of social norms on behavior are well established (Goldstein et al., 2008; Hamann et al., 2015; Keizer et al., 2008; Nolan et al., 2008; Schultz et al., 2007, 2008), while we still know little about the underlying mechanisms and how different types of norms affect each other (see Section 5.1). Moreover, the process of norm internalization remains up to now a black box, missing experimental approaches. By identifying these research gaps, providing a more granular taxonomy clearly differentiating between different types of norms, and a conceptual model on how they are related, the presented work aimed at contributing to filling these gaps in future research.

3.4 Norm Terminology of the Present Work

The norm terminology that was chosen for the present work is presented in the following. It slightly simplifies the one presented in Dannenberg et al. (in press), while being consistent with the original research that is part of the present dissertation presented in Chapters 4 and 5. The respective norm definitions given in Dannenberg et al. (in press) also apply for the present thesis. Table 1 presents the norm terminology applied in the present dissertation as well as related concepts from literature referenced in the present work along the two central norm dimensions: the subject and the quality of a norm as defined in Dannenberg et al. (in press).

Social descriptive norms and social injunctive norms carry the subject as well as the quality in their name to maximize clarity and understanding. Personal descriptive norms are simply called habits, being a widely known psychological concept not needing further explanation. Therefore, the ‘injunctive’ in personal injunctive norms is omitted. Hence, in the present work the term ‘personal norms’ implies an injunctive quality, being clearly distinguished from the descriptive quality. This terminology will be applied from now on in the present dissertation to organize existing concepts and clarify conceptual confusion.

Table 1

Norm terminology of the present work and related concepts along the quality and subject of a norm

		QUALITY	
		Descriptive ‘is’	Injunctive ‘ought’
SUBJECT	Social	<p>Social descriptive norms</p> <p>Related concepts are <i>descriptive norms</i> (Cialdini et al., 1990), <i>empirical expectations</i> (Bicchieri & Dimant, 2019), <i>conventions</i> (Mahmoud et al., 2014)</p>	<p>Social injunctive norms</p> <p>Related concepts are <i>injunctive norms</i> (Cialdini et al., 1990), <i>normative expectations</i> (Bicchieri & Dimant, 2019), <i>subjective norms</i> (Ajzen, 1991), <i>obligations</i> (Broersen et al., 2001), <i>social constraints</i> (Shoham & Tennenholtz, 1992)</p>
	Personal	<p>Habits</p> <p>Related concepts are <i>personal descriptive norms</i> (Dannenberg et al., in press)</p>	<p>Personal norms</p> <p>Related concepts are <i>personal injunctive norms</i> (Dannenberg et al., in press), <i>personal (normative) beliefs</i> (Szekely et al., 2021), <i>moral norms</i> (Thøgersen, 1999), <i>moral rules</i> (Bicchieri & Dimant, 2019)</p>

4. NORM DYNAMICS IN AGENT-BASED MODELING

Norms have received considerable attention also in agent-based modeling (Andrighetto, Campenni, et al., 2010, Andrighetto, Villatoro, & Conte, 2010; Conte & Castelfranchi, 1995; Neumann, 2010a). Yet, as mentioned in Section 3.3, many existing dynamic models on norms focused on one norm specifically, as will be outlined in the following by shortly reviewing agent-based modeling approaches to studying norms (Section 4.1). Subsequently, the few normative agent-based models that also include norm internalization are presented (Section 4.2). Next, a summary of the research paper part of the present dissertation is given, presenting an agent-based model of the presented conceptual model (Dannenberg et al., in press). It contributes to existing literature by addressing questions regarding the conditions and independent effects of norm internalization over and above other types of norms as well as non-normative behavioral influences (Section 4.3). Lastly, a revised version of the agent-based model is presented, including new results (Section 4.4).

4.1 Norm Research in Social Simulation

To better understand which questions regarding norm dynamics have so far been addressed and which not, a short historical overview of norm research in the field of social simulation is provided, as historically two different research traditions have grown. This includes, apart from the agent-based modeling tradition coming from artificial intelligence that the present work focuses on, game theoretical simulation models. Although both apply an agent-based approach, they can be (loosely) categorized based on the different foci that they have been taken in studying norm dynamics (Neumann, 2014).

On the one hand, game theoretical simulation models have emphasized the development and evolution of norms in large-scale societies over a long lapse of time. This line of research originated with Robert Axelrod's (1986) seminal work on the evolution of norms. Therein, he investigated the conditions under which individuals act in a certain way, which he calls a social norm (relating to social descriptive norms according to the taxonomy of the present work, see Section 3.4). By implementing the possibility of agents punishing each other for norm violations, Axelrod provided evidence for a causal mechanism of norm evolution, maintenance, and change. On the individual level, agents were exclusively driven by the material payoff not

considering norms, which were treated solely as macrolevel epiphenomena (i.e., as the equilibrium that arises). Similar simulation models conceptualizing norms as self-enforcing behavioral regularities can be found in Binmore and Samuelson (1994) and Bowles and Gintis (2004).

More recent research has represented norms explicitly inside agents. This allowed for different dissemination and enforcement mechanisms to be studied, such as reputation (Savarimuthu & Cranefield, 2011), and for implementing learning mechanisms, enabling agents to dynamically adapt to perceived social norms (e.g., Sen & Airiau, 2007). Moreover, the effects of norms on agents' decision-making have been considered (Epstein, 2001). Agents were equipped with an active norm orientation, so to say the will to conform to what the majority does. This imitation mechanism has been shown to contribute to the spreading of norms. Yet, in a large amount of research, social conventions – corresponding to social descriptive norms in the present taxonomy (see Section 3.4) – are the only type of norm considered (for a review see Mahmoud et al., 2014).

On the other hand, agent-based models coming from the tradition of artificial intelligence focus on the cognitive representation of norms within agents (Neumann, 2008). They are characterized by rich mental architectures and representing heterogeneity between agents. In 1992, Shoham and Tennenholtz introduced the new approach to simulating norms by operationalizing them as social constraints to agents' decision-making, corresponding to social injunctive norms according to the presented taxonomy (see Section 3.4). The authors were able to show that the knowledge about how one should behave facilitates coordination in social systems (see also Moses & Tennenholtz, 1996; Shoham & Tennenholtz, 1995). Agents' social constraints were implemented as strict built-in behavioral laws. Hence, agents were unable to violate them, since being incapable to deliberate about them. Conte and Castelfranchi (1995) introduced social norms as mental objects. They placed them inside of autonomous agents that were able to reason upon normative concerns, deliberately consider them in their behavioral decisions, and choose whether to comply with them or not (see also Castelfranchi et al., 2000; Saam & Harrer, 1999).

This approach was elaborated in the *Belief-Obligations-Intentions-Desires* architecture (Broersen et al., 2001), wherein obligations represent social injunctive norms according to the present taxonomy (see Section 3.4). With agents mediating between personal goals and obligations and resolving conflicts between the two, a new level of complexity was introduced (Neumann, 2010b). Yet, obligations were designed offline (i.e., represented statically) and thus

agents did not adapt to societal changes. In fact, few authors have been concerned with the dynamics of the social injunctive norms (examples are Andrighetto, Villatoro, & Conte, 2010; Verhagen, 2001). Those who have, have often also been concerned with norm internalization. The *EMIL (Emergence In The Loop)* model provides such a framework, explaining how normative beliefs (corresponding to social injunctive norms, see Section 3.4) are developed through processing observed or communicated social information (Andrighetto, Campenni, et al., 2010). It was expanded by including norm internalization, as will be elaborated in the following section.

By applying the norm taxonomy of the present work (see Section 3.4), the presented overview of norms in social simulations illustrates that mainly two approaches have been taken, studying different types of norms: one was mainly concerned with the emergence of social descriptive norms and the other with the mental representation (and partly creation) of social injunctive norms. Although these lines of research have partly converged over the past years (Neumann, 2008), few simulation models represent different types of norms. Including different types of norms allows investigating the independent effects of each norm and the dynamic interplay of different norm types across time.

4.2 Norm Internalization in Agent-Based Modeling

Generally, few normative agent-based models include an internalization process of norms. Yet, the importance of internalization for norm compliance has often been declared (Axelrod, 1986; Hollander & Wu, 2011; Mahmoud et al., 2014; Neumann, 2008, 2010b; Saam & Harrer, 1999). Verhagen (2001) investigated the spread of group norms (being social injunctive norms according to the present taxonomy, see Section 3.4) through agents' communication of their internalized norms (corresponding to personal norms). Agents' group norms change depending on what other agents communicate about their internalized norms. When being part of a group, agents tend to internalize the group norm. Hence, Verhagen included two types of norms with both affecting decision-making. Yet, norm internalization was merely based on the choice whether to adopt the group norm or not. Neumann (2008) has raised the question, whether this adoption process is able to capture the essence of norm internalization according to Wallis and Poulton (2001), who stated that internalization rather describes taking something and turning it into something own. Also, norm internalization was measured by the divergence of group and individual norms, arguably representing the effect of norm internalization, but not the process itself (Neumann, 2008).

Andrighetto, Villatoro, and Conte (2010) expanded the above-mentioned EMIL architecture and presented a rich cognitive model of norm internalization (the *EMIL-I-A* model: *EMIL-Internalizer-Agent*). Therein, internalized normative goals (corresponding to personal norms, see Section 3.4) were added over and above normative beliefs (corresponding to social injunctive norms, see Section 3.4). The conceptualization of normative goals broadened previous approaches as they may be prohibiting, prescribing, or permitting. The translation of normative beliefs into internalized goals (i.e., internalization) was dependent on two parameters: the salience of the normative belief and a cost-benefit analysis. A dichotomous parameter determined whether an internalized goal is formed. Once it is formed and the normative belief is still salient, the EMIL-I-A agent stops the normative deliberation and complies with the internalized goal, using it to save time and calculation in the decision-making process. This conceptualizes internalization as an automatism similar to a habit, making other calculations superfluous and disregarding non-normative influences in decision-making. Similarly, Epstein (2006) has described internalization as blind conformism with a norm. The EMIL-I-A architecture was further developed by Villatoro and colleagues (2015), making the internalization process a multi-step process, rather than a binary question. Therein, only the deepest level of internalization led to complete norm compliance. As Verhagen (2001) has pointed out, the ability to violate an internalized norm is an essential aspect of norm internalization.

The presented literature on agent-based modeling approaches to norm internalization leaves several questions open. The independent effects of norm internalization remain to be investigated, while disentangling conceptually distinct constructs and representing them separately, such as habits and personal norms. This could be informed by a psychological theory of norm internalization (Neumann, 2010b), being embedded in a theory of decision-making representing normative and non-normative behavioral influences. As Verhagen (2001) stated, a central aspect of norm internalization is that it does not necessarily and in every situation lead to the corresponding behavior. Psychological empirical research supports the idea of an imperfect relation, showing that the correlation between personal norms and behavior is small to medium sized (e.g., Bamberg et al., 2007; Hines et al., 1987). A plausible explanation for the imperfect relation between personal norms and behavior might be that there are multiple internalized norms, simultaneously influencing decisions. Several, potentially conflicting personal norms may explain why one internalized norm does not always lead to the corresponding behavior. Yet, the presented simulation models only allow for one norm to be internalized at any one time. By allowing multiple norms to be internalized, inconsistencies

between personal norms and behavioral decisions may be investigated, such as when they arise and how they can be resolved (Batzke & Ernst, 2022). Moreover, the conditions of when and why a certain norm is internalized remain to be examined. Psychological theory on norm internalization states the importance of interindividual differences (Ryan & Deci, 2017; Vygotsky, 2004). By including personality differences that interact with the social context conditions of norm internalization may be investigated.

4.3 Paper II: Conditions and Effects of Norm Internalization

The second research paper that is part of the present dissertation titled *Conditions and Effects of Norm Internalization* was published by the *Journal of Artificial Societies and Social Simulation* (Batzke & Ernst, 2023c) and is shown in Appendix B. Therein, an explicit dynamic theory of norm-based decision-making is presented. The theory is largely based on the conceptual model of norm effects, interconnectedness of different types of norms, and norm internalization presented in Dannenberg et al. (in press; see Section 3.3). It was implemented into the agent-based *DINO* model (*Dynamics of Internalization and Dissemination of Norms*) and analyzed regarding the conditions and effects of norm internalization. Yet, implementation of the theory into an agent-based model demanded specifying and refining the conceptual model.

The paper presents a psychological theory of norm-based decision-making, including goals, social injunctive norms, social descriptive norms, habits, and personal norms. The theory is based on an expectancy-value model (cf. Ajzen, 1991; Fishbein & Ajzen, 1981), representing learning of situational expectations as well as interindividual differences via values. Grounded in personality psychological theory, seven agent types were developed that represent the range of interpersonal differences that the model may portray. The seven agent types categorize into cooperators, conditional cooperators, and defectors (cf. Fehr & Fischbacher, 2002; Fischbacher et al., 2001). By considering interindividual differences, the *DINO* model allows investigating their role in decision-making and norm internalization.

While situational expectations of goals and social norms were assumed to be adapted quickly by the agent, the *DINO* norm internalization process was assumed to be a slower process in decision-making. Norm internalization was assumed to store and aggregate part of the situational learning, interacting with interindividual differences. Personal norms, the product of norm internalization, were conceptualized as behavior-specific rules. This introduces a new level of complexity in agent-based models on norm internalization as it assumes the

existence of a personal norm for each behavioral action. The DINO model simulates three agents playing a 3-person prisoner's dilemma game (see Section 2.2). Hence, agents possess two personal norms: one for cooperation and one for defection. Each personal norm was assumed to develop independently and may be approved or disapproved. Depending on approval and disapproval, personal norms may strengthen or inhibit a behavioral action, influencing decision-making on a higher level (similar to values rather than social norms, goals, or habits).

Regarding the question when and why a norm is internalized, simulation results showed that approval of the cooperation norm is achievable by lots of different agent types, valuing different goals, and in a variety of different social settings. This effect is supported by the empirical finding that cooperative behavior may result from different motivations, such as environmental behavior may result from environmental as well as economic concerns (Brandon & Lewis, 1999; Thøgersen, 2003). DINO cooperator and defector agents tended to internalize a norm according to their value structure, while conditional cooperators showed to be strongly adaptive to the social setting (cf. Burlando & Guala, 2005). More cooperative conditional cooperators increased cooperation with cooperative others and defection with defective others (cf. de Oliveira et al., 2015; Fischbacher & Gächter, 2010). Yet, more defective conditional cooperators produced a behavioral pattern shown by Fischbacher et al. (2001) called “hump-shaped”. It describes conditional cooperation and a decay of cooperation once others cooperate above a certain degree. This is opportunistically rational in the game context. The DINO model showed that this effect may be caused by approving of the defection norm. It also showed that this phenomenon may be prevented when playing with conditional cooperators or defectors rather than cooperators. It prevents more defective agents from following through on their individualist and competitive goals and makes it difficult to free ride, as was similarly argued by Gächter and Thöni (2005). Across all agent types, simulation results showed that playing with cooperative other players facilitated norm approval. That includes approval of the cooperation as well as the defection norm.

Regarding the behavioral and social effects of norm internalization, the DINO model showed that norm internalization encouraged norm compliance and had diverse effects on behavioral stability and payoff equality. Although norm compliance generally increased through norm approval, relating, for instance, to Andrighetto, Villatoro, and Conte (2010), the DINO model suggested that norm disapproval had even stronger behavioral effects. In norm-based intervention studies, behavior-specific norms were rarely differentiated. Research predominantly focused on fostering norm approval (e.g., Hamann et al., 2015; Terrier &

Marfaing, 2015). Moreover, the DINO model showed that the relative importance of different personal norms matters – an argument relating to *goal-framing theory* (Lindenberg & Steg, 2007) not yet examined in the context of norms.

Norm internalization increased behavioral stability and payoff equality in the DINO model in cases of collective approval of the cooperation norm and disapproval of the defection norm. Hence, internalizing norms that support cooperation promoted behavioral persistence (cf. Andrighetto, Villatoro, & Conte, 2010) and social equality among agents (cf. Conte & Castelfranchi, 1995). Yet, approval of the defection norm tended to have contradictory effects in DINO simulation runs, increasing inequality (cf. Saam & Harrer, 1999) and behavioral instability. The DINO model norm internalization process offered possible explanations for the contradictory effects of norms on equality. Results indicated that equality in payoffs can be fostered through collective internalization of the same norm (with exception of approval of the defection norm). Moreover, simulations showed that internalized norms may and may not look like habits on the behavioral level, only partly supporting the assumed connection of habits and norm internalization (Epstein, 2001).

The presented research comes with several limitations, such as the simplicity and stability of the situational framework, specificity for the small group context, free parameters that miss external validation, missing heuristic decision-making processes, and so forth. Yet, the DINO model contributed to existing norm internalization research by combining the depth of psychological theories with social simulation. The model has proven a good candidate to replicate empirical findings, to provide insights to the underlying interactions of internal and external phenomena, and to generate testable hypotheses for future research.

4.4 DINO Model 1.1: Results from a Revised Model

Model reviewing during successive research stages revealed a flaw in the implementation of the DINO model. In the original DINO model 1.0, the mathematical logic on how norm internalization influences decision-making led to an artificial asymmetry in the model dynamics. Therefore, the formula describing the influence of personal norms on decision-making was revised in DINO model 1.1. The DINO model 1.1 and sensitivity analyses (on variations of start values, change rate parameters, and the influence of the different motivational factors) are available at: <https://www.comses.net/codebases/1bb193d4-6c9e-4f19-84d1-b95e2780e9ed/releases/1.1.0/>. In the following, first the mathematical problem and the

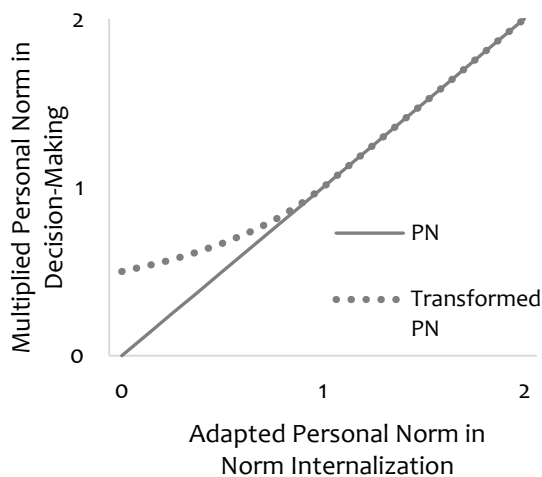
implemented revision in DINO model 1.1 (Section 4.4.1) and second new model results are presented (Section 4.4.2).

4.4.1 Revision from DINO Model 1.0 to 1.1

The mathematical assumption on how personal norms influence decision-making in the DINO model is that the expectation-value products of each motivational factor (i.e., goals, social norms, and habits) are multiplied by personal norms. A personal norm of 1 therefore relates to a neutral value, not influencing decision-making, while values larger than 1 strengthen the intention to show an action and vice versa, values below 1 weaken the intention. To represent a linear norm internalization process, adapting the personal norm value each time step, the range of personal norms was set to $[0;2]$. However, mathematically, multiplying by 0 and 2 are not equivalent, which led to an unintended asymmetry. The problem was solved in DINO model 1.1 by transforming the personal norm for the value range between $[0;1]$ before it is multiplied in the decision-making function. For personal norm values ≤ 1 , it is multiplied as: $1 / (2 - \text{Personal Norm})$ (see dotted line in Figure 1). That way, the personal norm value is still adapted linearly in norm internalization (see solid line in Figure 1), yet the transformed personal norm value in decision-making has the properties of ranging between $[0.5;2]$ (i.e., halving or duplicating the intention for a behavioral action) and its neutral midpoint being 1 (i.e., not influencing the decision).

Figure 1

Transformation of the personal norm for multiplication in decision-making in the DINO model 1.1



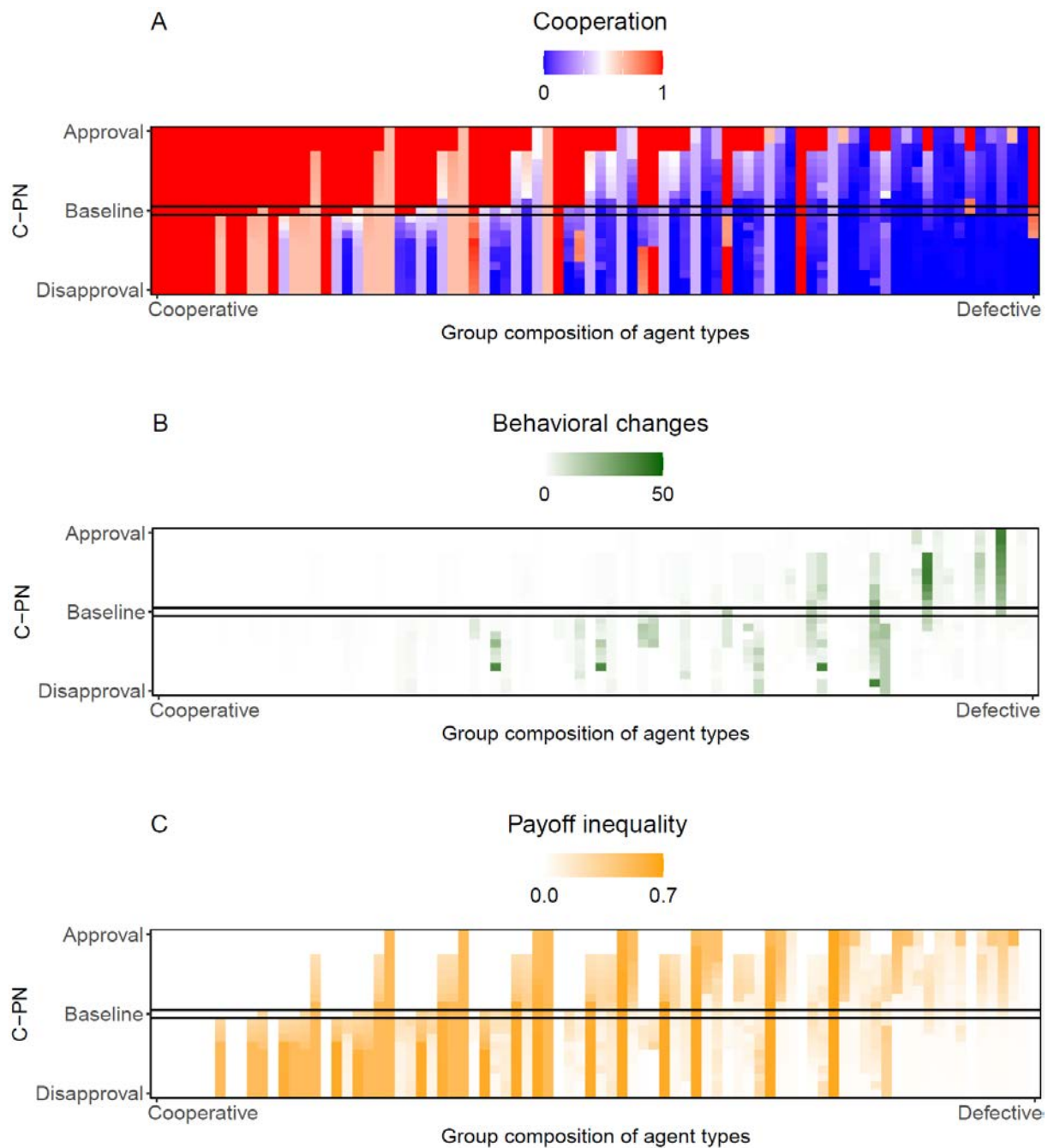
Note. PN = Personal norm.

4.4.2 Results from DINO Model 1.1

In Batzke and Ernst (2023c), two research questions were presented, addressing (a) the conditions and (b) the effects of norm internalization. As the results regarding the effects of norm internalization diverge from the ones presented in the paper, the same two experiments addressing the effects of norm internalization are presented in the following. Simulation results from DINO model 1.1 are compared to the ones from DINO model 1.0 (Batzke & Ernst, 2023c). Testing for the robustness of the simulation results, experiments were repeated with noise in agents' initial start values. These results are presented in Appendix C. They generally show robust behavior against minor variations in start values.

In the first experiment, the personal norm to cooperate was manipulated from approval to disapproval (see Figure 2). Hence, in the beginning of a model run the personal norm was once manipulated within its boundary values. Figure 2 shows the effects of having internalized the personal norm to cooperate depending on the group composition of agent types (ordered along the cooperativeness of the group), while results are aggregated across 200 time steps and all three agents. The group composition refers to the three agents playing the 3-person prisoner's dilemma game. In total, 84 different group compositions were analyzed as the implemented seven agent types can form 84 different groups of three. In the second experiment, the personal norm to cooperate and the personal norm to defect were manipulated independently from each other, showing results aggregated across time, agents as well as group compositions (see Figure 3). Results are each shown regarding agents' average cooperation, agents' number of behavioral changes, and inequality in payoffs between agents.

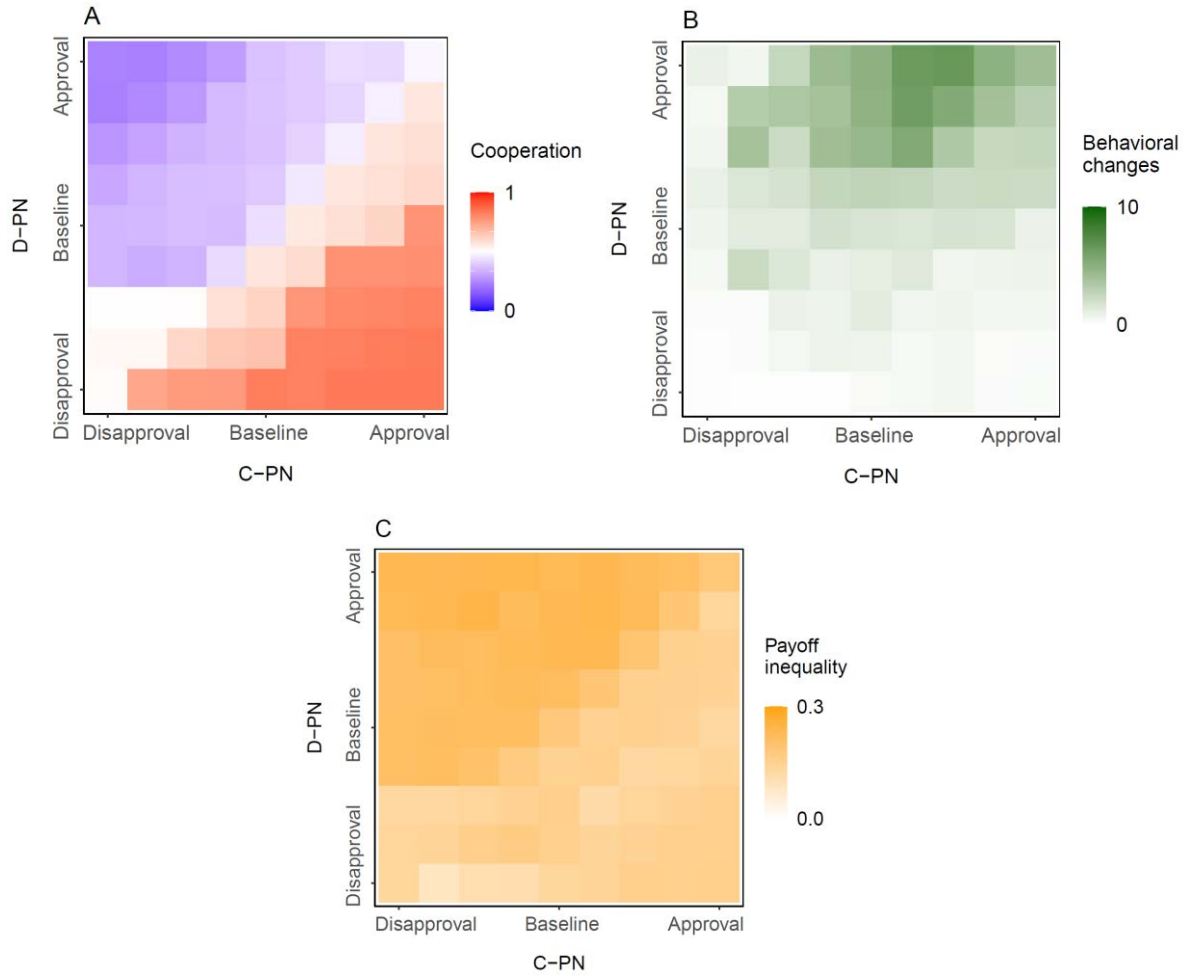
The central difference of the DINO model 1.1 results to the ones from DINO model 1.0 concerns the effects of norm internalization on cooperation (see Figures 2A and 3A). The effect of the DINO model 1.0 that norm disapproval generally has stronger effects than norm approval showed to be an artefact as the DINO model 1.1 did not show this effect. This can be observed when comparing Figure 2A (below) to Figure 5A from Batzke and Ernst (2023c) and Figure 3A (below) to Figure 6A from Batzke and Ernst (2023c). Figure 3A shows that agents' behavior is generally affected similarly by personal norm approval and disapproval. Figures 2B and 2C as well as Figures 3B and 3C (i.e., results regarding behavioral changes and payoff inequality in DINO model 1.1) generally show the same patterns as results from DINO model 1.0.

Figure 2*Manipulation of personal norms to cooperate (C-PN)*

Note. The personal norm to cooperate (C-PN) was varied from full disapproval to full approval. No manipulation was conducted in baseline model runs. Left to right shows agent group compositions, ordered along cooperativeness. Group compositions are defined by three-digit numbers, indicating the three agent types in the group, ordered by digit sum and largest single digit. Effects are shown regarding (A) cooperation (ranging between [0,1], averaged across agents and time), (B) absolute number of behavioral changes, and (C) inequality between agents' individual payoffs (ranging between [0,1], averaged across agents and time). Duration of model runs: 200 time steps.

Figure 3

Manipulation of personal norms to cooperate (C-PN) and to defect (D-PN)



Note. The personal norm to cooperate (C-PN) and to defect (D-PN) were varied from full disapproval to full approval. Effects are shown regarding (A) cooperation (ranging between [0,1], averaged across agents, time, and group compositions), (B) absolute number of behavioral changes (averaged across group compositions), and (C) inequality between agents' individual payoffs (ranging between [0,1], averaged across agents, time, and group compositions). Results were averaged across 84 group compositions of agent types. Duration of model runs: 200 time steps.

5. NORMS AND NORM DYNAMICS IN EMPIRICAL RESEARCH

The following chapter turns to empirical work on norm dynamics. A first overview of economic and psychological empirical norm research was provided in Dannenberg et al. (in press) in relation to the developed conceptual model (see Section 3.3). In the following, a more general overview of psychological empirical norm research is given, showing its emphasis on studying norm effects rather than dynamics (Section 5.1). Next, experimental approaches on studying norm dynamics are presented, also looking beyond the strictly psychological domain (Section 5.2). Lastly, the original research part of the present dissertation is summarized, contributing to psychological empirical norm research by studying the dynamics of different types of norms in a social dilemma game (Section 5.3).

5.1 Norms in Psychological Studies

Psychological studies on norms mostly followed one of the two presented theoretical traditions (see Section 3.2), being the focus theory of normative conduct (Cialdini et al., 1990) and the norm activation model (Schwartz, 1977; Schwartz & Howard, 1981, 1982). As the focus theory of normative conduct states the distinction of injunctive and descriptive quality of social norms, there is a large body of research that investigated their influence on behavioral decisions by varying the situational salience (Cialdini et al., 2006; Göckeritz et al., 2010; Kallgren et al., 2000). For instance, it was shown that participants, who were given the opportunity to litter, were more likely to do so in a littered environment than in a clean environment (Cialdini et al., 1990, Study 1). The variation in the state of the environment was meant to manipulate the perceived descriptive norm for littering (relating to the social descriptive norm, see Section 3.4). The effect of the environment was increased by a person walking by and behaving in accordance with the depicted norm (Cialdini et al., 1990). In a similar manner, Cialdini and colleagues showed that an anti-littering injunctive norm message (“[...] Please do not litter.”, relating to a social injunctive norm, see Section 3.4) reduced participants’ littering behavior in contrast to a non-normative message (Cialdini et al., 1990, Study 5).

Smith and colleagues (2012) manipulated both types of norms orthogonally, aiming at disentangling their independent effects in a laboratory experiment and found that aligned

supportive descriptive and injunctive norms were associated with the highest level of behavioral intentions to engage in energy conservation (Smith et al., 2012; see also Schultz et al., 2007, 2008). But what happens if social descriptive and social injunctive norms conflict with each other? In a seminal field experiment, Keizer and colleagues (2008) addressed this question and showed that a descriptive norm outweighs the effect of an injunctive norm on littering behavior if both are in conflict.

Referred to as the boomerang effect in literature, social descriptive norms have been shown to provoke behavioral assimilation in the sense of an anchoring effect. Schultz and colleagues (2007) communicated descriptive norms regarding energy conservation to households. Households consuming at a higher and a lower rate than the presented norm both assimilated to it. The undesired increase in consumption by those consuming less than the communicated norm prior the norm message is referred to as boomerang effect. Hence, the behavioral adaptation in response to the norm communication can be assumed as in direction of the presented norm (Schultz et al., 2007) and immediate (Cialdini & Goldstein, 2004; Nolan et al., 2008).

The line of research originating from the norm activation model and its extensions (see Section 3.2) evolves around the concept of personal norms. In line with the theory, several factors found large empirical support in influencing personal norms, such as social injunctive norms (under the term “subjective norms”, Bamberg et al., 2007; Hunecke et al., 2001), the attitude (Onwezen et al., 2013), anticipated feelings of pride and guilt (Han, 2014), the ascription of responsibility, and the awareness of consequences (Bamberg, 2013; de Groot & Steg, 2009; Steg & de Groot, 2010). In turn, personal norms have been shown to significantly contribute to explaining variance in a wide variety of behavioral decisions, such as environmental (e.g., Hunecke et al., 2001), economic (Bašić & Verrina, 2021), and cooperative behavior (e.g., Szekely et al., 2021). Even when including the factors from the theory of planned behavior (Ajzen, 1991), meaning attitude, perceived behavioral control, and subjective norms (relating to social injunctive norms along the present taxonomy, see Section 3.4), personal norms additionally contributed to explaining variance in behavior (e.g., Conner & Armitage, 1998; Harland et al., 1999; Shin et al., 2018). As there are few experimental approaches to studying personal norms (yet one example is: Steg & de Groot, 2010) with research being mostly survey based, there is little experimental evidence on the independent effects of personal norms.

To sum up, psychological norm research has repeatedly shown the importance of different types of norms on behavioral decisions or investigated factors that influence personal norms. Yet, there are fewer experimental approaches on studying the dynamics of norms from psychology and adjacent disciplines. In the following section, three experiments are presented, focusing on those that explicitly measured norms.

5.2 Experiments on the Dynamics of Norms

Szekely and colleagues (2021) investigated norm and behavior change under collective risk. Participants played a 28-day collective-risk social dilemma game in groups of six participants. Each day participants were randomly allocated in a group, decided about contributing hypothetical money to collectively reach a certain threshold, and rated their norms. Results showed that social norms (corresponding to a combination of social descriptive and social injunctive norms, see Section 3.4) causally affected behavior and were adapted along the within-subjects (high vs. low) collective risk manipulation. While the effect of the experimental manipulation on personal beliefs (corresponding to personal norms, see Section 3.4) was not reported, descriptive results seemed inconclusive. Yet, personal beliefs significantly increased the explained variance in behavior over and above social norms by 9%.

Bicchieri and colleagues (2022) investigated the influence of observing other participants' actions in a repeated take-or-give donation game on cooperation. Therein, participants could decide each round in a non-strategic setting whether they wanted to take money from a charity, give money to it, or do neither. Results from experiment four showed that normative expectations (relating to social injunctive norms, Section 3.4) significantly decreased after the game, while personal normative beliefs (relating to personal norms, Section 3.4) did not change during the average game duration of 10 minutes (Bicchieri et al., 2022, Appendix). These results raise the question whether personal norms do in fact change over time.

Results from Tverskoi and colleagues (2023) lead to suggest that they do. The authors studied the influence of norms, as well as other social, cognitive, and material behavioral determinants of cooperative behavior. Participants played an online nonlinear common pool resources game for 35 days either with or without messaging. Personal norms as well as normative expectations and empirical expectations (relating to social injunctive and social descriptive norms, see Section 3.4) all significantly increased towards higher resource extraction. The authors additionally showed that personal norms and empirical expectations had the largest behavioral impacts with material benefits and normative expectations being less

important. Therefore, it was concluded that “[...] one can hardly understand social behavior without understanding the dynamics of personal beliefs and beliefs of others [...]” (p. 10). The results allow assuming that personal norms do change. But how?

5.3 Paper III: Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation

The third research article that is part of the present dissertation titled *Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation* was submitted for publication (Batzke & Ernst, 2023b). The manuscript is shown in Appendix D. Therein, the temporal dynamics of different types of norms were investigated, using the temporality as an indicator of different underlying processes of norm change. As introduced in Batzke and Ernst (2023c), the theoretical assumption was investigated that an individual adapts its social norms (referring to both social descriptive and social injunctive norms) quickly whenever the social situation changes, while personal norms are adapted more slowly and gradually, resulting from a more complex learning process that is affected by situational as well as personal factors. In the context of the present article, the term ‘personal norm change’ rather than ‘norm internalization’ was used. This is because the average duration of the experimental game was 20 minutes and hence a rather short period of time. While the underlying process is assumed to be qualitatively similar, norm internalization is rather used for longer time spans.

To test the assumptions, an experimental design that allows investigating the temporal dynamics of norms in a social dilemma game was introduced. In an online experiment, participants played a repeated social dilemma game with artificial co-players. The co-players represented an experimentally controlled, yet realistic social setting, which differed depending on the experimental group. The cooperative group was characterized by a majority of cooperative co-players, the defective group more defective co-players. The overall cooperativity was assumed to affect slow changes in personal norms. Additionally, the social situation changed repeatedly over time within each experimental group, which was assumed to affect quick changes in social norms. Participants’ personal and social norms were repeatedly assessed throughout the game.

The results provided evidence that personal norms and social norms differ in their temporal dynamics, possibly indicating qualitatively different underlying processes of norm change. As predicted, social norms changed quickly, mirroring the social situation, and they were explained by the experimental group. Also as predicted, personal norms changed more slowly, and they were explained by situational and personal factors as well as self-oriented social norms, being the expectation of others regarding the individual's behavior. Yet, the assumed linear trends towards cooperativity (in the cooperative group) and defectivity (in the defective group) could not be found. Generally, personal norms rather developed towards more defectivity in both groups. While a significant difference in personal norms between groups was found during the game, it was lost after the game. Several theoretical and methodological potential reasons for these results were discussed, while *prospect theory* (Tversky & Kahneman, 1992) offers an especially plausible explanation: negative experiences loom larger than positive ones. Accordingly, personal norms might develop asymmetrically regarding positive and negative experiences (see Chapter 6).

Based on the study results, it yet remains unclear whether personal norm change was due to belief change or change in the norm activation level (in the sense of the norm activation theory, cf. Schwartz, 1977), and how it can be directed towards more cooperativeness in the long run. As personal norms showed to be highly predictive of behavioral decisions once again (e.g., Tverskoi et al., 2023), it seems relevant further investigating the mechanisms of personal norms change.

6. PREVIEW ON COMPARING NORM DYNAMICS IN AGENT-BASED MODELING AND EXPERIMENT

The present dissertation represents an approach towards the psychological study of the dynamics of norms via theory, agent-based modeling, and experimental research. While these approaches are intertwined with each other, data from agent-based modeling and experiment can also be compared directly. The combination of simulation and experimental data represents a way of externally validating agent-based models. By comparing results from both data, assumptions about underlying mechanisms that generate the observed empirical results can be investigated. This allows testing as well as improving the agent-based model as part of the

scientific process. This process is part of understanding and gaining knowledge about the underlying mechanisms of norm change (see Section 2.3.3). As preliminary work was already conducted on the comparison of DINO model simulation data and experimental data, it is presented in the following, representing – so to say – the closing of the circle of the present dissertation.

The working paper titled *An Experimental Attempt at Validating an Agent-Based Model on Decision-Making, Social Norm Change, and Norm Internalization* was accepted to be presented at the *Social Simulation Conference 2023* (Batzke & Ernst, 2023a). The unpublished manuscript is shown in Appendix E. It provides a first attempt at comparing data on norm internalization from the agent-based DINO model with time-series experimental data. This was made possible due to designing study and agent-based model highly similarly. Both experiment and agent-based model provide data about individuals' behavior, social descriptive norm change¹, and personal norm change (i.e., norm internalization) in the context of the same social dilemma game. Hence, apart from norm internalization processes, the approach also allowed for comparing simulated and experimental behavioral data and social descriptive norm change processes.

To make the comparison, the agent-based model was modified so that one agent plays the game with two predefined behavioral sequences – like study participants did. The experimental design could then be applied to agents as it was to participants. As the present work has shown that interindividual differences have an impact on norm dynamics, simulation and experimental results were analyzed with respect to participants' and agents' interindividual differences represented by the concept of trait cooperativeness. Along the continuum of trait cooperativeness, agents and participants were categorized into three groups: cooperators, conditional cooperators, and defectors.

The comparison showed that DINO agents and study participants categorized as conditional cooperators showed similarities in behavior, social descriptive norm change, and norm internalization, suggesting the plausibility of the assumed mechanisms in the DINO model. This also led to suggest that DINO cooperator and defector agents might be unrealistically extreme, needing stronger resemblance with conditional cooperators. Norm internalization processes showed a major difference between agents and participants: agents' norm internalization was generally more towards cooperativeness. In participants' personal

¹ While in the experiment social injunctive norms were assessed as well, they were represented as constants in the DINO model. Therefore, the comparison could only be conducted for social descriptive norms.

norm change, there was no learning of a cooperative norm under any circumstances, but rather of a defective norm (see Section 5.3). Based on that observation, it was assumed that participants' norm internalization in a social dilemma may be asymmetrical regarding the learning of cooperativeness and defectivity (see Batzke & Ernst, 2023b). Personal norm change may underly a negativity bias, meaning that negative experiences are more strongly (or even qualitatively differently) accounted for than positive ones. Possibly, this also affects norm internalization.

Since that mechanism was so far missing in the DINO norm internalization process, the agent-based model was adjusted by introducing asymmetry in the sense of a negativity bias, making a cooperative norm more difficult to approve in norm internalization than a defective norm. The adjustment improved the overall similarity of agents' and participants' norm internalization processes across agent types and experimental groups, supporting the notion of asymmetry being an aspect in norm internalization in the social dilemma context. While there are numerous limitations to the comparison, including missing statistical validation and lack of generalizability across different situations, the approach presented promising starting points for future research.

7. GENERAL DISCUSSION

The present dissertation represents an attempt at studying norm dynamics in decision-making. In the following, the contribution of the present work to answering the two central research questions is revisited (Section 7.1). Subsequently, limitations of the present work are discussed relating to future research (Section 7.2). Section 7.3 provides a brief outlook to the continuation of the present work and concludes.

7.1 Addressing the Research Questions

In Section 2.1, two main research questions were presented: (1) How do different types of norms change? and (2) How are norms internalized and what are the effects of internalization? To address these questions, targeting a better understanding of the norm dynamics in decision-making, a multi-method approach was chosen, combining a conceptual model, an agent-based model, and an experimental study. First, in the theoretical contribution a conceptual model of

the interrelations of different types of norms as well as their influence of and on decision-making was presented. Second, the conceptual model was refined towards a dynamic norm-based theory of decision-making and implemented into an agent-based model. Via agent-based modeling, the consequences of the theoretical assumptions over time were investigated. Third, an experimental study tested a theoretical assumption concerning differences in norm processes empirically. These three elements partly build on each other, such as the theoretical work was prerequisite for both other elements. At the same time, each of these three elements was intended to individually contribute to addressing the two research questions, as will be discussed in the following sections.

7.1.1 How do Different Types of Norms Change?

The first research question addressed norm change in different types of norms. It incorporated the questions of how, when, and why each type of norm changes as well as the question of potential differences between different types of norms. As existing theoretical frameworks tend to focus on the behavioral effects of norms rather than describing how change in different types of norms occurs (see Section 3.2), the first contribution to addressing these questions was made by introducing a conceptual model on norm change and decision-making (Dannenberg et al., in press). Therein, a taxonomy was presented that clearly differentiates different types of norms, being a prerequisite to address their distinct processes. Based thereon, the presented conceptual model states assumptions grounded in psychological research on how different types of norms influence each other and the decision-making process.

These theoretical assumptions were implemented into the agent-based DINO model representing different types of norms against the backdrop of a psychological theory of decision-making. This allowed investigating the questions of when and why a certain norm changes and examining the individual effects of a norm over time in a dynamic decision-making framework. The DINO model contributes to existing social simulations on norms by presenting a psychologically grounded assumptions on norm internalization, introducing the possibility to internalize more than one norm at the same time, and incorporating interpersonal differences in norm internalization and decision-making (see Sections 4.1 and 4.2). To simulate the theoretical assumptions presented in the conceptual model (Dannenberg et al., in press), the theory was extended and made computationally explicit (Batzke & Ernst, 2023c). The theoretical assumptions stated different temporal dynamics underlying social and personal norms. Social norms were assumed to be adapted quickly, whenever the social situation changes. Personal

norm change was assumed to underlie a more complex norm internalization process, depending on situational and personal factors. Therefore, personal norms were assumed to change more slowly. Simulation results of Batzke and Ernst (2023c) focused on the conditions and effects of norm internalization, which are discussed with respect to the second research question in the following section.

As comparing the conceptual model with empirical norm research showed, there are research gaps concerning the processes of norm change. An attempt at addressing norm change was made in the presented experimental study. Therein, the theoretical assumptions of fast change in social norms and slow change in personal norms (i.e., norm internalization) were experimentally tested (Batzke & Ernst, 2023b). The experiment addressed the question of potential differences in how different types of norms change, using the temporal dynamics of norms as a proxy for different underlying processes of norm change. This contributes to psychological norm research by focusing on norm processes and introducing an approach to examining differences between norm processes experimentally (see Sections 5.1 and 5.2). The results supported the idea of temporal differences in different types of norms. Personal norms changed slower than social norms and were influenced not just by situational (as in the case of social norms) but also personal factors, supporting the assumption of different mechanisms underlying social and personal norms.

The question of how social and personal norms change was further addressed by comparing data from agent-based modeling and the experiment (Batzke & Ernst, 2023a). Comparing both data types is not only an essential step in validating model assumptions but may also provide generative (i.e., model-based) explanations to empirical observations and thus potentially a deeper understanding of the mechanisms of norm change (see Chapter 6). Simulated social descriptive norm change generally showed a high resemblance with empirical patterns, supporting the assumption of a fast adaptation process. Comparing simulated and empirical personal norm change led to developing hypotheses regarding the mechanisms of norm internalization, as will be further elaborated in the following section. While first differences between social and personal norms change could be revealed, understanding the change processes of different types of norms requires further investigation.

7.1.2 How are Norms Internalized and What are the Effects of Internalization?

A special focus of the present work was put on norm internalization with the second research question addressing how norm internalization may function and what its effects are. The theoretical paper presented assumptions regarding how norms are internalized with respect to different types of norms and a dynamic decision-making process (Dannenberget al., in press). The subsequent literature review revealed that the norm internalization process lacks experimental evidence. In Batzke and Ernst (2023c), theoretical assumptions about how norms are internalized and affect decision-making were refined. As mentioned in the previous section, personal norms were assumed to change slower than social norms, resulting from the interaction of situational and personal factors. Personal norms were furthermore stated to influence decision-making on a higher level than motivational factors, such as social norms or goals. Hence, they were assumed to influence decisions on the same level as values, while being able to strengthen as well as weaken behavioral intentions.

The questions regarding the conditions of norm internalization as well as its effects on the behavioral and social dynamics were addressed via agent-based modeling (Batzke & Ernst, 2023c). Simulation results showed that conditions for norm internalization are diverse: different agent types, also defective ones, may internalize cooperation as appropriate, depending on the situation. Simulations further supported the widely accepted notion that norm internalization increases norm compliance, while agents also maintained the ability to violate internalized norms. Personal norms were shown to resemble habits on the behavioral level under specific circumstances, such as approving of the cooperative personal norm, while they may also increase behavioral instability. Moreover, the simulations offered a possible explanation for the inconclusive results of how norms affect social inequality: approving of the personal norm to defect might be one reason.

The experimental study addressed the questions of which factors influence personal norms and of the effect of personal norms on behavioral decisions (Batzke & Ernst, 2023b). Personal norms showed to be influenced by situational factors (i.e., the experimental group) as well as interindividual differences (i.e., participants' trait cooperativeness). Moreover, *self-oriented* social descriptive and social injunctive norms showed to affect personal norms. Hence, what an individual perceives as expectations of others regarding the *own* behavior (vs. general behaviors of others) showed to be strongly related to personal norms, offering a leverage point for future norm-based intervention studies. In turn, personal norms showed highly behaviorally

predictive, emphasizing the importance of better understanding norm internalization. The experimental results however did not support the assumption that personal norms change in a slow upward or downward trend across different social settings – at least not within the short time frame of the experimental situation (see Section 7.2.3).

Norm internalization was further investigated by comparing personal norm change in the agent-based model and the experimental study (Batzke & Ernst, 2023a). The generated data from the agent-based model provide an object of comparison to the experimental data, being based on theoretical assumptions of how norms are internalized (Batzke & Ernst, 2023c). A first comparison showed that agents' norm internalization did not show the empirically observed trend towards defectivity. This led to assume the existence of a negativity bias in norm internalization in the social dilemma context. By implementing asymmetry into the DINO norm internalization process, meaning that internalizing the appropriateness of cooperation is more difficult compared to defectivity, the fit of simulation to experimental data strongly improved. While the comparison suggested one plausible mechanism in norm internalization in a social dilemma situation, our understanding of the complex process of norm internalization remains fragmentary (cf. Conte et al., 2010).

7.2 Limitations and Future Research

The present work comes with several limitations, offering numerous leverage points for future research. In the following, limitations regarding the external validity (Section 7.2.1), situational generalizability (Section 7.2.2), and internal validity (Section 7.2.3) are addressed.

7.2.1 External Validity

The external validity of agent-based modeling results is directly linked to the well-known issue of model validation (e.g., Moss & Edmonds, 2005; Windrum et al., 2007). Simulating a norm-based theory of decision-making demands for a high degree of specificity in assumptions – a degree which exceeds verbal psychological theories by far (Schlüter et al., 2017; Smaldino, 2020). Moreover, it demands for theoretical assumptions about how psychological constructs, such as norms and goals, change over time. These assumptions can be rarely found in psychological norm theory, which predominantly describes how a behavior comes about in one point in time (see Section 3.2). As a result, most agent-based models, including the presented DINO model, are based in part on free parameters, meaning assumptions which could neither

empirically nor theoretically be derived. For instance, what exactly is the level of conviction one needs to feel to approve of a behavior? Making these decisions when implementing a theory affects the simulation results. Hence, this issue not only relates to the validity but also the robustness of the simulation results.

Yet, simulations can contribute to understanding the role of those assumptions in the dynamic interplay by testing their influence. As the implementation of the negativity bias into the DINO model showed (Batzke & Ernst, 2023a), the threshold of level of conviction matters and deserves further investigation. By conducting sensitivity analyses, the effects of parameter variations on simulation results can be tested (Gilbert, 2008). This aims at understanding the conditions under which the simulation model shows the theoretically expected results. That way, agent-based modeling can serve as a tool for theory development and improvement (Jager, 2017). There is much value in an explicit theory, guiding empirical research and allowing to generalize findings (Eberlen et al., 2017). The generative ability of a simulated theory allows investigating the consequences of a set of assumptions over time, in a way that the human mind is unable to do by solely thinking about it (Resnick, 1994; Troitzsch, 2017). This provides testable hypotheses, such as the conditions when norm internalization increases and decreases social inequality. The hypotheses presented in this work yet remain to be examined empirically in future research.

Apart from missing theoretical specificity, a major hurdle for validating agent-based models is often a lack of suitable empirical data on the investigated phenomenon (Smith & Conrey, 2007). Psychological empirical data tend to differ from what is needed for validation or can be assessed (Jager & Ernst, 2017). The conducted empirical study (see Section 5.3) aimed at testing one aspect of the norm theory, being the difference in temporal dynamics between social and personal norms. Yet, there are other untested aspects of the norm theory, such as norm internalization being a higher-level process in decision-making (cf. Piaget, 1970; Vygotsky, 1930/1981)

While untested assumptions are related to uncertainty in agent-based modeling, they can also be understood as blind spots in empirical research. Hence, missing data to validate agent-based models is not only an issue to modelers wanting to validate their models, but also shows a general lack of knowledge on how psychological constructs change or even come about. Psychological research tends to focus on investigating relationships between variables (via correlation-based statistical methods) or group differences, indicating the causal effect of one variable on another. This tends to neglect processes, namely change over time (Jager & Ernst,

2017). Agent-based modeling may facilitate a change in the way of thinking towards processes and thereby open up new questions. That way, in a recursive process, theory, data, and simulation may be improved (Lorenz et al., 2021). The present dissertation represents a first cycle of model development, consisting of model implementation, analysis, validation, and improvement. Many more cycles are necessary to solidify the gained knowledge.

7.2.2 Generalizability

The presented results from agent-based modeling and experiment are specific for a highly simplified and standardized social dilemma situation (Dawes, 1980). The social dilemma applied in the present work considers only three players and is stable across time. While this was chosen as a first test of the presented theory, it limits the generalizability of the results. Real-world situations are often characterized by more than two behavioral alternatives, larger and more dynamic groups, changing behavioral consequences, uncertainty, et cetera. While norms arguably influence decisions in most social situations, one can assume that the context influences how and when norms change (Conte et al., 2014). For instance, it might be possible that the observed negativity bias in participants' personal norm change in the experiment was related to frustration due to the social dilemma situation. It remains for future research to test the presented theoretical assumptions in different situational frameworks and investigate context-specific effects in the dynamics of norms.

In this sense, a transfer to the environmental domain seems particularly relevant. Mitigating global warming and massive environmental destruction represent some of the greatest challenges of the human species (Intergovernmental Panel on Climate Change [IPCC], 2023). The contexts of environmental behavior are typically characterized by a high degree of anonymity and invisibility of the consequences, which are delayed in time and/or located elsewhere (Gardner & Stern, 2002). As this comes with challenges in fostering sustainable behavior, it might also require dealing with norm transgressions differently than the context of a social dilemma game allows. In any case, it demands for specific solutions matching these contexts. Like social conflicts, environmental challenges can also be described in standardized terms via so-called resource or commons dilemmas (Hardin, 1968; Ostrom et al., 2002). Therein, the element of a resource that changes over time, depending on its usage by the agent population, is added to basic conflict represented in the social dilemma. This adds to the complexity of the dilemma since overusing the resource not only affects all other individuals but might also lead to the irreversible destruction of the ecosystem represented by the resource

(Ernst, 2010). As norms strongly influence environmental behavior (Nyborg, 2018), they represent one possible solution to the dilemma situation (Ostrom, 1990, 1999; Thøgersen, 2008) and to preventing the *tragedy of the commons* (Hardin, 1968).

7.2.3 Internal Validity

Theoretical work on norm internalization traces back to developmental psychologists such as Jean Piaget and Lawrence Kohlberg, describing how moral principles are learned via moral reasoning across the period of childhood and early adulthood. Social psychologists have considered it a continuous process, focusing on the social aspect in learning norms (e.g., Kelly & Davis, 2018), similar to motivational psychologists, interested in the motivational underpinnings of internalization (e.g., Ryan & Deci, 2000). These psychological theories, in combination with theorizing from other fields (e.g., Durkheim, 1893; Parsons, 1937; Ullmann-Margalit, 1977), laid the foundation to what is understood as and known about norm internalization. As these theories would suggest, norm internalization is a challenging research object. Being a slow, internal cognitive process, it is arguably rather stable across time and difficult to measure (Neumann, 2010b). In the present experiment (see Section 5.3), although time-series data were assessed, participants' average duration to complete the study was only about 20 minutes. Game duration was even shorter. Moreover, assessing norm internalization was conducted via self-reports. However, it is possible that personal norms are not fully accessible via conscious deliberation about them. These issues raise the questions whether the experiment was able to assess personal norm change and whether experimental data is comparable to simulation data.

Up till now, there is missing experimental evidence on the process of norm internalization. While it may be assumed that a slow process can in part also be observed in a short period of time, there is a need for more and longer empirical studies on norm internalization (cf. Tverskoi et al., 2023) that combines self-report questionnaires with more implicit or indirect measures of personal norms (cf. Bicchieri et al., 2014). In continuation of the present work, further investigating the notion of asymmetry in norm internalization may represent an interesting starting point (see Chapter 6): are cooperation norms easier to disapprove than non-cooperative norms in norm internalization? This idea relates to empirical studies showing that social norm adherence has weaker positive effects, than observing social norm violations has negative behavioral effects (Charness et al., 2019; Thöni & Gächter, 2015). The asymmetry in the behavioral effects of social norms has been shown to disappear by increasing group identification among participants in a donation game (Bicchieri et al., 2022).

Does this finding translate into personal norms? Is personal norm change differently affected by positive and negative experiences? Can (mis)trust in the other players account for potentially different effects?

A combination of long-term studies and agent-based modeling may allow addressing these questions and the complex interplay of person, situation, and communication variables. Modern technology may assist in that endeavor, facilitating the assessment people's norms in the field over longer periods of time via smartphone applications. Better understanding the conditions for and the mechanisms of norm acceptance and rejection processes as well as a potential asymmetry between the two would allow designing interventions that, for instance, specifically target the approval of pro-environmental or disapproval of environmentally harmful norms.

7.3 Outlook and Conclusion

The present work aimed at taking a step towards bridging the gap between agent-based modeling and psychological theory and behavioral experiments, by taking a psychological approach towards studying norm dynamics in decision-making. A psychologically grounded agent-based model on norm internalization and decision-making was presented. It allowed formulating numerous hypotheses concerning the behavioral and social effects of norm change for future research. This was based on a differentiated taxonomy of norms and a conceptual model on the interplay between norms and behavior and norm internalization. An experimental design was introduced that helped disentangling the temporal dynamics of social norm and personal norm change. Experimental findings suggested that there are in fact fast and slow norm change. Finally, combining results from experiment and agent-based modeling led to assume that in the social dilemma cooperative norms are more difficult to internalize as appropriate than defective norms.

Questions concerning norm change and norm internalization were touched upon in this work. Yet, up till now, norm internalization remains a puzzle. Vygotsky stated in his theory of norm internalization that “the barest outline of this process is known” (Vygotsky, 1978, p. 57). This still seems to be relevant today. Many more questions regarding how people internalize norms await consideration (Andrighetto et al., 2014). Which factors elicit norm internalization? In which type of situations are norms internalized? How are new personal norms created? Open questions also go well beyond norm internalization. For policy design, better understanding the role of social and legal authorities in norm change matters. While it seems possible that

authorities may positively influence intended norm change, the effect might as well as backfire and lead to erosion of existing norms. This issue relates to the larger question of how norm dynamics and change in other cognitive elements are interrelated. How do norms change, depending on individuals' change in political opinions, social identities, and values?

Addressing individual level norm dynamics calls for psychological answers. Psychological theoretical, quantitative, and qualitative research is a prerequisite for understanding intrapersonal complexity of psychological phenomena. Social simulation may represent an important added value, deriving complexity from including the temporal and social dimensions of norm change (see Section 2.3.1). It allows studying intrapersonal cognitive dynamics and the two-way relations of cognitive and social norm processes (Neumann, 2014).

Norms regulate most spheres of human activity. Ubiquitous and fascinating as they are, norms continue to receive attention from various scientific disciplines. Yet, social scientists have no common understanding of what a norm is and have applied very different approaches to studying them. The present work has attempted to overcome some of the disciplinary boundaries in norm research and aimed for an innovative approach to studying the dynamics of norms from a psychological perspective. Thinking about norms in terms of their cognitive and social interplay may assist in approaching the next frontiers in norm research.

REFERENCES

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Andrighetto, G., Campenni, M., Cecconi, F., & Conte, R. (2010). The complex loop of norm emergence: A simulation model. In K. Takadama, C. Cioffi-Revilla, & G. Deffuant (Eds.), *Simulating Interacting Agents and Social Phenomena* (pp. 19–35). Springer. https://doi.org/10.1007/978-4-431-99781-8_2
- Andrighetto, G., Grieco, D., & Tummolini, L. (2015). Perceived legitimacy of normative expectations motivates compliance with social norms when nobody is watching. *Frontiers in Psychology*, 6, 1413. <https://doi.org/10.3389/fpsyg.2015.01413>
- Andrighetto, G., Villatoro, D., & Conte, R. (2010). Norm internalization in artificial societies. *AI Communications*, 23(4), 325–339. <https://doi.org/10.3233/AIC-2010-0477>
- Andrighetto, G., Villatoro, D., & Conte, R. (2014). The role of norm internalizers in mixed populations. In R. Conte, G. Andrighetto, & M. Campenni (Eds.), *Minding norms: Mechanisms and dynamics of social order in agent societies* (pp. 153–174). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199812677.003.0010>
- Andrighetto, G., & Vriens, E. (2022). A research agenda for the study of social norm change. *Philosophical Transactions of the Royal Society A*, 380(2227), 20200411. <https://doi.org/10.1098/rsta.2020.0411>
- Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.
- Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review*, 80(4), 1095–1111. <https://doi.org/10.2307/1960858>
- Bamberg, S. (2013). Changing environmentally harmful behaviors: A stage model of self-regulated behavioral change. *Journal of Environmental Psychology*, 34, 151–159. <https://doi.org/10.1016/j.jenvp.2013.01.002>
- Bamberg, S., Hunecke, M., & Blöbaum, A. (2007). Social context, personal norms and the use of public transportation: Two field studies. *Journal of Environmental Psychology*, 27(3), 190–203. <https://doi.org/10.1016/j.jenvp.2007.04.001>
- Bamberg, S., & Schmidt, P. (2003). Incentives, morality, or habit? Predicting students' car use for university routes with the models of Ajzen, Schwartz, and Triandis. *Environment and Behavior*, 35(2), 264–285. <https://doi.org/10.1177/0013916502250134>
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52(1), 1–26. <https://doi.org/10.1111/1467-839X.00024>

- Bašić, Z., & Verrina, E. (2021). Personal norms – and not only social norms – shape economic behavior. *MPI Collective Goods Discussion Paper*, (2020/25). <http://dx.doi.org/10.2139/ssrn.3720539>
- Batzke, M. C. L., & Ernst, A. (2022). Explaining and Resolving Norm-Behavior Inconsistencies – A Theoretical Agent-Based Model. In M. Czupryna & B. Kamiński (Eds.), *Advances in Social Simulation* (pp. 41–52). Springer International Publishing. https://doi.org/10.1007/978-3-030-92843-8_4
- Batzke, M. C. L., & Ernst, A. (2023a). *An Experimental Attempt at Validating an Agent- Based Model on Decision-Making, Social Norm Change, and Norm Internalization* [Unpublished manuscript]. Center for Environmental Systems Research, University of Kassel.
- Batzke, M. C. L., & Ernst, A. (2023b). *Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation* [Manuscript submitted for publication]. Center for Environmental Systems Research, University of Kassel.
- Batzke, M. C. L., & Ernst, A. (2023c). Conditions and Effects of Norm Internalization. *Journal of Artificial Societies and Social Simulation*, 26(1), 1–31. <https://doi.org/10.18564/jasss.5003>
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
- Bicchieri, C., & Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public Choice*, 191, 1–22. <https://doi.org/10.1007/s11127-019-00684-6>
- Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132, 59-72. <https://doi.org/10.1016/j.geb.2021.11.012>
- Bicchieri, C., Dimant, E., Gelfand, M., & Sonderegger, S. (2023). Social norms and behavior change: The interdisciplinary research frontier. *Journal of Economic Behavior & Organization*, 205, A4-A7. <https://doi.org/10.1016/j.jebo.2022.11.007>
- Bicchieri, C., Lindemans, J. W., & Jiang, T. (2014). A structured approach to a diagnostic of collective practices. *Frontiers in Psychology*, 5, 1418. <https://doi.org/10.3389/fpsyg.2014.01418>
- Bicchieri, C., & Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, 22(2), 191-208. <https://doi.org/10.1002/bdm.621>
- Binmore, K., & Samuelson, L. (1994). An economist's perspective on the evolution of norms. *Journal of Institutional and Theoretical Economics*, 150(1), 45-63.
- Bossel, H. (1994). *Modeling and simulation*. A K Peters. <https://doi.org/10.1201/9781315275574>

- Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, 65(1), 17-28. <https://doi.org/10.1016/j.tpb.2003.07.001>
- Brandon, G., & Lewis, A. (1999). Reducing household energy consumption: A qualitative and quantitative field study. *Journal of Environmental Psychology*, 19(1), 75–85. <https://doi.org/10.1006/jevp.1998.0105>
- Broersen, J., Dastani, M., Hulstijn, J., Huang, Z., & van der Torre, L. (2001). The BOID architecture: conflicts between beliefs, obligations, intentions and desires. In *Proceedings of the fifth international conference on Autonomous agents* (pp. 9-16). <http://dx.doi.org/10.1145/375735.375766>
- Bruch, E., & Atwell, J. (2015). Agent-based models in empirical social research. *Sociological Methods and Research*, 44(2), 186–221. <https://doi.org/10.1177/0049124113506405>
- Burlando, R., & Guala, F. (2005). Heterogeneous agents in public goods experiments. *Experimental Economics*, 8(1), 35-54. <https://doi.org/10.1007/s10683-005-0436-4>
- Castelfranchi, C., Dignum, F., Jonker, C. M., & Treur, J. (2000). Deliberative normative agents: Principles and architecture. In N. R. Jennings & Y. Lespérance (Eds.), *Lecture notes in computer science: Vol. 1757. Intelligent Agents VI. Agent Theories, Architectures, and Languages: ATAL 1999*. (pp. 364-378). Springer. https://doi.org/10.1007/10719619_27
- Charness, G., Naef, M., & Sontuoso, A. (2019). Opportunistic conformism. *Journal of Economic Theory*, 180, 100-134. <https://doi.org/10.1016/j.jet.2018.12.003>
- Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. *Social Influence*, 1(1), 3-15. <https://doi.org/10.1080/15534510500181459>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591-621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015–1026. <https://psycnet.apa.org/doi/10.1037/0022-3514.58.6.1015>
- Conner, M., & Armitage, C. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology*, 28(15), 1429–1464. <https://doi.org/10.1111/j.1559-1816.1998.tb01685.x>
- Conte, R., Andrighetto, G., & Campenni, M. (2010). Internalizing norms: A cognitive model of (social) norms' internalization. *International Journal of Agent Technologies and Systems*, 2(1), 63–73. <https://doi.org/10.4018/jats.2010120105>

Conte, R., Andrighetto, G., & Campenni, M. (2014). *Minding norms: Mechanisms and dynamics of social order in agent societies*. Oxford University Press.

Conte, R., & Castelfranchi, C. (1995). Understanding the functions of norms in social groups through simulation. In N. Gilbert & R. Conte (Eds.), *Artificial Societies: The Computer Simulation of Social Life* (pp. 213–226). Routledge.

Conte, R., & Dellarocas, C. (2001). Social order in info societies: An old Challenge for innovation. In R. Conte & C. Dellarocas (Eds.), *Social Order in Multiagent Systems* (pp. 1-16). Kluwer. https://doi.org/10.1007/978-1-4615-1555-5_1

Dannenberg, A., Gutsche, G., Batzke, M. C. L., Christens, S., Engler, D., Mankat, F., Möller, S., Weingärtner, E., Ernst, A., Lumkowsky, M., von Wangenheim, G., Hornung, G., & Ziegler, A. (in press). The effects of norms on environmental behavior. *Review of Environmental Economics and Policy*.

Davison, L., Littleford, C., & Ryley, T. (2014). Air travel attitudes and behaviours: The development of environment-based segments. *Journal of Air Transport Management*, 36, 13-22. <https://doi.org/10.1016/j.jairtraman.2013.12.007>

Dawes, R. (1980). Social dilemmas. *Annual Review of Psychology*, 31, 169–193. <https://doi.org/10.1146/annurev.ps.31.020180.001125>

de Groot, J. I., & Steg, L. (2009). Morality and prosocial behavior: The role of awareness, responsibility, and norms in the norm activation model. *The Journal of Social Psychology*, 149(4), 425-449. <https://doi.org/10.3200/SOCP.149.4.425-449>

de Oliveira, A. C. M., Croson, R. T. A., & Eckel, C. (2015). One bad apple? Heterogeneity and information in public good provision. *Experimental Economics*, 18(1), 116-135. <https://doi.org/10.1007/s10683-014-9412-1>

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, 19(2), 109–134. [https://doi.org/10.1016/0092-6566\(85\)90023-6](https://doi.org/10.1016/0092-6566(85)90023-6)

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, 51(3), 629-636. <https://doi.org/10.1037/h0046408>

Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulation & Games*, 11(1), 87-106. <https://doi.org/10.1177/104687818001100108>

Dowling, D. (1999). Experimenting on theories. *Science in Context*, 12(2), 261–273. <https://doi.org/10.1017/S0269889700003410>

Durkheim, E. (1893). *Über soziale Arbeitsteilung. Studie über die Organisation höherer Gesellschaften* [The Division of Labour in Society]. Suhrkamp.

Eberlen, J., Scholz, G., & Gagliolo, M. (2017). Simulate this! An introduction to agent-based models and their power to improve your research practice. *International Review of Social Psychology*, 30(1), 149-160. <https://doi.org/10.5334/irsp.115>

Edmonds, B. (2014). Agent-based social simulation and its necessity for understanding socially embedded phenomena. In R. Conte, G. Andrighetto, & M. Campenni (Eds.), *Minding norms: Mechanisms and dynamics of social order in agent societies* (pp. 34–49). Oxford University Press.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives*, 3(4), 99–117.

Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, 4(5), 41–60. [https://doi.org/10.1002/\(SICI\)1099-0526\(199905/06\)4:5%3C41::AID-CPLX9%3E3.0.CO;2-F](https://doi.org/10.1002/(SICI)1099-0526(199905/06)4:5%3C41::AID-CPLX9%3E3.0.CO;2-F)

Epstein, J. M. (2001). Learning to be thoughtless: Social norms and individual computation. *Computational Economics*, 18, 9–24. <https://doi.org/10.1023/A:1013810410243>

Epstein, J. M. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press. <https://doi.org/10.1515/9781400842872>

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, 11(4), 12.

Ernst, A. (1997). *Ökologisch-soziale Dilemmata: Psychologische Wirkmechanismen des Umweltverhaltens* [Ecological-social dilemmas: Psychological mechanisms of action of environmental behavior]. Psychologie Verlags Union.

Ernst, A. (2010). Social simulation: A method to investigate environmental change from a social science perspective. In M. Gross & H. Heinrichs (Eds.), *Environmental Sociology* (pp. 109–122). Springer. https://doi.org/10.1007/978-90-481-8730-0_7

Fehr, E., & Fischbacher, U. (2002). Why social preferences matter – The impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal*, 112(478), 1–33. <https://doi.org/10.1111/1468-0297.00027>

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1), 541–556. <https://doi.org/10.1257/aer.100.1.541>

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397–404. [https://doi.org/10.1016/S0165-1765\(01\)00394-9](https://doi.org/10.1016/S0165-1765(01)00394-9)

Fishbein, M., & Ajzen, I. (1981). Attitudes and voting behavior: An application of the theory of reasoned action. *Progress in Applied Social Psychology*, 1(1), 253–313.

Flache, A., & Macy, M. W. (2005). Social life from the bottom up: Agent modeling and the New Sociology. In C. M. Macal, D. Sallach, & M. J. North (Eds.), *Proceedings of the Agent 2004 Conference. Social Dynamics: Interaction, Reflexivity and Emergence* (pp. 275–303). The University Argonne National Laboratory of Chicago.

Gächter, S., & Thöni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, 3(2–3), 303–314.

Gardner, G. T., & Stern, P. (2002). *Environmental problems and human behavior* (2nd ed.). Allyn and Bacon.

Gelfand, M. J. (2018). *Rule makers, rule breakers: How culture wires our minds, shapes our nations and drive our differences*. Robinson.

Gilbert, N. (2008). *Agent-based models*. Sage Publications.
<https://doi.org/10.4135/9781412983259>

Gilbert, N., & Troitzsch, K. G. (1999). *Simulation for the social scientist*. Open University Press.

Göckeritz, S., Schultz, P. W., Rendón, T., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2010). Descriptive normative beliefs and conservation behavior: The moderating roles of personal involvement and injunctive normative beliefs. *European Journal of Social Psychology, 40*(3), 514-523. <https://doi.org/10.1002/ejsp.643>

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35*(3), 472-482. <https://doi.org/10.1086/586910>

Hamann, K. R., Reese, G., Seewald, D., & Loeschinger, D. C. (2015). Affixing the theory of normative conduct (to your mailbox): Injunctive and descriptive norms as predictors of anti-ads sticker use. *Journal of Environmental Psychology, 44*, 1-9.
<https://doi.org/10.1016/j.jenvp.2015.08.003>

Han, H. (2014). The norm activation model and theory-broadening: Individuals' decision-making on environmentally-responsible convention attendance. *Journal of Environmental Psychology, 40*, 462-471. <https://doi.org/10.1016/j.jenvp.2014.10.006>

Hardin, G. (1968). The tragedy of the commons. *Science, 162*(3859), 1243-1248.
<https://doi.org/10.1126/science.162.3859.1243>

Harland, P., Staats, H., & Wilke, H. A. (1999). Explaining proenvironmental intention and behavior by personal norms and the Theory of Planned Behavior. *Journal of Applied Social Psychology, 29*(12), 2505-2528. <https://doi.org/10.1111/j.1559-1816.1999.tb00123.x>

Hines, J. M., Hungerford, H. R., & Tomera, A. N. (1987). Analysis and synthesis of research on responsible environmental behavior: A meta-analysis. *The Journal of Environmental Education, 18*(2), 1-8. <https://doi.org/10.1080/00958964.1987.9943482>

Hoffman, M. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.

Holland, J. H. (2000). *Emergence: From chaos to order*. Oxford University Press.

Hollander, C. D., & Wu, A. S. (2011). The current state of normative agent-based systems. *Journal of Artificial Societies and Social Simulation, 14*(2), 6.
<https://doi.org/10.18564/jasss.1750>

Howard, J. A., & Renfrow, D. G. (2003). Social cognition. In J. Delamater (Ed.), *Handbook of Social Psychology* (pp. 259-281). Kluwer Academic/Plenum Publishers.

Hunecke, M., Blöbaum, A., Matthies, E., & Höger, R. (2001). Responsibility and environment: Ecological norm orientation and external factors in the domain of travel mode choice behavior. *Environment and Behavior*, 33(6), 830-852.
<https://doi.org/10.1177/00139160121973269>

Intergovernmental Panel on Climate Change. (2023). Summary for Policymakers. In Core Writing Team, H. Lee, & J. Romero (Eds.), *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1-34). IPCC.
<https://doi.org/10.59327/IPCC/AR6-9789291691647.001>

Jackson, J. C., Rand, D., Lewis, K., Norton, M. I., & Gray, K. (2017). Agent-based modeling: A guide for social psychologists. *Social Psychological and Personality Science*, 8(4), 387-395. <https://doi.org/10.1177/1948550617691100>

Jacobson, R. P., Mortensen, C. R., & Cialdini, R. B. (2011). Bodies obliged and unbound: Differentiated response tendencies for injunctive and descriptive social norms. *Journal of Personality and Social Psychology*, 100(3), 433. <https://doi.org/10.1037/a0021470>

Jager, W. (2017). Enhancing the realism of simulation (EROS): On implementing and developing psychological theory in social simulation. *Journal of Artificial Societies and Social Simulation*, 20(3), 14. <https://doi.org/10.18564/jasss.3522>

Jager, W., & Ernst, A. (2017). Introduction of the special issue "Social simulation in environmental psychology". *Journal of Environmental Psychology*, 52, 114–118.
<https://doi.org/10.1016/j.jenvp.2017.07.002>

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, 26(8), 1002-1012. <https://doi.org/10.1177/01461672002610009>

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science*, 322(5908), 1681-1685. <https://doi.org/10.1126/science.1161405>

Kelly, D., & Davis, T. (2018). Social norms and human normative psychology. *Social Philosophy and Policy*, 35(1), 54-76. <https://doi.org/10.1017/S0265052518000122>

Klein, S. B. (2014). What can recent replication failures tell us about the theoretical commitments of psychology?. *Theory & Psychology*, 24(3), 326-338.
<https://doi.org/10.1177/0959354314529616>

Kohlberg, L. (1964). Development of moral character and moral ideology. In M. Hoffman & L. W. Hoffman (Eds.), *Review of Research in Child Development* (Vol. 1, pp. 383-431). Russell Sage Foundation.

Kohlberg, L. (1978). Revisions in the theory and practice of moral development. *New Directions for Child and Adolescent Development*, 1978(2), 83–87.
<https://doi.org/10.1002/cd.23219780207>

Kohlberg, L. (1984). *Essays on Moral Development: The Psychology of Moral Development*. Row Publishers, Inc.

- Liebrand, W. B. G., Messick, D. M., & Wilke, H. A. M. (1992). *Social dilemmas: Theoretical issues and research findings*. Pergamon Press.
- Lindenberg, S., & Steg, L. (2007). Normative, gain and hedonic goal frames guiding environmental behavior. *Journal of Social Issues*, 63(1), 117-137.
<https://doi.org/10.1111/j.1540-4560.2007.00499.x>
- Lorenz, J., Neumann, M., & Schröder, T. (2021). Individual attitude change and societal dynamics: Computational experiments with psychological theories. *Psychological Review*, 128(4), 623-642. <https://doi.org/10.1037/rev0000291>
- Luce, R. D., & Raiffa, H. (1957). *Games and Decisions: Introduction and Critical Survey*. John Wiley & Sons.
- Mahmoud, M. A., Ahmad, M. S., Mohd Yusoff, M. Z., & Mustapha, A. (2014). A review of norms and normative multiagent systems. *The Scientific World Journal*, 2014, 684587.
- Miller, N., & Dollard, J. (1941). *Social Learning and Imitation*. Yale University Press.
- Moses, Y., & Tennenholtz, M. (1996). Off-line reasoning for on-line efficiency: Knowledge bases. *Artificial Intelligence*, 83(2), 229-239.
[https://doi.org/10.1016/0004-3702\(95\)00015-1](https://doi.org/10.1016/0004-3702(95)00015-1)
- Moss, S., & Edmonds, B. (2005). Sociology and simulation: Statistical and qualitative cross-validation. *American Journal of Sociology*, 110(4), 1095-1131.
<https://doi.org/10.1086/427320>
- Neumann, M. (2008). Homo Socionicus: A case study of simulation models of norms. *Journal of Artificial Societies and Social Simulation*, 11(4), 6.
- Neumann, M. (2010a). A classification of normative architectures. In K. Takadama, C. Cioffi-Revilla, & G. Deffuant (Eds.), *Agent-Based Social Systems: Vol. 7. Simulating Interacting Agents and Social Phenomena* (pp. 3–18). Springer.
https://doi.org/10.1007/978-4-431-99781-8_1
- Neumann, M. (2010b). Norm internalisation in human and artificial intelligence. *Journal of Artificial Societies and Social Simulation*, 13(1), 12.
- Neumann, M. (2014). How are norms brought about? A state of the art of current research. In R. Conte, G. Andrighetto, & M. Campenni (Eds.), *Minding norms: Mechanisms and dynamics of social order in agent societies* (pp. 50–67). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199812677.003.0004>
- Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin*, 34(7), 913-923. <https://doi.org/10.1177/0146167208316691>
- Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97(3), 362–376.
<https://doi.org/10.1037/0033-295X.97.3.362>

- Nyborg, K. (2018). Social norms and the environment. *Annual Review of Resource Economics*, 10, 405-423. <https://doi.org/10.1146/annurev-resource-100517-023232>
- Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W.N., Arrow, K. J., Barrett, S., Carpenter, S., Chapin III, F. S., Crépin, A.-S., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S. A., Madsen, O. J., ... de Zeeuw, A. (2016). Social norms as solutions. *Science*, 354(6308), 42-43. <https://doi.org/10.1126/science.aaf8317>
- Onwezen, M. C., Antonides, G., & Bartels, J. (2013). The Norm Activation Model: An exploration of the functions of anticipated pride and guilt in pro-environmental behaviour. *Journal of Economic Psychology*, 39, 141-153. <https://doi.org/10.1016/j.joep.2013.07.005>
- Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511807763>
- Ostrom, E. (1999). Coping with tragedies of the commons. *Annual Review of Political Science*, 2(1), 493-535. <https://doi.org/10.1146/annurev.polisci.2.1.493>
- Ostrom, E., Dietz, T., Dolšák, N., Stern, P., Stonich, S., & Weber, E. (2002). *The Drama of the Commons*. National Academy Press.
- Parsons, T. (1937). *The Structure of Social Action*. Free Press.
- Piaget, J. (1965). *The Moral Judgment of the Child* (M. Gabain, Trans.). Free Press. (Original work published 1932)
- Piaget, J. (1970). Piaget's theory. In P. Mussen (Ed.), *Carmichaels' Manual of Child Psychology* (Vol. 1, pp. 703-732). John Wiley & Sons.
- Railsback, S. F., & Grimm, V. (2019). *Agent-based and individual-based modeling: A practical introduction*. Princeton University Press.
- Resnick, M. (1994). *Turtles, termites, and traffic jams: Explorations in massively parallel microworlds*. MIT Press.
- Resnick, M. (1996). Beyond the centralized mindset. *Journal of the Learning Sciences*, 5(1), 1-22. https://doi.org/10.1207/s15327809jls0501_1
- Resnick, M., & Wilensky, U. (1998). Diving into complexity: Developing probabilistic decentralized thinking through role-playing activities. *Journal of the Learning Sciences*, 7(2), 153-172. https://doi.org/10.1207/s15327809jls0702_1
- Rogers, E. M. (2003). *Diffusion of innovations*. Free Press.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54-67. <https://doi.org/10.1006/ceps.1999.1020>
- Ryan, R. M., & Deci, E. L. (2017). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. Guilford Press.

- Saam, N., & Harrer, A. (1999). Simulating norms, social inequality, and functional change in artificial societies. *Journal of Artificial Societies and Social Simulation*, 2(1), 2.
- Savarimuthu, B. T. R., & Cranefield, S. (2011). Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1), 21-54. <https://doi.org/10.3233/MGS-2011-0167>
- Schahn, J., & Bertsch, H. J. (2003). Normdiskrepantes Verhalten im Umweltbereich: Empirischer Test einer Integration des Normaktivationsmodells von Schwartz und der Neutralisationstheorie von Sykes und Matza [Neutralizing norm-discrepant ecological behavior: Testing an integration of Schwartz's norm activation model and Sykes & Matza's techniques of neutralization]. *Umweltpsychologie*, 7(1), 128-148.
- Schlüter, M., Baeza, A., Dressler, G., Frank, K., Groeneveld, J., Jager, W., Janssen, M., McAllister, R., Müller, B., Orach, K., Schwarz, N., & Wijermans, N. (2017). A framework for mapping and comparing behavioural theories in models of social-ecological systems. *Ecological Economics*, 131, 21–35. <https://doi.org/10.1016/j.ecolecon.2016.08.008>
- Schultz, W. P., Khazian, A. M., & Zaleski, A. C. (2008). Using normative social influence to promote conservation among hotel guests. *Social Influence*, 3(1), 4-23. <https://doi.org/10.1080/15534510701755614>
- Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5), 429-434. <https://doi.org/10.1111/j.1467-9280.2007.01917.x>
- Schwartz, S. H. (1977). Normative influences on altruism. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 221-279). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60358-5](https://doi.org/10.1016/S0065-2601(08)60358-5)
- Schwartz, S. H., & Howard, J. (1981). A normative decision-making model of altruism. In J. Rushton (Ed.), *Altruism and Helping Behaviour: Social, Personality and Developmental Perspectives* (pp. 189–211). Lawrence Erlbaum Associates Inc.
- Schwartz, S. H., & Howard, J. (1982). Helping and cooperation: A self-based motivational model. In V. Derlega & J. Grzelak (Eds.), *Cooperation and Helping Behavior: Theories and Research* (pp. 327–353). Academic Press. <https://doi.org/10.1016/B978-0-12-210820-4.50019-8>
- Sen, S., & Airiau, S. (2007). Emergence of norms through social learning. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence* (Vol. 1507, pp. 1507-1512). AAAI Press.
- Sherif, M., & Sherif, C. (1953). *Groups in Harmony and Tension: An integration of studies of intergroup relations*. Harper & Brothers.
- Shin, Y. H., Im, J., Jung, S. E., & Severt, K. (2018). The theory of planned behavior and the norm activation model approach to consumer behavior regarding organic menus. *International Journal of Hospitality Management*, 69, 21-29. <https://doi.org/10.1016/j.ijhm.2017.10.011>

- Shoham, Y., & Tennenholtz, M. (1992). On the synthesis of useful social laws for artificial agent societies. In *AAAI'92: Proceedings of the tenth national conference on artificial intelligence* (pp. 276-281). AAAI Press.
- Shoham, Y., & Tennenholtz, M. (1995). On social laws for artificial agent societies: Off-line design. *Artificial Intelligence*, *73*(1-2), 231-252.
[https://doi.org/10.1016/0004-3702\(94\)00007-N](https://doi.org/10.1016/0004-3702(94)00007-N)
- Smaldino, P. E. (2017). Models are stupid, and we need more of them. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational social psychology* (pp. 311-331). Routledge.
- Smaldino, P. E. (2020). How to translate a verbal theory into a formal model. *Social Psychology*, *51*(4), 207-218. <https://doi.org/10.1027/1864-9335/a000425>
- Smaldino, P. E., Calanchini, J., & Pickett, C. L. (2015). Theory development with agent-based models. *Organizational Psychology Review*, *5*(4), 300-317.
<https://doi.org/10.1177/2041386614546944>
- Smith, E. R., & Conrey, F. R. (2007). Agent-based modeling: A new approach for theory building in social psychology. *Personality and Social Psychology Review*, *11*(1), 87-104. <https://doi.org/10.1177/1088868306294789>
- Smith, J. R., Louis, W. R., Terry, D. J., Greenaway, K. H., Clarke, M. R., & Cheng, X. (2012). Congruent or conflicted? The impact of injunctive and descriptive norms on environmental intentions. *Journal of Environmental Psychology*, *32*(4), 353-361.
<https://doi.org/10.1016/j.jenvp.2012.06.001>
- Steg, L., & de Groot, J. (2010). Explaining prosocial intentions: Testing causal relationships in the norm activation model. *British Journal of Social Psychology*, *49*(4), 725-743. <https://doi.org/10.1348/014466609X477745>
- Stern, P. C. (2000). New environmental theories: Toward a coherent theory of environmentally significant behavior. *Journal of Social Issues*, *56*(3), 407-424.
<https://doi.org/10.1111/0022-4537.00175>
- Szekely, A., Lipari, F., Antonioni, A., Paolucci, M., Sánchez, A., Tummolini, L., & Andrighetto, G. (2021). Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications*, *12*(1), 1-7.
<https://doi.org/10.1038/s41467-021-25734-w>
- Terrier, L., & Marfaing, B. (2015). Using social norms and commitment to promote pro-environmental behavior among hotel guests. *Journal of Environmental Psychology*, *44*, 10-15. <https://doi.org/10.1016/j.jenvp.2015.09.001>
- Theriault, J. E., Young, L., & Barrett, L. F. (2021). The sense of should: A biologically-based framework for modeling social pressure. *Physics of Life Reviews*, *36*, 100-136. <https://doi.org/10.1016/j.plrev.2020.01.004>
- Thøgersen, J. (1999). The ethical consumer: Moral norms and packaging choice. *Journal of Consumer Policy*, *22*(4), 439-460. <https://doi.org/10.1023/A:1006225711603>

- Thøgersen, J. (2003). Monetary incentives and recycling: Behavioural and psychological reactions to a performance-dependent garbage fee. *Journal of Consumer Policy*, 26, 197–228. <https://doi.org/10.1023/A:1023633320485>
- Thøgersen, J. (2006). Norms for environmentally responsible behaviour: An extended taxonomy. *Journal of Environmental Psychology*, 26(4), 247–261. <https://doi.org/10.1016/j.jenvp.2006.09.004>
- Thøgersen, J. (2008). Social norms and cooperation in real-life social dilemmas. *Journal of Economic Psychology*, 29(4), 458–472. <https://doi.org/10.1016/j.joep.2007.12.004>
- Thöni, C., & Gächter, S. (2015). Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis. *Journal of Economic Psychology*, 48, 72–88. <https://doi.org/10.1016/j.joep.2015.03.001>
- Troitzsch, K. (2017). Axiomatic theory and simulation. A philosophy of science perspective on Schelling's segregation model. *Journal of Artificial Societies and Social Simulation*, 20(1), 10. <https://doi.org/10.18564/jasss.3372>
- Tverskoi, D., Guido, A., Andrighetto, G., Sánchez, A., & Gavrilets, S. (2023). Disentangling material, social, and cognitive determinants of human behavior and beliefs. *Humanities and Social Sciences Communications*, 10(1), 1–13. <https://doi.org/10.1057/s41599-023-01745-4>
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323. <https://doi.org/10.1007/BF00122574>
- Ullmann-Margalit, E. (1977). *The Emergence of Norms*. Clarendon Press.
- Verhagen, H. (2001). Simulation of the learning of norms. *Social Science Computer Review*, 19(3), 296–306. <https://doi.org/10.1177/089443930101900305>
- Villatoro, D., Andrighetto, G., Conte, R., & Sabater-Mir, J. (2015). Self-policing through norm internalization: A cognitive solution to the tragedy of the digital commons in social networks. *Journal of Artificial Societies and Social Simulation*, 18(2), 2. <https://doi.org/10.18564/jasss.2759>
- von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 147–188). Sharpe, Inc. (Original work published 1930)
- Vygotsky, L. S. (2004). Analysis of sign operations of the child. In R. Rieber & D. Robinson (Eds.), *The Essential Vygotsky* (pp. 557–569). Kluwer Academic/Plenum Press.

Waldherr, A., & Wettstein, M. (2019). Bridging the gaps: Using agent-based modeling to reconcile data and theory in computational communication science. *International Journal of Communication*, 13, 3976-3999. <https://doi.org/10.5167/uzh-186794>

Wallis, K., & Poulton, L. (2001). *Internalization: The Origins and Construction of Internal Reality*. Open University Press.

Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. The MIT Press.

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, 8(1), 3-19. <https://doi.org/10.1023/A:1009421303064>

Windrum, P., Fagiolo, G., & Moneta, A. (2007). Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, 10(2), 8.

APPENDIX

Appendix A

Paper I: The Effects of Norms on Environmental Behavior

The following paper was accepted for publication by the *Review of Environmental Economics and Policy* on 18.10.22. The following shows the revised manuscript version from 16.08.2023.

Dannenberg, A., Gutsche, G., Batzke, M. C. L., Christens, S., Engler, D., Mankat, F., Möller, S., Weingärtner, E., Ernst, A., Lumkowsky, M., von Wangenheim, G., Hornung, G., & Ziegler, A. (in press). The effects of norms on environmental behavior. *Review of Environmental Economics and Policy*.

The Effects of Norms on Environmental Behavior

Astrid Dannenberg^a, Gunnar Gutsche^a, Marlene Batzke^b, Sven Christens^a, Daniel Engler^a,
Fabian Mankat^c, Sophia Möller^a, Eva Weingärtner^a, Andreas Ernst^b, Marcel Lumkowsky^a,
Georg von Wangenheim^c, Gerrit Hornung^c, Andreas Ziegler^a

^a University of Kassel, Institute of Economics, 34109 Kassel, Germany

^b University of Kassel, Center for Environmental Systems Research, 34117 Kassel, Germany

^c University of Kassel, Institute of Economic Law, 34127 Kassel, Germany

Abstract

The study of norms is of paramount importance in understanding human behavior. An interdisciplinary literature, using varying definitions and conceptions, shows when and why norms emerge and spread, what form they can take, and how they are enforced. Here, we focus on theoretical and empirical literature that treats norms as a factor influencing human behavior. We first present a new taxonomy of norms, which builds upon and merges previous taxonomies, to distinguish between different types of norms and enforcement mechanisms. We then provide a conceptual framework that identifies reasons for the effects of norms. This framework is based on psychological theories, which serve as a foundation for much of the empirical economic literature measuring norm effects. Finally, we present an overview of empirical economic papers that study the effects of norms on environmentally relevant behavior, as a particularly relevant area for the study of norms. The aim of this overview is to highlight which effects have been insufficiently studied and to give a sense of the potential of norms. This can help policymakers intervene in a more targeted way to address environmental problems.

JEL classification codes: C9, D7, D8, D9, H4, Q0

Competing interests statement: The authors declare no competing interests.

1. Introduction

The interest in understanding social or behavioral norms has increased greatly in recent years. This is particularly true for the analysis of environmental behavior, which appears to be strongly affected by norms, from decisions about food and transportation, to energy and water use, to waste separation and recycling.

While the concept of norms varies across disciplines (Cialdini and Goldstein 2004; Young 2015), they often refer to a shared understanding in society about appropriate behaviors and wide participation in implementing and enforcing these behaviors. Norms exist in all human societies in various forms, contexts, and dimensions. Knowing how norms work and can be influenced provides policymakers with a powerful tool (Nyborg et al. 2016). In the environmental field, political processes are often accompanied or even enabled by changes in norms (Nyborg 2018). Climate change is a prominent example, where progress on international climate policy is limited, while much activity takes place at the local and individual level.

In this paper, we compile theoretical and empirical literature that interprets norms as a factor influencing human behavior, with a primary focus on environmental and prosocial behavior. To address the problem of different definitions in this literature, we begin with a new taxonomy that captures the key dimensions of norms and their conceivable combinations. We then elaborate a conceptual structure that presents reasons for the effects of various types of norms on human behavior based on psychological theories. These theories have found their way into a variety of empirical studies, including the empirical economics literature on the measurement of the effects of norms, which we review in the last part of the paper. This last part focuses on environmental and prosocial behavior because this appears to be an area where people are increasingly paying attention and potentially influencing each other, and where there is a pressing need for policymakers to better understand how norms guide behavior. Our focus on norms as a factor influencing behavior leaves out the theoretical literature in economics, game theory, and evolutionary biology; these typically treat norms as a description of equilibrium or steady-state behavior, which is thoroughly summarized in other overviews (e.g., Farrow, Grolleau, and Ibanez 2017; Nyborg 2018; Ehrlich and Levin 2005; Nowak and Sigmund 2005; Okada 2020).¹

¹ Farrow, Grolleau, and Ibanez (2017) provide an overview of how norms have entered economics theoretically, picking up on topics such as self-image (e.g., Elster 1989), identity economics (e.g., Akerlof and Kranton 2000), normative expectations (e.g., Sugden 2000), or prosocial behavior (e.g., Bénabou and Tirole 2006). Nyborg (2018) explains how social norms are understood in game theory (which has actors play structured games as models of real-world situations) and evolutionary game theory (applying a similar approach to population dynamics; see, e.g., Young 1998, 2015, Nyborg and Rege 2003, Rege 2004). Nowak and Sigmund (2005) and Okada (2020)

2. Taxonomy of norms

Several taxonomies have been developed and employed in the social sciences to distinguish between the various forms, functions, and dimensions of norms (Farrow, Grolleau, and Ibanez 2017). Table 1 presents the taxonomy we use in this paper, which we have developed by merging, adjusting, and refining previous conceptualizations (Schwartz 1977; Schwartz and Howard 1982; Cialdini, Reno, and Kallgren 1990; Stern et al. 1999; Farrow, Grolleau, and Ibanez 2017; Nyborg 2018; Bicchieri and Dimant 2019).²

We define a norm as the rule that characterizes a subset of all possible behaviors as either appropriate or customary. By ‘quality’ of the norm, we refer in our taxonomy to the important distinction between ‘injunctive norms,’ which describe what behavior is appropriate in a given situation, and ‘descriptive norms,’ which describe what people actually do.

If the goal of studying norms is to understand their effects on behavior, we also need to consider the ‘subject,’ who acts on the norm or believes in it. Is it an individual who shows a certain behavior or finds it appropriate? Is it a community that evaluates or exhibits behavior? Or is it a legislative authority that prescribes or prohibits behavior? Norms held by these different types of actors may closely interact and emerge from one another, yet keeping a conceptually clear distinction facilitates the study of norms and their effects on behavior.

The next dimension in our taxonomy, ‘perspective,’ involves the difference between ‘perceived norms,’ which refer to an individual’s subjective beliefs, and ‘objective norms,’ which refer to actual behaviors or attitudes. Differences between perceived and objective norms arise because people sometimes send misleading signals, misinterpret signals from others, have biased beliefs, or simply lack information about others in the population. Although any information that individuals use to make decisions is ultimately subjective, it is useful to distinguish conceptually between perceptions and objective facts.

summarize theoretical work in evolutionary biology on norms as assessment rules in indirect reciprocity (e.g., Ohtsuki and Iwasa 2006), while Ehrlich and Levin (2005) cover norms as conventions from a cultural evolution perspective.

² For reasons of space, we can only describe the components of the taxonomy here. The appendix to this paper explains the relationships between the components and presents an illustrative example.

Table 1
Taxonomy of norms

Dimension	Type of norm			
	<i>Descriptive norm</i>		<i>Injunctive norm</i>	
Quality of the norm	Perceivable situation-specific behavior aggregated over time and/or individuals	Situation-specific behavior that is seen as (in)appropriate		
	<i>Personal descriptive norm</i>	<i>Social descriptive norm</i>	<i>Personal injunctive norm</i>	<i>Legal injunctive norm</i>
Quality ×	An individual regularly follows a behavioral pattern	A significant proportion of individuals regularly follows a behavioral pattern	An individual considers a behavior as (in)appropriate for him-/herself (self-oriented) and/or others (other-oriented)	A significant proportion of individuals considers a behavior as (in)appropriate for themselves (self-oriented) and/or others (other-oriented)
Subject of the norm				
[Quality × Subject ×]	<i>Objective</i>	<i>Objective</i>	<i>Perceived</i>	<i>Perceived</i>
Perspective on the norm	Actually prevalent norm	Actually prevalent norm	Subjective perception of a norm	Subjective perception of a norm

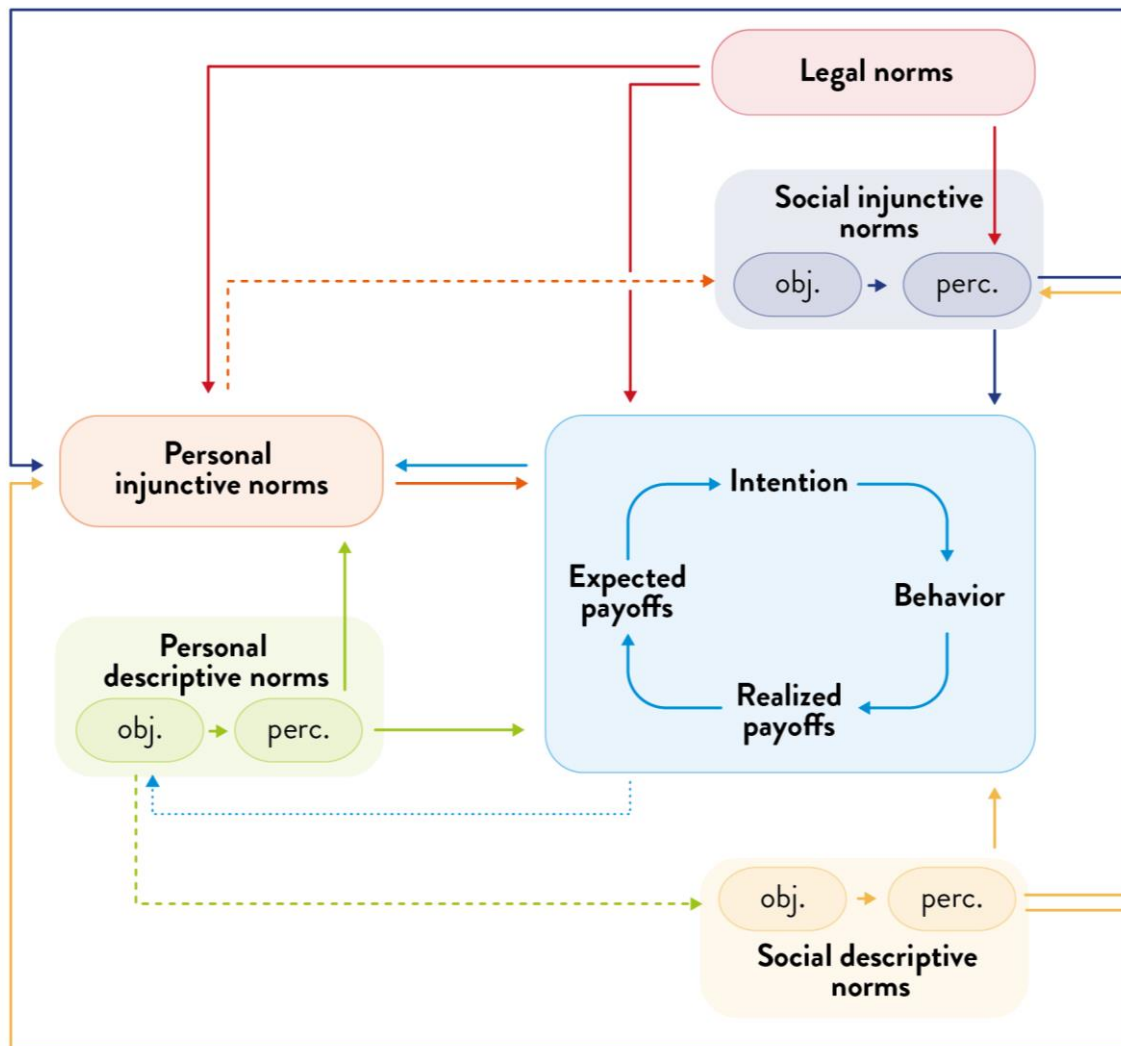
Finally, compliance with existing norms crucially depends on the ‘enforcement’ of norms. We call purely internal motives such as satisfaction, inner peace, or avoidance of guilt ‘personal enforcement.’ Seeking social approval or avoiding social disapproval or sanctions is called ‘social enforcement.’ ‘Legal enforcement’ is carried out by state coercion, where individuals are subject to measures set by law if they deviate from a legal norm. These different enforcement mechanisms occur separately or simultaneously, reinforcing or displacing each other; again, conceptual separation is important for understanding the different effects.

While all four dimensions in our taxonomy can in principle be combined with each other, some dimensions are closely correlated or overlapping. For example, legal enforcement applies only to legal norms. The separation between the subject of the norm and its enforcement thus contains some redundancies, but it aids clear thinking about what characterizes and differentiates norms.

3. A conceptual structure of the causes of norm effects

Figure 1 provides a graphical illustration of the various interactions between different types of norms and their influence on decision making. The aim of this figure is to show the most important channels of influence and to explain our current state of knowledge on the basis of theories from social, cognitive, developmental, and motivational psychology.³

³ In the appendix to this paper, we provide additional explanations of the conceptual structure in Figure 1 and examples of how this general framework can also be used to classify models from behavioral economics and evolutionary game theory.

Figure 1*The effects of norms*

Note. The figure shows a conceptual structure of norm influences. The arrows represent the influence of the source variable on the change of the targeted variable. The blue rectangle in the center right illustrates the basic decision-making process. All other rectangles refer to a specific type of norm. The personal descriptive, social descriptive, and social injunctive norms are differentiated between their objective and perceived components. This distinction does not apply to personal injunctive norms, which we assume to be inherently subjective. The figure captures how legal norms may exert influence on other norms but not how those other norms in turn affect legal norms, which is beyond the scope of this paper. A norm may influence decision-making, be influenced by it, or influence other norms. Each assumption concerning either of the three is depicted as a solid arrow. The objective components of norms are aggregations over time (dotted arrow) or individuals and time (dashed arrows).

An individual's decision-making process is at the center of the illustration. This process is a result of a person's past experiences, knowledge, and situation, expressed in terms of expectations about payoffs from the available behavioral options. Payoffs in this context include not just material benefits and costs but also non-material benefits and costs such as positive or negative feelings. The intentions that result from the situation-specific weighing of expected payoffs are finally translated into actions (Fishbein and Ajzen 1981). It is useful to distinguish between an intention – an unobservable outcome of an internal cognitive process – and external behavior, which is visible to the social world (Ajzen 1991). Because intentions may not translate into actions, there may be an intention-behavior gap (Sheeran 2002). The action taken, together with the actions of others, determines the individual's payoff. The payoff provides feedback about the success of one's own choice given the circumstances; this gives rise to a learning process, in which a person changes her expectations, depending on how successful she was in leading to desired payoffs in a given situation (Bandura 1999).

Let's first look at the connections between an individual's current behavior, shown in the center of the figure, and the individual's personal norms, shown on the left. The current individual behavior is added to the objective personal descriptive norm. The objective norm provides the basis for how the person perceives herself, which in turn influences her future behavior, as people tend to strive for consistency (Elliott 1986; Aronson 1992). The perception of one's own behavior also influences the personal injunctive norm. According to the self-perception theory (Bem 1967; Bem 1972) and other consistency theories (Heider 1946; Festinger 1957; Harmon-Jones et al. 2009), past behavior is evaluated in an internal after-the-fact reasoning process and may be attributed to normative beliefs. This post-hoc process is more likely to happen if a person's behavior is associated with high payoffs; the more successful one's past behavior has been, the more likely it is that she will decide it was appropriate. The development of the personal injunctive norm is a complex process also involving other norms, which will be described further below. Once formed, the personal injunctive norm exerts a decisive influence on an individual's behavior, because deviations from the personal injunctive norm create feelings of inner conflict, failure, guilt, or shame (Schwartz 1977; Thøgersen 2007).

In the figure, social norms are shown above and below the individual decision-making process. Individual behaviors, aggregated over time and people, constitute the objective social descriptive norm. Likewise, the aggregation of all individual attitudes over time and people forms the objective social injunctive norm (Cooter 1998; Carbonara, Parisi, and von Wangenheim 2008). Objective social norms are the bases for how individuals perceive social norms, but perceived social norms may deviate due to observation errors or motivated

information seeking and reasoning – for example, an individual may seek and interpret social information in a way that fits her prior beliefs or self-image (e.g., Kunda 1990). The perceived social descriptive norm affects individuals' decision-making through conformism and imitation (Smith 2012; White and Simpson 2013; White, Habib, and Hardisty 2019). Observation of others provides clues as to what behavior is effective and adaptive, and thus encourages imitation (Cialdini, Reno, and Kallgren 1990; Hamann et al. 2015; Dimant 2019). In addition to the tendency to follow others, observing others also can lead individuals to adjust their personal view of what is appropriate behavior (Bandura 2001).

Social injunctive norms influence people's behavior by affecting their expectations of social responses to their behavior, such as sanctions, disapproval, or recognition (Elster 1989; Ajzen 1991; Sunstein 1996; Ellickson 2001; Schultz et al. 2007; Jacobson, Mortensen, and Cialdini 2011). They also play an important role in the development and change of personal injunctive norms. This process of norm internalization is a learning process in which the perceived social injunctive norms are integrated through moral cognition and reasoning (Kohlberg 1964; Piaget 1965[1932]; Kohlberg 1978; Hoffman 2000).

Because of the close link between personal injunctive norms and individual behavior, there is an equally close link between social injunctive norms and collective behavior. Behavior in a society is often used as an indicator of what is considered right and wrong in that society (Bicchieri 2005; Morris et al. 2015; Nyborg 2018). People's tendency to assign causes to behaviors and to see their environment as more controllable than it is (Heider 1958; Kelley 1967) may even lead them to equate the perceived social descriptive norm with the social injunctive norm.

Legal norms are shown at the top of the figure. Their purpose is to increase or decrease the expected and actual payoffs associated with available behaviors and to regulate individuals' behaviors. But they also have effects that go beyond that. According to the theory of expressive law (Cooter 1998; Cooter 2000), people will see a problem as more widespread and pressing if a law regulates it. A newly introduced law thus signals a serious problem and a consensus in the society to address it, thereby changing people's perceptions of social injunctive norms, independent of the altered payoffs (Tyler 1990; Tyler and Huo 2002). In addition, many people ascribe normative power to the law and adapt their personal injunctive norms to legal norms, just because these norms are prescribed by law (for a review, see Larcom, Panzone, and Swanson 2019). These shifts at the individual level, which are due to altered payoffs and attitudes, add up to shifts at the societal level and change the objective social descriptive and injunctive norms.

The conceptual structure in Figure 1 also provides guidance on how policy can intervene to influence behavior. We can distinguish between policies that directly target individual behavior (the blue box) and those measures that aim to affect (perceived) personal or social norms and thus indirectly influence individual behavior. Policies that directly target behavior include price regulation (such as gasoline taxes), command-and-control regulation (for instance, requirements to use or not use a particular technology), and adjustments to the set of choices that individuals face (such as electricity pricing systems that charge extra to run appliances at times of high electricity demand). These measures affect individuals' expected payoffs by making environmentally harmful behavior more costly, less beneficial, or less convenient. Policies that influence behavior indirectly through manipulation of norms, which are called "active norm management" by Kinzig et al. (2013), include the provision of information, for example about other people's behavior or attitudes; publicizing neighbors' successful water conservation is an example. Personal injunctive norms might be activated by making consumers aware of the circumstances under which a product was manufactured; for instance, product labels may announce that the packaging is made from recycled material.

Of course, the better the interactions are empirically tested, the more valuable Figure 1 is to policy. This is the subject of the next section.

4. Empirical analyses of social and personal norms

The literature on the empirical analysis of norms is growing rapidly, and it is virtually impossible to provide a comprehensive overview. We limit our review to studies that meet the following three criteria. First, the behaviors and norms analyzed can be ranked in terms of environmental friendliness or prosociality. Second, the behaviors analyzed have consequences for oneself or other people and are not purely hypothetical. Third, the studies examine behaviors and norms by either using randomized treatments (comparing control groups to groups that receive an experimental treatment) or by measuring norms in an incentive-compatible manner (where subjects have an incentive to reveal their true preferences or beliefs). In the case of field experiments, we further limit the selection to studies dealing with environmental behavior and use of natural resources.

Table 2 provides an overview of the selected lab and field experiments. It contains information on how the different types of norms are implemented, what the main results are, and how the studies are distributed across the different types of norms.

Objective social descriptive norms are implemented by giving subjects information about the behavior of others or having them observe the behavior of others. In the field

experimental settings, subjects receive information on other people's energy conservation (Schultz et al. 2007; Allcott 2011), residential water usage (Ferraro, Miranda, and Price 2011; Tiefenbeck et al. 2013), towel reuse (Goldstein, Cialdini, and Griskevicius 2008; Reese, Loew, and Steffgen 2014), recycling or littering behavior (Reiter and Samuel 1980; Cialdini, Reno, and Kallgren 1990; Schultz 1999), use of public transport (Gravert and Olsson Collentine 2021), or food choices (Demarque et al. 2015; Sparkman and Walton 2018; Einhorn 2020; Griesoph et al. 2021). The information can be provided in writing, for example as an enclosure to the energy or water bill, or by manipulating the decision context. As an example of the latter, studies of littering behavior vary the amount of litter already lying around to present visual clues about the behavior of others (Reiter and Samuel 1980; Sagebiel et al. 2020). Participants in lab experiments directly observe or receive information about how other people behave in a variety of games. These include dictator games (Bicchieri and Xiao 2009; Krupka and Weber 2009) and modified dictator games (Schram and Charness 2015; Goeschl et al. 2018), where the player is asked to make a sharing decision; public goods games, in which players have to decide between their individual interest and the public interest (Carpenter 2004; Dal Bó and Dal Bó 2014); ultimatum games, where players have to agree on how to share a pie (Bicchieri and Chavez 2010); and gift exchange games, which are used to measure reciprocal actions (Thöni and Gächter 2015).

In the case of objective social injunctive norms, studies use messages or smileys to inform subjects about what other people consider to be appropriate or important (Bicchieri and Xiao 2009) – for example, whether saving energy is an important value (Bonan et al. 2020). Some of the techniques, such as smileys, not only convey social injunctive norms, but also generate a positive feeling, triggered by the nice picture or the feeling of being rewarded or being better than others (Bhanot 2021). Alternatively, subjects are confronted with third party advice, which refers to moral principles or appropriate behavior (Keizer, Lindenberg, and Steg 2008; Ferraro and Price 2013; Dal Bó and Dal Bó 2014; Schram and Charness 2015; Einhorn 2020), the harmful consequences of certain behaviors (Dannenberg and Weingärtner 2022), or the moral duty to avoid such harms (Panzone et al. 2021a). Using lab experiments, Dal Bó and Dal Bó (2014) show how cooperative behavior and expectations in a public goods game change when subjects are told that, for moral reasons, one should treat others as one would like to be treated or in ways that maximize the benefit to all. Schram and Charness (2015) study behavior in a modified dictator game in which subjects receive advice on what they “ought” to do from a group of uninvolved participants.

Providing objective information about social norms changes the subjective beliefs and, through them, the behavior of individuals. However, most studies only measure the change in behavior and not the changed perceptions. It is also important to note that what subjects learn about objective social norms through information, observation, or advice relates to a fraction of the population, sometimes only individual members. Many studies deliberately don't tell the participants how many and which people are involved in the behavior or attitudes that are presented. This is partly for methodological reasons – for example, when the communication of virtuous and harmful behavior is to be compared in different treatments. A systematic investigation of the reference group therefore is an important task for future research to study which persons are regarded as relevant (Knight Lapinski and Rimal 2005).

Lab experiments have been used to elicit perceived social norms, using a focusing technique where subjects are asked to guess how others behave (in the case of descriptive social norms) or what others consider to be appropriate (in the case of injunctive social norms). Correct guesses are rewarded, to provide an incentive to guess correctly and reveal the true beliefs (Krupka and Weber 2009; Bicchieri and Chavez 2010).

Objective personal descriptive norms are put into effect by providing subjects with information about their own past behavior – for example, their past energy consumption (Allcott 2011; Allcott and Rogers 2014; Andor et al. 2020) or recycling behavior (Schultz 1999). Schultz (1999) finds that reminding subjects of their past recycling behavior leads them to behave in a more environmentally friendly manner, compared to a control group that receives no information. This information about one's own past behavior is often provided in combination with information about what others do or find appropriate (Goldstein, Cialdini, and Griskevicius 2008; Allcott 2011). Although providing information about both one's own and others' behavior helps subjects assess their own behavior (Bonan et al. 2020), it impedes a clean differentiation between norms. In the future, it might be useful to think about other ways of assessing personal behavior that are not so much based on other people as on environmental necessities or hypothetical best-practice scenarios.

There are relatively few studies of personal injunctive norms (Kantola, Syme, and Campbell 1984; Panzone et al. 2021b). Panzone et al. (2021b) have subjects recall environmental protection measures they took in the past week, before they are asked to go grocery shopping at an online supermarket. These researchers find that remembering past environmentally friendly actions leads to more climate-friendly grocery purchases. Kantola, Syme, and Campbell (1984) observe that drawing consumers' attention to a contradiction between their previously measured attitudes toward electricity conservation and their actual

high electricity consumption is more effective in inducing energy savings than merely informing them that they are high electricity consumers (or providing no information).

The studies presented in Table 2 show that subjects tend to adjust their behavior in the direction of the presented norm and behave in a more environmentally friendly or prosocial manner. In certain cases, however, prosocial behavior decreases – for example, when highly cooperative individuals adjust their behavior toward a less cooperative norm. Using social injunctive norms, in addition to social descriptive norms, reduces this ‘boomerang effect’ (Schultz et al. 2007). Some studies find that adaptation to selfish norms is stronger than to cooperative norms (Thöni and Gächter 2015). Adaptation towards social descriptive and injunctive norms increases with the social proximity between decision makers and the reference group (Dimant 2019; Bicchieri et al. 2022). Adaptation also increases when a punishment option is available (Dal Bó and Dal Bó 2014; Bicchieri, Dimant, and Xiao 2021) or when choices are made public (Schram and Charness 2015). Naturally, people differ in their inclinations to adapt toward the norm (Ayres, Raseman, and Shih 2013). For example, Costa and Kahn (2013) find that liberals and environmentalists react more strongly than conservatives to home energy reports.

Our review also shows that the effects of social norms depend on which environmental behavior is studied. Social norms appear to have robust effects on mostly private environmental behaviors that cannot be observed by outsiders, that can be adapted relatively easily, and that are usually associated with saving money, like energy or water consumption, reducing waste, or returning bottles and cans for recycling. However, when it comes to how we eat, dress, or get around, there don’t seem to be robust effects from experimentally providing information about what other people think or do. There are few studies in these areas and they suggest small effects. Providing injunctive messages leads to only small changes in people’s food consumption (Einhorn 2020; Panzone et al. 2021a; Dannenberg and Weingärtner 2022). The findings for social descriptive norms are mixed; some studies find no effect (Einhorn 2020; Griesoph et al. 2021), while others show that specific ways of framing the descriptive norm can foster sustainable choices, even when the norm is described as something that only a minority of people practice (Demarque et al. 2015; Sparkman and Walton 2018). Informing newcomers who have recently moved to a new neighborhood about the use of public transport by locals does not have much effect (Gravert and Olsson Collentine 2021). Existing studies on sustainable clothing are largely based on self-reported consumption decisions, or do not inform participants of others’ norms, which is why they are not listed in Table 2 (Hustvedt and Bernard 2010; Kumar, Manrai, and Manrai 2017; Lin and Niu 2018; Kim and Seock 2019; Lo, Tsarenko,

and Tojib 2019; Park and Lin 2020). The results are again mixed. Hiller, Connell, and Kozar (2012) find that members of a sorority did not report making more sustainable clothing choices after having received injunctive messages about the implications of clothing for sustainability in general, human rights, or environmental protection. Frick et al. (2021) show that a sufficiency-promoting message by an online clothing store leads to more sustainable clothing decisions. However, a high number of likes and comments, signaling social endorsement of the message, does not further increase the effect.

It has yet to be determined whether this difference in the effects of norm interventions really exists between unobservable and observable environmental behavior, and, if so, what factors are responsible. The unobservable environmental behavior that has been studied so far is relatively easy to adapt and mostly produces economic gains, which could facilitate the operation of norms. In contrast, changing the way one gets around, dresses, or eats often comes at a monetary cost, or causes a loss of individual welfare or utility. Another possible explanation is that people are less knowledgeable about unobservable behavior and the norms provided are more likely to contain new information. The norms regarding observable behaviors are more likely to be already factored into the respective decisions. Choices that are visible to the people around us are arguably harder to change because they are more relevant to our social identity. On the other hand, the visibility of behavior could mean that social norms, once they exist, are more easily enforced in these areas. The interplay of social and personal norms and of visibility and identity would certainly be worth exploring in more detail (Gromet et al. 2013).

Table 2

Operationalization of norms in laboratory and field studies

Norm	Norm implementation	Main results	Laboratory studies	Field studies
Objective social descriptive norm	Subjects receive information about the past behavior of others	Subjects adapt their behavior and belief about others towards the norm Previously cooperative people reduce their prosocial behavior (boomerang effect); previously non-cooperative people increase their prosocial behavior	Bicchieri and Xiao 2009; Krupka and Weber 2009; Raihani and McAuliffe 2014; Goeschl et al. 2018; Bicchieri and Dimant 2021	Goldstein, Cialdini, and Griskevicius 2008 ^{a)} ; Schultz, Khazian, and Zaleski 2008 ^{a)} ; Bohner and Schlüter 2014 ^{a)} ; Reese, Loew, and Steffgen 2014 ^{a)} Schultz et al. 2007 ^{b)} ; Nolan et al. 2008 ^{b)} ; Allcott 2011 ^{b)} ; Carrico and Riemer 2011 ^{b)} ; Peschiera and Taylor 2012 ^{b)} ; Ayres, Raseman, and Shih 2013 ^{b)} ; Costa and Kahn 2013 ^{b)} ; Delmas and Lessem 2014 ^{b)} ; Dolan and Metcalfe 2015 ^{b)} ;

Norm	Norm implementation	Main results	Laboratory studies	Field studies
		Social proximity strengthens adaptation Larger gap between own behavior and norm strengthens adaptation		Schultz et al. 2015 ^{b)} ; Shen, Cui, and Fu 2015 ^{b)} ; Alberts et al. 2016 ^{b)} ; Anderson et al. 2017 ^{b)} ; De Dominicis et al. 2019 ^{b)} ; Andor et al. 2020 ^{b)} ; Bonan et al. 2020 ^{b)} Ferraro, Miranda, and Price 2011 ^{c)} ; Ferraro and Price 2013 ^{c)} ; Tiefenbeck et al. 2013 ^{c)} ; Bernedo, Ferraro, and Price 2014 ^{c)} ; Seyranian, Sinatra, and Polikoff 2015 ^{c)} ; Hahn et al. 2016 ^{c)} ; Sparkman and Walton 2017 ^{c)} ; Jaime Torres and Carlsson 2018 ^{c)} ; Bhanot 2021 ^{c)} Demarque et al. 2015 ^{d)} ; Sparkman and Walton 2017 ^{d)} ; Richter, Thøgersen, and Klöckner 2018 ^{d)} ; Einhorn 2020 ^{d)} ; Griesoph et al. 2021 ^{d)} Gravert and Olsson Collentine 2021 ^{e)} Schultz 1999 ^{f)}
	Subjects observe the behavior of others	Subjects adapt their behavior towards the norm Social proximity strengthens adaptation Observing selfish behavior strengthen adaptation	Carpenter 2004; Thöni and Gächter 2015; Gächter, Gerhards, and Nosenzo 2017; Dimant 2019	Oceja and Berenguer 2009 ^{a)} ; Delmas and Lessem 2014 ^{a)} Sussman and Gifford 2013 ^{f)} Cialdini, Reno, and Kallgren 1990 ^{g)} ; Keizer, Lindenberg, and Steg 2008 ^{g)} ; Bator, Bryan, and Schultz 2011 ^{g)} Reese et al. 2013 ^{h)} ; Hamann et al. 2015 ^{h)}
Perceived social descriptive norm	Subjects are asked to guess the behavior of others	Subjects adapt their behavior towards their guess about the descriptive norm	Krupka and Weber 2009	Griesoph et al. 2021 ^{d)}
Objective social injunctive norm	Subjects get third-party advice on appropriate behavior or moral principles	Subjects adapt their behavior toward the advice Accompanying information about others' behavior strengthens the effect of the advice	Dal Bó and Dal Bó 2014; Schram and Charness 2015	Schultz, Khazian, and Zaleski 2008 ^{a)} ; Bohner and Schlüter 2014 ^{a)} Nolan et al. 2008 ^{b)} ; Ito, Ida, and Tanaka 2018 ^{b)} Ferraro, Miranda, and Price 2011 ^{c)} ; Ferraro and Price 2013 ^{c)} ; Tiefenbeck et al. 2013 ^{c)} ; Bernedo, Ferraro, and Price

Norm	Norm implementation	Main results	Laboratory studies	Field studies
		Punishment option strengthens the effect of the advice		2014 ^{c)} ; Seyranian, Sinatra, and Polikoff 2015 ^{c)} Einhorn 2020 ^{d)} ; Panzone et al. 2021a ^{d)} Sussman and Gifford 2013 ^{f)} Keizer, Lindenberg, and Steg 2008 ^{g)} de Groot, Abrahamse, and Jones 2013 ^{h)} ; Kallbekken and Sælen 2013 ^{h)} ; Hamann et al. 2015 ^{h)} ; Jagau and Vyrastekova 2017 ^{h)} ; Stöckli, Dorn and Liechti 2018 ^{h)}
	Subjects receive information about what others consider to be appropriate	Subjects adapt their behavior and belief about others' attitudes toward the norm	Bicchieri and Xiao 2009; Raihani and McAuliffe 2014 d'Adda et al. 2020; Bicchieri, Dimant, and Xiao 2021	Bonan et al. 2020 ^{b)} Linder, Lindahl, and Borgström 2018 ^{f)} de Groot, Abrahamse, and Jones 2013 ^{g)} Stöckli, Dorn and Liechti 2018 ^{h)}
	Subjects receive information about how appropriate their behavior is by comparing it to the behavior of others	Subjects adapt their behavior toward the norm Combination of social injunctive and descriptive norm avoids the boomerang effect	-	Schultz, Khazian, and Zaleski 2008 ^{a)} Schultz et al. 2007 ^{b)} ; Oceja and Berenguer 2009 ^{b)} ; Allcott 2011 ^{b)} ; Ayres, Raseman, and Shih 2013 ^{b)} ; Costa and Kahn 2013 ^{b)} ; Handgraaf, van Lidth de Jeude, and Appelt 2013 ^{b)} ; Delmas and Lessem 2014 ^{b)} ; Dolan and Metcalfe 2015 ^{b)} ; Andor et al. 2020 ^{b)} ; Bonan et al. 2020 ^{b)} Jaime Torres and Carlsson 2018; Bhanot 2021 ^{e)}
Perceived social injunctive norm	Subjects are asked to guess what others consider to be appropriate	Subjects adapt their behavior towards their guess about the injunctive norm	Krupka and Weber 2009; Bicchieri and Chavez 2010	-
Objective personal descriptive norm	Subjects receive information about their own past behavior	Information increases prosocial behavior Increase is larger for subjects with low baseline behavior	-	Kantola, Syme, and Campbell 1984 ^{b)} ; Schultz et al. 2007 ^{b)} ; Allcott 2011 ^{b)} ; Peschiera and Taylor 2012 ^{b)} ; Ayres, Raseman, and Shih 2013 ^{b)} ; Costa and Kahn 2013 ^{b)} ; Handgraaf, van Lidth de Jeude, and Appelt 2013 ^{b)} ; Delmas and Lessem 2014 ^{b)} ; Dolan and Metcalfe 2015 ^{b)} ;

Norm	Norm implementation	Main results	Laboratory studies	Field studies
				Schultz et al. 2015 ^{b)} ; Shen, Cui, and Fu 2015 ^{b)} ; Alberts et al. 2016 ^{b)} ; Anderson et al. 2017 ^{b)} ; Andor et al. 2020 ^{b)} ; Bonan et al. 2020 ^{b)} Ferraro, Miranda, and Price 2011 ^{c)} ; Ferraro and Price 2013 ^{c)} ; Tiefenbeck et al. 2013 ^{c)} ; Bernedo, Ferraro, and Price 2014 ^{c)} ; Seyranian, Sinatra, and Polikoff 2015 ^{c)} ; Hahn et al. 2016 ^{c)} ; Jaime Torres and Carlsson 2018 ^{c)} ; Bhanot 2021 ^{c)} Schultz 1999 ^{f)}
Perceived personal descriptive norm	Subjects are asked to remember or report their own past behavior	-	-	Panzone et al. 2021b ^{d)}
Perceived personal injunctive norm	Subjects receive information about what they themselves considered to be appropriate in the past and how they actually behave	Subjects adapt their behavior if their previously stated personal injunctive norm is in conflict with their actual behavior	-	Kantola, Syme, and Campbell 1984 ^{b)}

Note. Letters indicate the environmentally relevant behavior. a) reuse; b) energy consumption; c) water consumption; d) food choice; e) transportation; f) recycling; g) littering; h) waste avoidance.

5. Future research on norms

Compared to the effects of norms on individual behavior, we still know little about how norms affect each other. Only a few studies have examined how information about objective social norms affects individuals' perceptions of social norms (Bicchieri and Xiao 2009; Goeschl et al. 2018) or how social injunctive norms affect personal injunctive norms (Bertoldo and Castro 2016; d'Adda et al. 2020). The analysis of personal norms is generally more challenging than the analysis of social norms because we need different information about the same person: either how their behavior changes over time or how their views and behavior differ. The latter comparison may entail the problem that subjects may give socially desirable answers. One solution might be to measure the extent to which subjects intervene in the environmental or prosocial decisions of others, with the assumption that interventions are made only when one's

own personal injunctive norms have been violated (Fehr and Fischbacher 2003, 2004; Lieberman and Linke 2007). A perhaps more reliable method consists of the application of neuroscience to measure brain activity when subjects change their behavior after receiving norm messages (Falk et al. 2010), shift or refuse to shift their attitudes (Berns 2005; Yomogida 2017) or punish norm violators (de Quervain et al. 2004).

When and why people accept a norm as their own personal standard for appropriate behavior is probably one of the most important questions, because only then can we assume that people will adapt their behavior and help enforce the norm in society in the long term. Relatively little is known about how personal injunctive norms are constructed, and under what circumstances, in what ways, and how often they change. Research on whether the introduction or abolition of rules changes people's views about appropriate behavior may serve as a stimulus. For example, plastic bag consumption in England was reduced after the introduction of a fee, not only because of the higher price, but also because of changes in consumer attitudes (Larcom, Panzone, and Swanson 2019). Voter turnout in Switzerland was influenced more by the abolition of the voting duty than by the possibility of postal voting, although the latter had a much larger effect on the costs (Funk 2007). Lab experiments confirm that the introduction of rules and fees affects willingness to cooperate and beliefs in others' cooperation, even when they do not eliminate incentives to be free-riders (Tyran and Feld 2006; Galbiati and Vertova 2014; Romaniuc 2016; Dannenberg and Gallier 2020). The duration of the behavioral adjustment must also be taken into account. For example, a public library's reminder to return books on time had only a short-term effect on return behavior (Apesteguia, Funk, and Iriberrri 2013). A better understanding of this black box of internalization processes will be key to providing social information and using other interventions to change environmental behaviors at sufficient scale for more sustainable development. Empirical research can help better predict the impact of interventions, such as when an intervention will promote norm internalization and reinforce desired behavior, as observed for the reduction of plastic bags (Convery, McDonnell, and Ferreira 2007), or when it will do the opposite, as in the famous example of late pickups at the Haifa school (Gneezy and Rustichini 2000) – in which imposing a fee for picking up children late resulted in parents deciding to pay for the extra child care time. Such research can also help determine what kind of interventions will spark consumer experimentation (Larcom, Rauch, and Willems 2017) and willingness to try new things not yet widely used.

It is also essential to examine the effects of norms when they are in conflict with each other. Perceived social descriptive norms seem to have a stronger effect than perceived social injunctive norms when they are in conflict with each other (Cialdini, Reno, and Kallgren 1990;

Keizer, Lindenberg, and Steg 2008; Bicchieri and Xiao 2009). Social injunctive norms may be more influential when subjects must expect social reactions, such as approval or disapproval, while social descriptive norms may be more influential when such reactions do not occur or are not visible. Lab experiments use a wide range of tools to investigate and compare different combinations of norms and contexts (Dal Bó and Dal Bó 2014; Schram and Charness 2015; Bicchieri, Dimant, and Xiao 2021).

Finally, the alignment of views and behavior through social norms can also have adverse effects when this happens within segregated groups in a polarized society (Stewart et al. 2019; Green et al. 2020; Druckman et al. 2021; Bühren and Dannenberg 2021; Vasconcelos et al. 2021). In the case of climate protection, or the fight against the coronavirus pandemic, measures are sometimes taken or not taken for political or ideological reasons. These are telling examples of how the emergence of a social norm in one group can reduce the likelihood of the same norm emerging in other groups. A systematic investigation of social networks and reference groups, variations in how groups adhere to norms, the interactions between political elites and the public, and the role of the media and institutions remain important issues for future research on norms.

6. Conclusion

A better understanding of how and under what circumstances norms influence environmental behavior is important, not only for those studying human behavior but also for policymakers. When the effects of social norms are taken into account, stronger and sometimes different interventions are generally appropriate (Kinzig et al. 2013; Nyborg et al. 2016; Frank 2020). A seemingly inefficient policy to promote the diffusion of low-emission cars can become efficient if one considers that a person's decision to buy a low-emission car also influences his or her friends' car choices in that direction (e.g., Müller and von Wangenheim 2017; Ulph and Ulph 2021). We hope that this paper will help identify research gaps in this important area and equip policymakers with better knowledge to take more targeted actions to influence environmental behavior.

References

- Ajzen, Icek. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50 (2): 179–211.
- Akerlof, George A., and Rachel E. Kranton. 2000. Economics and identity. *The Quarterly Journal of Economics* 115 (3): 715–53.
- Alberts, Genevieve, Zeynep Gurguc, Pantelis Koutroumpis, Ralf Martin, Mirabelle Muûls, and Tamaryn Napp. 2016. Competition and norms: a self-defeating combination? *Energy Policy* 96: 504–23.
- Allcott, Hunt. 2011. Social norms and energy conservation. *Journal of Public Economics* 95 (9-10): 1082–95.
- Allcott, Hunt, and Todd Rogers. 2014. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. *American Economic Review* 104 (10): 3003–37.
- Anderson, Kyle, Kwonsik Song, SangHyun Lee, Erin Krupka, Hyunsoo Lee, and Moonseo Park. 2017. Longitudinal analysis of normative energy use feedback on dormitory occupants. *Applied Energy* 189: 623–39.
- Andor, Mark A., Andreas Gerster, Jörg Peters, and Christoph M. Schmidt. 2020. Social norms and energy conservation beyond the US. *Journal of Environmental Economics and Management* 103: 102351.
- Apestequia, Jose, Patricia Funk, and Nagore Iriberry. 2013. Promoting rule compliance in daily-life: evidence from a randomized field experiment in the public libraries of Barcelona. *European Economic Review* 64: 266–84.
- Aronson, Elliot. 1992. The return of the repressed: Dissonance theory makes a comeback. *Psychological Inquiry* 3: 303–311.
- Ayres, Ian, Sophie Raseman, and Alice Shih. 2013. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics, & Organization* 29 (5): 992–1022.
- Bandura, Albert. 1999. Social cognitive theory: an agentic perspective. *Asian Journal of Social Psychology* 2 (1): 21–41.
- Bandura, Albert. 2001. Social cognitive theory: an agentic perspective. *Annual Review of Psychology* 52: 1–26.
- Bator, Renée J., Angela D. Bryan, and P. Wesley Schultz. 2011. Who gives a hoot?: Intercept surveys of litterers and disposers. *Environment and Behavior* 43 (3): 295–315.

Bem, Daryl J. 1967. Self-perception: an alternative interpretation of cognitive dissonance phenomena. *Psychological Review* 74 (3): 183–200.

Bem, Daryl J. 1972. Self-perception theory. In *Advances in Experimental Social Psychology*. ed. Berkowitz, L., 1–62. Academic Press.

Bénabou, Roland, and Jean Tirole. 2006. Incentives and prosocial behavior. *American Economic Review* 96 (5): 1652–78.

Bernedo, María, Paul J. Ferraro, and Michael Price. 2014. The persistent impacts of norm-based messaging and their implications for water conservation. *Journal of Consumer Policy* 37 (3): 437–52.

Berns, Gregory S., Jonathan Chappelow, Caroline F. Zink, Giuseppe Pagnoni, Megan E. Martin-Skurski, and Jim Richards. 2005. Neurobiological correlates of social conformity and independence during mental rotation. *Biological Psychiatry* 58 (3): 245–53.

Bertoldo, Raquel, and Paula Castro. 2016. The outer influence inside us: exploring the relation between social and personal norms. *Resources, Conservation and Recycling* 112: 45–53.

Bhanot, Syon P. 2021. Isolating the effect of injunctive norms on conservation behavior: new evidence from a field experiment in California. *Organizational Behavior and Human Decision Processes* 163: 30–42.

Bicchieri, Cristina. 2005. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Bicchieri, Cristina, and Alex K. Chavez. 2010. Behaving as expected: public information and fairness norms. *Journal of Behavioral Decision Making* 23 (2):161–78.

Bicchieri, Cristina, and Eugen Dimant. 2019. Nudging with care: the risks and benefits of social information. *Public Choice* 191: 443–64.

Bicchieri, Cristina, Eugen Dimant, Simon Gächter, and Daniele Nosenzo. 2022. Social proximity and the erosion of norm compliance. *Games and Economic Behavior* 132: 59–72.

Bicchieri, Cristina, Eugen Dimant, and Erte Xiao. 2021. Deviant or wrong? The effects of norm information on the efficacy of punishment. *Journal of Economic Behavior & Organization* 188: 209–35.

Bicchieri, Cristina, and Erte Xiao. 2009. Do the right thing: but only if others do so. *Journal of Behavioral Decision Making* 22 (2): 191–208.

Bohner, Gerd, and Lena E. Schlüter. 2014. A room with a viewpoint revisited: descriptive norms and hotel guests' towel reuse behavior. *PloS one* 9 (8): e104086.

- Bonan, Jacopo, Cristina Cattaneo, Giovanna d'Adda, and Massimo Tavoni. 2020. The interaction of descriptive and injunctive social norms in promoting energy conservation. *Nature Energy* 5 (11): 900–09.
- Bühren, Christoph, and Astrid Dannenberg. 2021. The demand for punishment to promote cooperation among like-minded people. *European Economic Review*, 138: 103862.
- Carbonara, Emanuela, Francesco Parisi, and Georg von Wangenheim. 2008. Lawmakers as norm entrepreneurs. *Review of Law & Economics* 4 (3): 779–99.
- Carpenter, Jeffrey P. 2004. When in Rome: conformity and the provision of public goods. *Journal of Socio-Economics* 33 (4): 395–408.
- Carrico, Amanda R., and Manuel Riemer. 2011. Motivating energy conservation in the workplace: an evaluation of the use of group-level feedback and peer education. *Journal of Environmental Psychology* 31 (1): 1–13.
- Cialdini, Robert B., and Noah J. Goldstein. 2004. Social influence: compliance and conformity. *Annual Review of Psychology* 55: 591–621.
- Cialdini, Robert B., Reno R. Reno, and Carl A. Kallgren. 1990. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58 (6): 1015–26.
- Convery, Frank, Simon McDonnell, and Susana Ferreira. 2007. The most popular tax in Europe? Lessons from the Irish plastic bags levy. *Environmental and Resource Economics* 38 (1): 1–11.
- Cooter, Robert. 1998. Expressive law and economics. *Journal of Legal Studies* 27 (S2): 585–607.
- Cooter, Robert. 2000. Do good laws make good citizens? An economic analysis of internalized norms. *Virginia Law Review* 86 (8): 1577–601.
- Costa, Dora L., and Matthew E. Kahn. 2013. Energy conservation “nudges” and environmentalist ideology: evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association* 11 (3): 680–702.
- d'Adda, Giovanna, Martin Dufwenberg, Francesco Passarelli, and Guido Tabellini. 2020. Social norms with private values: theory and experiments. *Games and Economic Behavior* 124: 288–304.
- Dal Bó, Ernesto, and Pedro Dal Bó. 2014. “Do the right thing:” the effects of moral suasion on cooperation. *Journal of Public Economics* 117: 28–38.
- Dannenberg, Astrid, and Carlo Gallier. 2020. The choice of institutions to solve cooperation problems: a survey of experimental research. *Experimental Economics* 23 (3): 716–49.

- Dannenberg, Astrid, and Eva Weingärtner. 2022. The effects of observability and an information nudge on food choice. MAGKS Discussion Paper Series.
- De Dominicis, Stefano, Rebecca Sokoloski, Christine M. Jaeger, and P. Wesley Schultz. 2019. Making the smart meter social promotes long-term energy conservation. *Palgrave Communications* 5 (1): 1-8.
- de Groot, Judith I. M., Wokje Abrahamse, and Kayleigh Jones. 2013. Persuasive normative messages: the influence of injunctive and personal norms on using free plastic bags. *Sustainability* 5 (5): 1829–44.
- Delmas, Magali A., and Neil Lessem. 2014. Saving power to conserve your reputation? The effectiveness of private versus public information. *Journal of Environmental Economics and Management* 67 (3): 353–70.
- Demarque, Christophe, Laetitia Charalambides, Denis J. Hilton, and Laurent Waroquier. 2015. Nudging sustainable consumption: the use of descriptive norms to promote a minority behavior in a realistic online shopping environment. *Journal of Environmental Psychology* 43: 166–74.
- de Quervain, Dominique J.-F., Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, Alfred Buck, and Ernst Fehr. 2004. The neural basis of altruistic punishment. *Science* 305 (5688): 1254-58.
- Dimant, Eugen. 2019. Contagion of pro- and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology* 73: 66–88.
- Dolan, Paul, and Robert Metcalfe. 2015. Neighbors, knowledge, and nuggets: two natural field experiments on the role of incentives on energy conservation. Becker Friedman Institute for Economics Working Paper No. 2589269, Becker Friedman Institute for Economics, Chicago. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2589269.
- Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. 2021. Affective polarization, local contexts and public opinion in America. *Nature Human Behavior* 5 (1): 28–38.
- Ehrlich, Paul R., and Simon A. Levin. 2005. The evolution of norms. *PLoS Biology* 3 (6): e194.
- Einhorn, Laura 2020. Normative social influence on meat consumption. MPIfG Discussion Paper (20/1). Max-Planck-Institut für Gesellschaftsforschung (MPIfG), Köln. Available at EconStor: <https://www.econstor.eu/handle/10419/215424>.
- Ellickson, Robert C. 2001. The market for social norms. *American Law and Economics Review* 3 (1): 1–49.
- Elliott, Gregory C. 1986. Self-esteem and self-consistency: a theoretical and empirical link between two primary motivations. *Social Psychology Quarterly* 49 (3): 207-18.

Elster, Jon 1989. Social norms and economic theory. *Journal of Economic Perspectives* 3 (4): 99–117.

Falk, Emily B., Elliot T. Berkman, Traci Mann, Brittany Harrison, and Matthew Lieberman. 2010. Predicting persuasion-induced behavior change from the brain. *Journal of Neuroscience* 30 (25): 8421–24.

Farrow, Katherine, Gilles Grolleau, and Lisette Ibanez. 2017. Social norms and pro-environmental behavior: a review of the evidence. *Ecological Economics* 140: 1–13.

Fehr, Ernst, and Urs Fischbacher. 2003. The nature of human altruism. *Nature* 425 (6960): 785–91.

Fehr, Ernst, and Urs Fischbacher. 2004. Third-party punishment and social norms. *Evolution and Human Behavior* 25 (2): 63–87.

Ferraro, Paul J., Juan Jose Miranda, and Michael K. Price. 2011. The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy Experiment. *American Economic Review* 101 (3): 318–22.

Ferraro, Paul J., and Michael K. Price. 2013. Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics* 95 (1): 64–73.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Fishbein, Martin, and Icek Ajzen. 1981. Attitudes and voting behavior: an application of the theory of reasoned action. In *Progress in applied social psychology*, eds. Stephenson, G. M., and J. M. Davis, vol. 2, 253–313. London: Wiley.

Frank, Robert H. 2020. *Under the Influence – Putting Peer Pressure to Work*. Princeton and Oxford: Princeton University Press.

Frick, Vivian, Maike Gossen, Tilman Santarius, and Sonja Geiger. 2021. When your shop says #lessismore. Online communication interventions for clothing sufficiency. *Journal of Environmental Psychology* 75: 101595.

Funk, Patricia. 2007. Is there an expressive function of law? An empirical analysis of voting laws with symbolic fines. *American Law and Economics Review* 9 (1): 135–59.

Gächter, Simon, Leonie Gerhards, and Daniele Nosenzo. 2017. The importance of peers for compliance with norms of fair sharing. *European Economic Review* 97: 72–86.

Galbiati, Roberto, and Pietro Vertova. 2014. How laws affect behavior: obligations, incentives and cooperative behavior. *International Review of Law and Economics* 38: 48–57.

Gneezy, Uri, and Aldo Rustichini. 2000. A fine is a price. *The Journal of Legal Studies* 29 (1): 1-17.

Goeschl, Timo, Sara Elisa Kettner, Johannes Lohse, and Christiane Schwieren. 2018. From social information to social norms: evidence from two experiments on donation behaviour. *Games* 9 (4): 91.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius. 2008. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research* 35 (3): 472–82.

Gravert, Christina, and Linus Olsson Collentine. 2021. When nudges aren't enough: Norms, incentives and habit formation in public transport usage. *Journal of Economic Behavior & Organization* 190: 1–14.

Green, Jon, Jared Edgerton, Daniel Naftel, Kelsey Shoub, and Skyler J. Cranmer. 2020. Elusive consensus: polarization in elite communication on the COVID-19 pandemic. *Science Advances* 6 (28): eabc2717.

Griesoph, Amelie, Stefan Hoffmann, Christine Merk, Katrin Rehdanz, and Ulrich Schmidt. 2021. Guess what ...?—How guessed norms nudge climate-friendly food choices in real-life settings. *Sustainability* 13 (15): 8669.

Gromet, Dena M., Howard Kunreuther, and Richard P. Larrick. 2013. Political ideology affects energy-efficiency attitudes and choices. *Proceedings of the National Academy of Sciences of the United States of America* 110 (23): 9314–19.

Hahn, Robert, Robert Metcalfe, David Novgorodsky, and Michael K. Price. 2016. The behavioralist as policy designer: the need to test multiple treatments to meet multiple targets. NBER Working Paper No. 22886, National Bureau of Economic Research, Cambridge, MA.

Hamann, Karen R. S., Gerhard Reese, Daniel Seewald, and Daniel C. Loeschinger. 2015. Affixing the theory of normative conduct (to your mailbox): injunctive and descriptive norms as predictors of anti-ads sticker use. *Journal of Environmental Psychology* 44: 1–9.

Handgraaf, Michael J. J., Margriet A. Van Lidth de Jeude, and Kirstin C. Appelt. 2013. Public praise vs. private pay: effects of rewards on energy conservation in the workplace. *Ecological Economics* 86: 86–92.

Harmon-Jones, Eddie, David M. Amodio, and Cindy Harmon-Jones. 2009. Action-based model of dissonance: A review, integration, and expansion of conceptions of cognitive conflict. *Advances in Experimental Social Psychology* 41: 119–166. Heider, F. 1946. Attitudes and cognitive organization. *Journal of Psychology* 21: 107–12.

Heider, Fritz. 1958. *The Psychology of Interpersonal Relations*. New York: Wiley.

- Hiller Connell, Kim Y., and Joy M. Kozar. 2012. Social normative influence: An exploratory study investigating its effectiveness in increasing engagement in sustainable apparel-purchasing behaviors. *Journal of Global Fashion Marketing* 3 (4): 172-79.
- Hoffman, Martin L. 2000. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press.
- Hustvedt, Gwendolyn, and John C. Bernard. 2010. Effects of social responsibility labelling and brand on willingness to pay for apparel. *International Journal of Consumer Studies* 34 (6): 619-26.
- Ito, Koichiro, Takanori Ida, and Makoto Tanaka. 2018. Moral suasion and economic incentives: field experimental evidence from energy demand. *American Economic Journal: Economic Policy* 10 (1): 240–67.
- Jacobson, Ryan P., Chad R. Mortensen, and Robert B. Cialdini. 2011. Bodies obliged and unbound: differentiated response tendencies for injunctive and descriptive social norms. *Journal of Personality and Social Psychology* 100 (3): 433–48.
- Jagau, Henrik L., and Jana Vyrastekova. 2017. Behavioral approach to food waste: an experiment. *British Food Journal* 119 (4): 882–94.
- Jaime Torres, Mónica M., and Fredrik Carlsson. 2018. Direct and spillover effects of a social information campaign on residential water-savings. *Journal of Environmental Economics and Management* 92: 222–43.
- Kallbekken, Steffen, and Håkon Sælen. 2013. ‘Nudging’ hotel guests to reduce food waste as a win-win environmental measure. *Economics Letters* 119 (3): 325-27.
- Kantola, Steven J., Geoff J. Syme, and Norm A. Campbell. 1984. Cognitive dissonance and energy conservation. *Journal of Applied Psychology* 69 (3): 416–21.
- Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2008. The spreading of disorder. *Science* 322 (5908): 1681–85.
- Kelley, Harold H. 1967. Attribution theory in social psychology. In *Nebraska Symposium on Motivation* 15, ed. Levine, D., 192–238. University of Nebraska Press.
- Kim, Soo H., and Yoo-Kyoung Seock. 2019. The roles of values and social norm on personal norms and pro-environmentally friendly apparel product purchasing behavior: The mediating role of personal norms. *Journal of Retailing and Consumer Services* 51: 83-90.
- Kinzig, Ann P., Paul R. Ehrlich, Lee J. Alston, Kenneth Arrow, Scott Barrett, Timothy G. Buchman, Gretchen C. Daily, et al. 2013. Social norms and global environmental challenges: The complex interaction of behaviors, values, and policy, *BioScience* 63 (3): 164-75.
- Knight Lapinski, Maria, and Rajiv N. Rimal. 2005. An explication of social norms, *Communication Theory* 15 (2): 127–47.

Kohlberg, Lawrence. 1964. Development of moral character and moral ideology. In *Review of Research in Child Development*, eds. Hoffman, M. L., and L. W. Hoffman, 383-432. New York: Russell Sage Foundation.

Kohlberg, Lawrence. 1978. Revisions in the theory and practice of moral development. *New Directions for Child and Adolescent Development* 1978 (2): 83–87.

Krupka, Erin, and Roberto A. Weber. 2009. The focusing and informational effects of norms on prosocial behavior. *Journal of Economic Psychology* 30 (3): 307–20.

Kumar, Bipul, Ajay K. Manrai, and Lalita A. Manrai. 2017. Purchasing behaviour for environmentally sustainable products: A conceptual framework and empirical study. *Journal of Retailing and Consumer Services* 34: 1-9.

Kunda, Ziva. 1990. The case for motivated reasoning. *Psychological Bulletin* 108 (3): 480–98.

Larcom, Shaun, Luca A. Panzone, and Timothy Swanson. 2019. Follow the leader? Testing for the internalization of law. *The Journal of Legal Studies* 48 (1): 217-44.

Larcom, Shaun, Ferdinand Rauch, and Tim Willems. 2017. The benefits of forced experimentation: Striking evidence from the London underground network. *Quarterly Journal of Economics* 132 (4): 2019-55.

Lieberman, Debra, and Lance Linke. 2007. The effect of social category on third party punishment. *Evolutionary Psychology* 5 (2): 289-305.

Lin, Szu-Tung, and Han-Jen Niu. 2018. Green consumption: Environmental knowledge, environmental consciousness, social norms, and purchasing behavior. *Business Strategy and the Environment* 27 (8): 1679-88.

Linder, Noah, Therese Lindahl, and Sara Borgström. 2018. Using behavioural insights to promote food waste recycling in urban households - evidence from a longitudinal field experiment. *Frontiers in Psychology* 9: 352.

Lo, Carolyn J., Yelena Tsarenko, and Dewi Tojib. 2019. To tell or not to tell? The roles of perceived norms and self-consciousness in understanding consumers' willingness to recommend online secondhand apparel shopping. *Psychology & Marketing* 36 (4): 287-304.

Morris, Michael W., Ying-Yi Hong, Chi-Yue Chiu, and Zhi Liu. 2015. Normology: integrating insights about social norms to understand cultural dynamics. *Organizational Behavior and Human Decision Processes* 129: 1–13.

Müller, Stephan, and Georg von Wangenheim. 2017. The impact of market innovations on the dissemination of social norms: the sustainability case. *Journal of Evolutionary Economics* 27 (4): 663–90.

- Nolan, Jessica M., P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein, and Vidas Griskevicius. 2008. Normative social influence is underdetected. *Personality and Social Psychology Bulletin* 34 (7): 913–23.
- Nowak, Martin A., and Karl Sigmund. 2005. Evolution of indirect reciprocity. *Nature* 437 (7063): 1291–98.
- Nyborg, Karine. 2018. Social norms and the environment. *Annual Review of Resource Economics* 10: 405–23.
- Nyborg, Karine, John M. Anderies, Astrid Dannenberg, Therese Lindahl, Caroline Schill, Maja Schlüter, W. Neil Adger, et al. 2016. Social norms as solutions. *Science* 354 (6308): 42–43.
- Nyborg, Karine, and Mari Rege. 2003. On social norms: the evolution of considerate smoking behavior. *Journal of Economic Behavior & Organization* 52 (3): 323–40.
- Oceja, Luis, and Jaime Berenguer. 2009. Putting text in context: the conflict between pro-ecological messages and anti-ecological descriptive norms. *Spanish Journal of Psychology* 12 (2): 657–66.
- Ohtsuki, Hisashi, and Yoh Iwasa. 2004. How should we define goodness? - Reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology* 231 (1): 107–20.
- Okada, Isamu. 2020. A review of theoretical studies on indirect reciprocity. *Games* 11 (3): 27.
- Panzone, Luca A., Alistair Ulph, Denis Hilton, Ilse Gortemaker, and Ibrahim A. Tajudeen. 2021a. Sustainable by design: choice architecture and the carbon footprint of grocery shopping. *Journal of Public Policy & Marketing* 40 (4): 463–86.
- Panzone, Luca A., Alistair Ulph, Daniel J. Zizzo, Denis Hilton, and Adrian Clear. 2021b. The impact of environmental recall and carbon taxation on the carbon footprint of supermarket shopping. *Journal of Environmental Economics and Management* 109: 102137.
- Park, Hyun J., and Li M. Lin. 2020. Exploring attitude–behavior gap in sustainable consumption: Comparison of recycled and upcycled fashion products. *Journal of Business Research* 117: 623–28.
- Peschiera, Gabriel, and John E. Taylor. 2012. The impact of peer network position on electricity consumption in building occupant networks utilizing energy feedback systems. *Energy and Buildings* 49: 584–90.
- Piaget, Jean. 1965[1932]. *The Moral Judgment of the Child*. New York: The Free Press.
- Raihani, Nichola J., and Katherine McAuliffe. 2014. Dictator game giving: the importance of descriptive versus injunctive norms. *PloS one* 9 (12): e113826.

- Reese, Gerhard, Daniel C. Loeschinger, Karen Hamann, and Sebastian Neubert. 2013. Sticker in the box! Object-person distance and descriptive norms as means to reduce waste. *Ecopsychology* 5 (2): 146–48.
- Reese, Gerhard, Kristina Loew, and Georges Steffgen. 2014. A towel less: social norms enhance pro-environmental behavior in hotels. *Journal of Social Psychology* 154 (2): 97–100.
- Rege, Mari. 2004. Social norms and private provision of public goods. *Journal of Public Economic Theory* 6 (1): 65-77.
- Reiter, Susan M., and William Samuel. 1980. Littering as a function of prior litter and the presence or absence of prohibitive signs. *Journal of Applied Social Psychology* 10 (1): 45-55.
- Richter, Isabel, John Thøgersen, and Christian Klöckner. 2018. A social norms intervention going wrong: boomerang effects from descriptive norms information. *Sustainability* 10 (8): 2848.
- Romaniuc, Rustam. 2016. What makes law to change behavior? An experimental study. *Review of Law & Economics* 12 (2): 447-75.
- Sagebiel, Julian, Lukas Karok, Julian Grund, and Jens. Rommel. 2020. Clean environments as a social norm: a field experiment on cigarette littering. *Environmental Research Communications* 2 (9): 091002.
- Schram, Arthur, and Gary Charness. 2015. Inducing social norms in laboratory allocation choices. *Management Science* 61 (7): 1531–46.
- Schultz, P. Wesley. 1999. Changing behavior with normative feedback interventions: a field experiment on curbside recycling. *Basic and Applied Social Psychology* 21 (1): 25–36.
- Schultz, P. Wesley, Mica Estrada, Joseph Schmitt, Rebecca Sokoloski, and Nilmini Silva-Send. 2015. Using in-home displays to provide smart meter feedback about household electricity consumption: a randomized control trial comparing kilowatts, cost, and social norms. *Energy* 90 (1): 351–58.
- Schultz, P. Wesley, Azar M. Khazian, and Adam C. Zaleski. 2008. Using normative social influence to promote conservation among hotel guests. *Social Influence* 3 (1): 4–23.
- Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vlaslas Griskevicius. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychological Science* 18 (5): 429–34.
- Schwartz, Shalom H. 1977. Normative influences on altruism. *Advances in Experimental Social Psychology* 10: 221–79.
- Schwartz, Shalom H., and Judith A. Howard. 1982. Helping and cooperation: a self-based motivational model. In *Cooperation and Helping Behavior: Theories and Research*, eds. Derlega, V. J., and J. Grzelak, 327–53. New York: Academic Press.

- Seyranian, Viviane, Gale M. Sinatra, and Morgan S. Polikoff. 2015. Comparing communication strategies for reducing residential water consumption. *Journal of Environmental Psychology* 41: 81–90.
- Sheeran, Paschal. 2002. Intention-behavior relations: a conceptual and empirical review. *European Review of Social Psychology* 12 (1): 1–36.
- Shen, Meng, Qingbin Cui, and Liping Fu. 2015. Personality traits and energy conservation. *Energy Policy* 85: 322–34.
- Smith, Joanne R., Winnifred R. Louis, Deborah J. Terry, Katharine Greenaway, Miranda R. Clarke, and Xiaoliang Cheng. 2012. Congruent or conflicted? The impact of injunctive and descriptive norms on environmental intentions. *Journal of Environmental Psychology* 32 (4): 353–61.
- Sparkman, Gregg, and Gregory M. Walton. 2017. Dynamic norms promote sustainable behavior, even if it is counternormative. *Psychological Science* 28 (11): 1663–74.
- Stern, Paul C., Thomas Dietz, Troy D. Abel, Gregory A. Guagnano, and Linda Kalof. 1999. A value-belief-norm theory of support for social movements: the case of environmentalism. *Human Ecology Review* 6 (2): 81–97.
- Stewart, Alexander J., Mohsen Mosleh, Marina Diakonova, Antonio A. Arechar, David G. Rand, and Joshua B. Plotkin. 2019. Information gerrymandering and undemocratic decisions. *Nature* 573 (7772): 117–21.
- Stöckli, Sabrina, Michael Dorn, and Stefan Liechti. 2018. Normative prompts reduce consumer food waste in restaurants. *Waste Management* 77: 532–36.
- Sugden, Robert. 2000. The motivating power of expectations. In *Rationality, Rules, and Structure. Theory and Decision Library*, eds. Nida-Rümelin, J., and W. Spohn, vol. 28, 103–29. Dordrecht: Springer.
- Sunstein, Cass R. 1996. Social norms and social roles. *Columbia Law Review* 96 (4): 903–68.
- Sussman, Reuven, and Robert Gifford. 2013. Be the change you want to see. *Environment and Behavior* 45 (3): 323–43.
- Thöni, Christian, and Simon Gächter. 2015. Peer effects and social preferences in voluntary cooperation: a theoretical and experimental analysis. *Journal of Economic Psychology* 48: 72–88.
- Tiefenbeck, Verena, Thorsten Staake, Kurt Roth, and Olga Sachs. 2013. For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy* 57: 160–71.

Thøgersen, John. 2006. Norms for environmentally responsible behaviour: An extended taxonomy. *Journal of environmental Psychology* 26 (4): 247-261.

Tyler, Tom R. 1990. *Why People Obey the Law*. Yale University Press.

Tyler, Tom R., and Yuen J. Huo. 2002. *Trust in the Law: Encouraging Public Cooperation with the Police and Courts*. Russell Sage Foundation.

Tyran, Jean-Robert, and Lars P. Feld. 2006. Achieving compliance when legal sanctions are non-deterrent. *The Scandinavian Journal of Economics* 108 (1): 135-56.

Ulph, Alistair, and David Ulph. 2021. Environmental policy when consumers value conformity, *Journal of Environmental Economics and Management* 109: 102172.

Vasconcelos, Vítor V., Sara M. Constantino, Astrid Dannenberg, Marcel Lumkowsky, Elke Weber, and Simon Levin. 2021. Segregation and Clustering of Preferences Erode Socially Beneficial Coordination. *Proceedings of the National Academy of Sciences USA* 118 (50): e2102153118.

White, Katherine, and Bonnie Simpson. 2013. When do (and don't) normative appeals influence sustainable consumer behaviors? *Journal of Marketing* 77 (2): 78-95.

White, Katherine, Rishad Habib, and David J. Hardisty. 2019. How to SHIFT consumer behaviors to be more sustainable: A literature review and guiding framework. *Journal of Marketing* 83 (3): 22-49.

Yomogida, Yukihiro, Madoka Matsumoto, Ryuta Aoki, Aayaka Sugiura, Adam N. Phillips, and Kenji Matsumoto. 2017. The neural basis of changing social norms through persuasion. *Scientific Reports* 7: 16295.

Young, H. Peyton. 1993. The evolution of conventions. *Econometrica* 61 (1): 57-84.

Young, H. Peyton. 2015. The evolution of social norms. *Annual Review of Economics* 7: 359–87.

Acknowledgments

The work was financially supported by the project ZumWert funded through the University of Kassel, Germany. We thank Marissa Reiserer for the design of Figure 1.

Appendix to

“The Effects of Norms on Environmental Behavior”

1. Taxonomy of norms
 2. A conceptual structure on the causes of norm effects
- References

1. Taxonomy of norms

Table 1, shown in the main paper, illustrates the taxonomy we use in our review. The taxonomy builds upon and merges existing conceptualizations of norms in economics and psychology (Schwartz 1977; Schwartz and Howard 1982; Cialdini, Reno, and Kallgren 1990; Stern et al. 1999; Farrow, Grolleau, and Ibanez 2017; Nyborg 2018; Bicchieri and Dimant 2019). In the following, we explain the individual components of the taxonomy and the existing connections between them. For better illustration, we use an example of a norm that refers to the use of a bicycle instead of a car for short distances.

1.1 Quality of the norm

We define a norm as the rule that characterizes a subset of all possible behaviors as either appropriate or normal. An ‘injunctive norm’ describes what is appropriate in a certain situation. Thus, an injunctive norm describes what one ought to do in a certain situation. An individual conforms to an injunctive norm, if his or her exhibited behavior is in line with what the norm prescribes or proscribes. A ‘descriptive norm’ describes what people actually do. It captures the frequency with which a certain behavior is executed. This distinction regarding the nature or quality of the norm (Cialdini, Reno, and Kallgren 1990) is fundamental to avoid confusion between ‘is’ and ‘ought’ as there is a large difference between observing that the majority uses the bicycle for short distances or thinking that they should use it.

1.2 Subject of the norm

The second dimension in our taxonomy concerns the subject of the norm, which is arguably the most commonly used distinction between norms (Schwartz 1977; Stern et al. 1999). A ‘personal norm’ describes what an individual finds appropriate (injunctive norm) or does (descriptive

norm). For example, the personal injunctive norm captures whether an individual considers commuting by bike the appropriate thing to do. The personal descriptive norm captures whether an individual actually commutes by bike regularly. A ‘social norm’ describes what a significant proportion of the population finds appropriate (injunctive norm) or does (descriptive norm) (Schwartz 1977; Stern et al. 1999). In terms of our example, the social injunctive norm captures whether a large group of people views commuting by bike the appropriate thing to do. The social descriptive norm captures how many individuals actually commute by bike regularly.

In addition to personal and social norms, we include ‘legal norms’ in our taxonomy. They contain what legislators or courts define as appropriate behavior. Hence, legal norms are by definition of an injunctive quality. If a legislator or an otherwise competent body, like a court, has proscribed cars in certain areas of a city, the norm is legal. Unlike other authors (e.g. Farrow, Grolleau, and Ibanez 2017 or Bicchieri and Dimant 2019), we include legal norms because norms held by these different subject types may closely interact and emerge from one another as is discussed further below in our conceptual model. Many individuals following a personal injunctive norm to use the bike for commuting form a social descriptive norm and may culminate in a legal norm restricting the use of cars in certain areas. Conversely, a legal norm prohibiting the use of cars in certain areas may induce people to adopt a personal injunctive norm of using the bike which may eventually result in a corresponding social descriptive norm (see Cooter 1998 for the example of cleaning up after dogs). Despite the close connections between the subjects of norms, the conceptual separation between them greatly facilitates the understanding of norms.

1.3 Perspective on the norm

The third dimension in our taxonomy refers to the perspective on the norms. We distinguish between a ‘perceived norm,’ which refers to an individual’s subjective beliefs about behavior or attitudes, and an ‘objective norm,’ which refers to actual behavior or attitudes. Objective norms are (statistical) facts. Perceived norms capture how an individual perceives the objective norm. They are subjective, can differ between individuals, and influence the behavior of individuals. For example, an individual may perceive that cycling is a widespread norm (perceived social descriptive norm), but in fact only a minority ride bicycles and thus the objective social descriptive norm differs. The distinction between objective and perceived norms allows for the possibility that people sometimes send misleading signals, misinterpret signals from others, or simply lack information about relevant others in the population. Although any information that individuals use to make decisions is ultimately subjective, it is

useful to distinguish conceptually between perceptions that affect human behavior and statistical facts. Note that this distinction does not apply to personal injunctive norms, which we assume to be inherently subjective.

1.4 Enforcement of the norm

The fourth and final dimension in our taxonomy refers to the enforcement of a norm. This dimension captures for which reasons a norm is followed. In contrast to previous approaches (e.g., Farrow, Grolleau, and Ibanez 2017; Nyborg, 2018; Bicchieri and Dimant 2019), we disentangle the enforcement mechanism from the subject of the norm. The separation between the subject of the norm and its enforcement contains some redundancies but it aids clear thinking about what characterizes and differentiates norms. We differentiate between personal, social, and legal enforcement. ‘Personal enforcement’ refers to purely internal motives such as satisfaction, inner peace, or avoidance of guilt. This may not feel like enforcement to the person at all. A person may commute by bike because of the good feeling of doing the right thing or to avoid a guilty conscience. Commuting by bike may also be motivated by social pressure, a desire to conform with others, or anticipated negative reactions from colleagues if one does not. This pursuit of social approval or avoidance of social disapproval or sanctions is called ‘social enforcement.’ The third mechanism is ‘legal enforcement’ by an executive authority in which individuals are subjected to legally defined measures if they deviate from a legal norm. In terms of our example, individuals may refrain from commuting by car in certain cases due to the threat of financial fines.

The clear conceptual distinction between the different enforcement mechanisms is very important for understanding the effects of norms and deriving policy implications. Not only may the various enforcement mechanisms differ in their effectiveness, but they may also reinforce or hinder each other. Cooter (1998), Tyler (1990), and Tyler and Huo (2002) showed how social enforcement of social norms supports enforcement of legal norms, even if the latter by itself lacks effective enforcement. Carbonara, Parisi, and von Wangenheim (2008) show how legal enforcement of norms that diverge too much from social norms can cause the latter to move further away from the legal norm and make their social and personal enforcement even stronger.

2. A conceptual structure on the causes of norm effects

Figure 1, shown in the main paper, provides a graphical illustration of the various interactions between different types of norms and their influence on decision making. The aim of this figure is to show important channels of influence and to explain our current state of knowledge mainly on the basis of theories from social, cognitive, developmental, and motivational psychology. In this way, we create a broad conceptual framework that shows how norms can influence each other. This general framework also allows for the classification of models from behavioral economics (Lindbeck, Nyberg, and Weibull 1999; Brekke, Kverndokk, and Nyborg 2003; Azar 2004, 2005; Bénabou and Tirole 2006; Nyborg, Howarth, and Brekke 2006; d'Adda et al. 2020) and evolutionary game theory (from economics: Sugden 1989; Young 1993, 1996, 2015; Binmore and Samuelson 1994; Sethi and Somanathan 1996; Nyborg and Rege 2003; Rege 2004; Mengel 2008; Traxler and Spichtig 2011; from biology: Henrich and Boyd 2001, Gintis 2003; Bowles and Gintis 2004) which we will discuss at the end of the section.

2.1 The decision-making process

A person's decision-making process is shown in the center of the figure, illustrating that decisions are influenced by factors such as personal preferences, experience, knowledge, and the situational context. These influences are captured in an individual's expectations about the payoffs from the available behavioral options. Expectations are based on what the individual has learned so far. In most situations a person's ultimate behavioral decision is the result of weighting different expected payoffs (Fishbein and Ajzen 1975). For example, a person may enjoy riding a bike due to the physical workout but does not want to cycle in the rain. Conceptually, it is useful to distinguish between the cognitive level, namely the intention, being an unobservable outcome of an internal cognitive process, and the behavioral level, being externally visible to observers (Ajzen 1991). In some cases, intentions may not translate into actions leading to an intention-behavior gap (Sheeran 2002). It may be that a person has the intention to take the bike but in the end takes the car for situational reasons such as time constraints.

An individual eventually chooses an action. The sum of all individual actions constitutes the collective outcome. An individual's realized payoff depends on both the personal and the collective behavior. The realized payoff informs individuals about the consequences of their behavioral choices given the circumstances, which in our example might be physical workout and enjoying the good weather via cycling. Based on this feedback information, a learning

process takes place in which individuals change and adapt their expectations about payoffs, regarding how expedient they are in the given situation for future decisions (Bandura 1999). With this feedback loop, the decision-making process comes full circle.

2.2 Personal descriptive norms

We will now look at the connections between an individual's current behavior, shown in the center of the figure, and the individual's personal norms, shown on the left. An individual's current behavior by definition constitutes part of their objective personal descriptive norm. Individuals' past behaviors (their objective personal descriptive norm) provides a basis for how they perceive themselves. Individuals who use their bikes relatively often are also likely to perceive themselves as regular bike riders. Yet, this assessment may also depend on the social environment, personal aspirations, biases, and so forth. The perception of the personal descriptive norm influences future behavior through strivings for self-consistency (Elliott 1986). Thus, regularly commuting by bike introduces incentives to continue doing so in order to align current with past behavior.

2.3 Personal injunctive norms

The perception of own behavior may also influence the personal injunctive norm. According to self-perception theory (Bem 1967, 1972) and other consistency theories (Heider 1946; Festinger 1957), past behavior is evaluated in an internal post-hoc reasoning process and may be ascribed to normative beliefs. Conversely, personal norms can also influence behavior. When a personal injunctive norm is inconsistent with current behavior, feelings of inner conflict, failure, guilt, or shame may arise (Schwartz 1977; Schwartz and Howard 1981, 1982). Such feelings generate incentives to align behavior with the personal injunctive norm. According to Festinger's (1957) theory of cognitive dissonance, dissonance between a personal injunctive norm and behavior can be alleviated either by changing behavior or the norm, depending on the circumstances (e.g., Heider 1946; Elliot and Devine 1994). Hence, a person feeling uncomfortable with commuting by car may either change to commuting by bike or adapt their personal norms. In the latter case, the person would adjust his or her views on the inappropriateness of driving a car. Such post-hoc reasoning processes are more likely to happen if a person's behavior is associated with high payoffs (Kunda 1990; Ryan and Deci 2017). In other words, the more successful one's past behavior has been, the more likely it will be judged as appropriate. If choosing the car to commute yields considerable benefits such as saving time and energy, this behavior is more likely to be considered appropriate.

2.4 Social descriptive norms

Social descriptive norms are shown below the individual decision-making process in Figure 1. Individual behaviors, aggregated over time and people, constitute the objective social descriptive norm. In our bike-riding example, mobility statistics correspond to the definition of an objective social descriptive norm. Likewise, the aggregation of all individual attitudes over time and people forms the objective social injunctive norm (Cooter 1998; Carbonara, Parisi, and von Wangenheim 2008). If many individuals consider commuting by bike as appropriate, it would be the objective social injunctive norm. Both types of objective social norms are the bases for how individuals perceive social norms. Yet, the objective information is always subject to the individual's perception, which may be biased due to observation errors, motivated information seeking and reasoning (e.g., Johnston and Dark 1986; Kunda 1990), or the individual's social network (e.g., Fishbein and Ajzen 1981). As only the perceived social norm is available to an individual, it is the one that ultimately influences behaviors and attitudes.

The perceived social descriptive norm affects individuals' decision-making through conformism and imitation (Schultz et al. 2007; Schultz, Khazian, and Zaleski 2008; Smith et al. 2012). Observation of others provides clues as to what behavior is effective and adaptive (Cialdini, Reno, and Kallgren 1990). If individuals observe other people commuting by bike, they may perceive biking as an effective way to get from one place to another and therefore will be more likely to ride their bike as well. In addition to the tendency to follow others, observing others can also lead individuals to adjust their personal view of what is appropriate behavior (Miller and Dollard 1941; Bandura 2001). In other words, an individual observing others riding the bike is also more likely to consider biking as appropriate.

2.5 Social injunctive norms

Social injunctive norms are shown above the individual decision-making process in Figure 1. Individuals are motivated to behave consistently with their perceived social injunctive norms due to social enforcement, as proposed by Ajzen (1991), Elster (1989), Sunstein (1996), and Ellickson (2001), among others. According to our taxonomy, shown in Table 1, such social enforcement can take various forms, ranging from unspoken social disapproval to communicated dislike (e.g., via raised eyebrows) to material or physical punishments (e.g., ostracism, threats, violence). Possible reactions of colleagues can be a motivation to take the bike instead of the car for the trip to work.

Moreover, perceived social injunctive norms influence an individual's personal injunctive norms. This idea has broad support in the psychological literature. The process of

internalizing social injunctive norms and integrating them into one's own normative system is usually understood as a learning process (e.g., Kohlberg 1984; Hoffman 2000). Moral development theories (first described by Piaget 1932[1965] and later refined by Kohlberg 1964) describe how an individual's moral reasoning develops over time, moving from the imitation of social norms to more mature stages, in which self-chosen moral principles are obtained through moral 'cognition' (judgment and reasoning) (Kohlberg 1978, p. 84). In terms of our example, if an individual operates in a society where many others believe that commuting by bike is appropriate, then the individual is more likely to adopt a similar belief.

Similar to the close link between personal injunctive norms and individual behavior, there is a link between social injunctive norms and collective behavior. Behavior in a society often serves as an indicator of what is considered right and wrong in that society (Bicchieri 2005; Morris et al. 2015; Nyborg 2018). Accordingly, an individual's perceived social injunctive norm is (at least) partly inferred from her perceived social descriptive norms. Attribution theory (e.g., Heider 1958; Kelley 1967; Kelley and Michela 1980) states that people are motivated to assign causes to behaviors to understand past and predict future behavior and thereby make their environment seem more controllable. Therefore, it is reasonable to assume that an individual's perceptions of how others behave influence their perception of what others regard as acceptable behavior, meaning their subjective social injunctive norm. For example, an individual observing many peers commuting by car on short distances may conclude that these peers find such behavior appropriate.

2.6 Legal norms

Legal norms are shown at the top of Figure 1. Their purpose is to directly increase or decrease the expected and actual payoffs associated with the available behavioral options and to thereby regulate individuals' behaviors. But they also have effects that go beyond that. According to the expressive law theory (Cooter 1998; Cooter 2000), a newly introduced law changes people's beliefs about which behaviors are seen as appropriate and which are seen as inappropriate in society. Such beliefs may change because legislators are prominent members of the society and act as role models on the adoption of personal injunctive norms. Reference to the law facilitates social enforcement of social injunctive norms (people say "but it's against the law") which may additionally strengthen the injunctive norms. A law thus changes the perceived social injunctive norm and by this may also change an individual's personal view about what is right and what is not, independent of the altered payoffs (Tyler 1990; Tyler and Huo 2002).

A direct effect of legal norms on personal norms goes in the same direction. Many people ascribe some normative power to the law and tend to adapt their personal injunctive norms to legal norms. Larcom, Panzone, and Swanson (2019) provide a recent literature review of channels through which individuals internalize legal norms and prove their (restricted) empirical relevance. These shifts at the individual level, due to altered payoffs and attitudes, then add up to shifts at the societal level and change the objective social descriptive and injunctive norms. Although we do not elaborate on the emergence of legal norms and legislative institutions (Parisi and von Wangenheim 2006; Carbonara, Parisi, and von Wangenheim 2008; Dannenberg and Gallier 2020), we have included legal norms in Table 1 and Figure 1 for completeness and note that social descriptive and injunctive norms are important determinants of their emergence.

2.7 Exemplary classification of existing dynamic norm models

As mentioned above, we can use our conceptual framework to classify existing models of norms, which we illustrate below using three different types of models as examples. Young (2015) uses a stochastic evolutionary coordination game, in which he defines a social norm as a pattern of behavior that each member of a group conforms to, is expected to conform to, and wants to conform to, if and when they expect everyone else to do so. In this coordination game, the behavioral pattern that arises in equilibrium represents the social norm. In terms of our taxonomy and framework, Young models a decision-making process where utility is solely subject to material payoff and unaffected by any norm-related concerns that do not directly change the payoff. This is depicted in the blue rectangle in the center-right of Figure 1. The behavior resulting from the decision-making process then determines personal descriptive and social descriptive norms. Thus, Young rationalizes how social descriptive norms arise if behavioral conformity is beneficial from a material perspective. The work of Young (1993, 1996), Binmore and Samuelson (1994), Sethi and Somanathan (1996), and Bowles and Gintis (2004) can be similarly classified as social descriptive norm models.

Nyborg and Rege (2003) and Rege (2004) present models that account for normative aspects and go beyond the rationalization of behavioral patterns. They define a social norm as a rule that captures a normative order of appropriate behavior and is enforced by social sanctions. Social sanctions are endogenous in the models as they increase with the share of individuals that follow the social norm. In terms of our taxonomy and framework, this relates to the social injunctive and social descriptive norms co-influencing the decision-making process. While the social injunctive norm prescribes appropriate behavior, it is exogenously

given and thus unaffected by other norms (i.e., arrows pointing at the social injunctive norm are not included). The social descriptive norm determines the degree to which violation of the social injunctive norm triggers social disapproval. Therefore, the models incorporate the feedback loop where aggregate behavior determines descriptive norms which in turn influence the decision-making process through sanctions.

Mengel (2008) presents a model that focuses on the process of social norm internalization. She defines a social norm as a code of conduct shared by a society, which, when internalized by an individual, is enforced via internal sanctions such as shame, guilt, embarrassment, and loss of self-esteem. Moreover, Mengel assumes that individuals copy the cultural traits of peers that are successful in terms of material payoff. Thus, her model focuses on the co-dependencies of the decision-making process and personal injunctive norms. In terms of our taxonomy and framework, personal injunctive norms coincide with Mengel's understanding of an internalized social norm. An individual's personal injunctive norm directly impacts the decision-making process and, thereby, material payoffs. These material payoffs then influence the individuals' personal injunctive norms. Social injunctive norms play a subordinate role in Mengel's model since they do not influence the decision-making process directly. Since the social injunctive norm describes the distribution of personal injunctive norms across society, a strong social injunctive norm increases the probability of observing and copying individuals with the according personal injunctive norms. Thus, Mengel also incorporates the link between personal injunctive norms and social injunctive norms.

References

- Ajzen, I. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50 (2): 179–211.
- Azar, O. H. 2004. What sustains social norms and how they evolve?: The case of tipping. *Journal of Economic Behavior & Organization* 54 (1): 49-64.
- Azar, O. H. 2005. The social norm of tipping: does it improve social welfare?. *Journal of Economics* 85 (2): 141-73.
- Bandura, A. 1999. Social cognitive theory: an agentic perspective. *Asian Journal of Social Psychology* 2 (1): 21–41.
- Bandura, A. 2001. Social cognitive theory: an agentic perspective. *Annual Review of Psychology* 52: 1–26.
- Bem, D. J. 1967. Self-perception: an alternative interpretation of cognitive dissonance phenomena. *Psychological Review* 74 (3): 183–200.
- Bem, D. J. 1972. Self-perception theory. In *Advances in Experimental Social Psychology*. ed. Berkowitz, L., 1–62. Academic Press.
- Bénabou, R., and J. Tirole. 2006. Incentives and prosocial behavior. *American Economic Review* 96 (5): 1652-78.
- Bicchieri, C. 2005. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.
- Bicchieri, C., and E. Dimant. 2019. Nudging with care: the risks and benefits of social information. *Public Choice* 191: 443-64.
- Binmore, K., and L. Samuelson. 1994. An economist's perspective on the evolution of norms. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft* 150 (1): 45-63.
- Bowles, S., and H. Gintis. 2004. The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical Population Biology* 65 (1): 17-28.
- Brekke, K. A., S. Kverndokk, and K. Nyborg. 2003. An economic model of moral motivation. *Journal of Public Economics* 87 (9-10): 1967-83.
- Carbonara, E., F. Parisi, and G. von Wangenheim. 2008. Lawmakers as norm entrepreneurs. *Review of Law & Economics* 4 (3): 779-99.
- Cialdini, R. B., R. R. Reno, and C. A. Kallgren. 1990. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58 (6): 1015–26.

- Cooter, R. 1998. Expressive law and economics. *Journal of Legal Studies* 27 (S2): 585–607.
- Cooter, R. 2000. Do good laws make good citizens? An economic analysis of internalized norms. *Virginia Law Review* 86 (8): 1577-601.
- d’Adda, G., M. Dufwenberg, F. Passarelli, and G. Tabellini. 2020. Social norms with private values: theory and experiments. *Games and Economic Behavior* 124: 288–304.
- Dannenbergh, A., and C. Gallier, 2020. The choice of institutions to solve cooperation problems: a survey of experimental research. *Experimental Economics* 23 (3): 716-49.
- Ellickson, R. C. 2001. The market for social norms. *American Law and Economics Review* 3 (1): 1–49.
- Elliott, G. C. 1986. Self-esteem and self-consistency: a theoretical and empirical link between two primary motivations. *Social Psychology Quarterly* 49 (3): 207-18.
- Elliot, A. J., and P. G. Devine. 1994. On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology*, 67 (3): 382-94.
- Elster, J. 1989. Social norms and economic theory. *Journal of Economic Perspectives* 3 (4): 99–117.
- Farrow, K., G. Grolleau, and L. Ibanez. 2017. Social norms and pro-environmental behavior: a review of the evidence. *Ecological Economics* 140: 1–13.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.
- Fishbein, M., and I. Ajzen. 1975. *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, Massachusetts: Addison-Wesley.
- Fishbein, M., and I. Ajzen. 1981. Attitudes and voting behavior: an application of the theory of reasoned action. In *Progress in Applied Social Psychology*, eds. Stephenson, G. M., and J. M. Davis, vol. 2, 253–313. London: Wiley.
- Gintis, H. (2003). The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms. *Journal of Theoretical Biology* 220 (4): 407-18.
- Heider, F. 1946. Attitudes and cognitive organization. *Journal of Psychology* 21: 107–12.
- Heider, F. 1958. *The Psychology of Interpersonal Relations*. New York: Wiley.
- Henrich, J., and R. Boyd. 2001. Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of theoretical biology* 208 (1): 79-89.

- Hoffman, M. L. 2000. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press.
- Johnston, W. A., and V. J. Dark. 1986. Selective attention. *Annual Review of Psychology* 37: 43–75.
- Kelley, H. H. 1967. Attribution theory in social psychology. In *Nebraska Symposium on Motivation* 15, ed. Levine, D., 192–238. University of Nebraska Press.
- Kelley, H. H. and J. L. Michela. 1980. Attribution Theory and Research. *Annual Review of Psychology* 31 (1): 457-501.
- Kohlberg, L. 1964. Development of moral character and moral ideology. In *Review of Research in Child Development*, eds. Hoffman, M. L., and L. W. Hoffman, 383-432. New York: Russell Sage Foundation.
- Kohlberg, L. 1978. Revisions in the theory and practice of moral development. *New Directions for Child and Adolescent Development* 1978 (2): 83–87.
- Kohlberg, L. 1984. *Essays on moral development: The psychology of moral development*, vol. 2. San Francisco: Harper & Row Publishers, Inc.
- Kunda, Z. 1990. The case for motivated reasoning. *Psychological Bulletin* 108 (3): 480–98.
- Larcom, S., L. A. Panzone, and T. Swanson. 2019. Follow the leader? Testing for the internalization of law. *The Journal of Legal Studies* 48 (1): 217-44.
- Lindbeck, A., S. Nyberg and J. W. Weibull. 1999. Social norms and economic incentives in the welfare state. *The Quarterly Journal of Economics* 114 (1): 1-35.
- Mengel, F. 2008. Matching structure and the cultural transmission of social norms. *Journal of Economic Behavior & Organization* 67 (3-4): 608-23.
- Miller, N. E., and J. Dollard. 1941. *Social Learning and Imitation*. New Haven: Yale University Press.
- Morris, M. W., Y.-Y. Hong, C.-Y. Chiu, and Z. Liu. 2015. Normology: integrating insights about social norms to understand cultural dynamics. *Organizational Behavior and Human Decision Processes*. 129: 1–13.
- Nyborg, K. 2018. Social norms and the environment. *Annual Review of Resource Economics*. 10: 405–23.
- Nyborg, K., and M. Rege. 2003. On social norms: the evolution of considerate smoking behavior. *Journal of Economic Behavior & Organization*. 52 (3): 323–40.

Nyborg, K., R. B. Howarth, and K. A. Brekke. 2006. Green consumers and public policy: On socially contingent moral motivation. *Resource and energy economics*, 28 (4): 351-66.

Parisi, F., and G. von Wangenheim. 2006. Legislation and countervailing effects from social norms. In *Evolution and Design of Institutions*, eds. Schubert, C., and G. von Wangenheim, 22-55. London: Routledge.

Piaget, J. 1965[1932]. *The Moral Judgment of the Child*. New York: The Free Press.

Rege, M. 2004. Social norms and private provision of public goods. *Journal of Public Economic Theory*, 6 (1): 65-77.

Ryan, R. M., and E. L. Deci. 2017. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York: The Guilford Press.

Schultz, P. W., A. M. Khazian, and A. C. Zaleski. 2008. Using normative social influence to promote conservation among hotel guests. *Social Influence* 3 (1): 4-23.

Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychological Science* 18 (5): 429-34.

Schwartz, S. H. 1977. Normative influences on altruism. *Advances in Experimental Social Psychology* 10: 221-79.

Schwartz, S. H., and J. A. Howard. 1981. A normative decision-making model of altruism. In *Altruism and Helping Behaviour: Social, Personality and Developmental Perspectives*, ed. Rushton, J. P., 189-211. Hillsdale: Erlbaum.

Schwartz, S. H., and J. A. Howard. 1982. Helping and cooperation: a self-based motivational model. In *Cooperation and Helping Behavior: Theories and Research*, eds. Derlega, V. J., and J. Grzelak, 327-53. New York: Academic Press.

Sethi, R., and E. Somanathan. 1996. The evolution of social norms in common property resource use. *The American Economic Review* 86 (4): 766-88.

Sheeran, P. 2002. Intention-behavior relations: a conceptual and empirical review. *European Review of Social Psychology* 12 (1): 1-36.

Smith, J. R., W. R. Louis, D. J. Terry, K. Greenaway, M. R. Clarke, and X. Cheng. 2012. Congruent or conflicted? The impact of injunctive and descriptive norms on environmental intentions. *Journal of Environmental Psychology* 32 (4): 353-61.

Stern, P. C., T. Dietz, T. D. Abel, G. Guagnano, and L. Kalof. 1999. A value-belief-norm theory of support for social movements: the case of environmentalism. *Human Ecology Review* 6 (2): 81-97.

Sugden, R. 1989. Spontaneous order. *Journal of Economic perspectives* 3 (4): 85-97.

- Sunstein, C. R. 1996. Social norms and social roles. *Columbia Law Review* 96 (4): 903-68.
- Traxler, C., and M. Spichtig. 2011. Social norms and the indirect evolution of conditional cooperation. *Journal of Economics* 102 (3): 237-62.
- Tyler, T. R. 1990. *Why People Obey the Law*. Yale University Press.
- Tyler, T. R., and Y. J. Huo. 2002. *Trust in the Law: Encouraging Public Cooperation with the Police and Courts*. Russell Sage Foundation.
- Young, H. P. 1993. The evolution of conventions. *Econometrica* 61 (1): 57-84.
- Young, H. P. 1996. The economics of convention. *Journal of economic perspectives* 10 (2): 105-22.
- Young, H. P. 2015. The evolution of social norms. *Annual Review of Economics* 7: 359–87.

Appendix B

Paper II: Conditions and Effects of Norm Internalization

The following paper was published by the *Journal of Artificial Societies and Social Simulation*.

Batzke, M. C. L., & Ernst, A. (2023c). Conditions and Effects of Norm Internalization. *Journal of Artificial Societies and Social Simulation*, 26(1), 1–31.
<https://doi.org/10.18564/jasss.5003>

Conditions and Effects of Norm Internalization

Marlene C. L. Batzke¹ and Andreas Ernst¹

¹Center for Environmental Systems Research, University of Kassel Wilhelmshöher Allee 47, 34119 Kassel, Germany

Correspondence should be addressed to batzke@uni-kassel.de

Journal of Artificial Societies and Social Simulation 26(1) 6, 2023

Doi: 10.18564/jasss.5003 Url: <http://jasss.soc.surrey.ac.uk/26/1/6.html>

Received: 09-07-2021 Accepted: 03-01-2023 Published: 31-01-2023

Abstract: Norm internalization refers to the process of adoption of normative beliefs by individuals, thus representing a link between individual and social change. However, there are several questions regarding norm internalization which need to be answered. These include understanding under which circumstances norm internalization does occur by considering the effects of internalizing either a certain norm or even conflicting norms. To investigate the conditions and effects of norm internalization, we developed a theoretical agent-based model called “DINO”, comprising a norm internalization process grounded on a psychological model of decision-making, considering different types of norms, goals, and habits as well as inter-individual differences. Our conceptualization of personal norms introduces a new level of complexity, allowing for more than one norm to be internalized and either approved or disapproved. Our conceptual model was implemented within the framework of a 3-person Prisoner’s Dilemma game. Results showed that playing with cooperative others generally facilitated norm internalization. Norm internalization encouraged norm compliance and affected behavioural stability and payoff equality. We discuss how our results relate to empirical findings and theoretical literature, providing a bridge between theory development and empirically testable hypotheses and between psychological micro-level phenomena and social dynamics.

Keywords: Norms, Internalization, Learning, Social Dilemma, Cooperation, Decision-Making

● Introduction

- 1.1 Norm internalization is a key mechanism in norm compliance and norm maintenance (Axelrod 1986; Gintis 2004; Horne 2003). In the social sciences, there is a long tradition of theorizing about and studying norm internalization (Parsons 1937; Piaget 1970; Ullmann-Margalit 1977). Norm internalization is considered crucial for learning social values and norms, being an important link between society and the individual (Hoffman 2000; Kohlberg 1984; Vygotsky 1930). This makes norm internalization not only relevant for the individual but also for social and societal change. Internalization has been associated with selfless behaviour (Durkheim 1893), taming the egoistic individual through socialization (Freud 1932), while social influences may arguably lead to “the creation of a storm trooper, a Buddhist monk or a civil rights activist” (Kohlberg & Hersh 1977, p.53). Empirical studies have shown the importance of personal norms for behavioural decisions (e.g., Harland et al. 1999; Shin et al. 2018). However, there is little research about how internalization proceeds (Neumann 2010b). This is where social simulation comes into play (Jager & Ernst 2017).
- 1.2 In social simulation, few researchers have studied the norm internalization process (Andrighetto et al. 2010b; Villatoro et al. 2015). So far, our understanding remains “fragmentary and insufficient” (Conte et al. 2010, p.64). There is a lack of a psychologically plausible, dynamic theory of norm internalization (Hollander & Wu 2011; Neumann 2010b; Saam & Harrer 1999). Social simulation is a suitable means for developing and rigorously testing a dynamic theory. It enables exploring behavioural effects of internal changes as well as their interaction with social dynamics of change. As norm internalization is challenging to pin down methodologically

(Neumann 2010b), the additional use of simulation methods to empirical approaches seems especially promising. Modelling norm internalization demands for cognitively rather complex and heterogeneous agents, since internalization is considered a higher mental function that is individually specific (Piaget 1970; Vygotsky 2004).

- 1.3** Here, we combine psychological theory on norm internalization with agent-based simulation methods. This has several challenges: there is a lack of theories concerning dynamic processes; there are few theoretical superstructures that combine different psychological fields, and simulation demands for a level of precision that psychological theories rarely provide. Moreover, one of the greatest issues for modelling is psychological complexity. We approached these challenges first through relying on psychological concepts and research as well as through integrating assumptions from different psychological and adjacent disciplines. Second, we limited complexity to those areas that are essential to address our research questions. We now will discuss related research on the dynamics of norms, focusing in particular on models that include norm internalization, before presenting the contribution of this work to the field.

Related research

- 1.4** So far, two very different approaches to simulating norms in agent-based systems have been taken. In short, the first is mainly concerned with the emergence of behavioural conventions, treating norms solely as macro level epiphenomena (e.g., Axelrod 1986). In this line of research, social convention norms are often the only quality of norms considered (Mahmoud et al. 2012; savarimuthu et al. 2007; Sen & Airiau 2007). In the second line of research, norms are modelled as social constraints to agent's decision-making (Shoham & Tennenholtz 1992, 1995). This idea of built-in behavioural laws was advanced by emphasizing the cognitive representations of social norms and building intelligent, autonomous agent architectures (Castelfranchi et al. 2000; Conte & Castelfranchi 1995; Saam & Harrer 1999). In the Belief-Obligations-Intentions-Desires (BOID) architecture, social norms are represented as perceived obligations and personal goals as desires (Broersen et al. 2001). Through differentiating between individual and social desires, a new complexity was introduced (Neumann 2010b).
- 1.5** Regarding the importance of norm internalization (Axelrod 1986; Hollander & Wu 2011; Mahmoud et al. 2014; Neumann 2008, 2010b; Saam & Harrer 1999), surprisingly few agent-based systems include such a process. Verhagen (2001) presented a model of norm internalization, defining internalization as the matching of personal norms to social norms. Whereas this operationalizes the effect of norm internalization, it "does not represent the process of norm internalization" (Neumann 2008, Section 7.6). Andrighetto et al. (2010b) introduced a rich cognitive model of norm internalization, the EMIL-I-A architecture (see also Conte et al. 2010), in which, a norm may be prohibiting, prescribing, or permitting. EMIL Internalizer Agents internalize a norm depending on two conditions: a salient social norm and a cost-benefit-computation exceeding a threshold. Based on a dichotomous parameter, successful internalization stops the EMIL-I-A agent's normative deliberation and starts a decision-making automatism, disregarding other normative and non-normative motivators. The internalized norm has become a goal in itself. As long as the social norm is still salient, the agent complies with its internalized norm. This conceptualizes an internalized norm similar to a habit that saves time and calculation effort in decision-making and is in line with Epstein (2006) assumption of internalization being blind conformism with a norm. Villatoro et al. (2015) enhanced the EMIL-I-A architecture by characterizing norm internalization as a multi-step process, whereas only the deepest level of internalization corresponds to Epstein's assumption of thoughtless conformity. This allows agents to have partly internalized a norm and still violate it, while there is always just one norm internalized at any one time.

The present research

- 1.6** The current state of research leaves several questions open that existing models on norm internalization do not address. First, it seems important to investigate facilitating conditions of norm internalization to eventually be able to promote internalization of cooperation norms. Whereas norm internalization is assumed a universal process, when and how a norm is internalized depends on a person's personality (Ryan & Deci 2017; Vygotsky 2004) interacting with its environment, but how? Which conditions facilitate norm internalization? As working hypotheses, we assumed that individuals' inherent cooperativeness and the cooperativeness of the social surrounding positively influence internalizing the cooperation norm as appropriate.
- 1.7** Second, we are interested in the *effects* of norm internalization, assuming an imperfect relation between personal norms (the product of norm internalization) and behaviour. Whereas personal norms have been shown to influence behaviour, empirical investigations mostly found small to medium sized relations (e.g., Bamberg

et al. 2007; Bamberg & Schmidt 2003; Hines et al. 1987). Hence, personal norms do not seem to translate one-to-one into behaviour (Bandura 2001; Schönbach 1990; Schwartz 1977a). One possible explanation could be that multiple internalized norms influence behavioural decisions at the same time. Assuming that there are several, potentially conflicting personal norms at work, what are the effects of norm internalization? As working hypotheses, we first hoped to replicate the finding that norm internalization increases norm compliance (Axelrod 1986; Gintis 2004). We secondly expected that individuals become more determined and persistent through norm internalization (Andrighetto et al. 2010b) and therefore less flexible in their behaviour. Regarding the macro level social effects of norm internalization, there is little existing research. Since social norms have been proposed as solutions to social inequality (Saam & Harrer 1999; Ullmann-Margalit 1977), we thirdly assumed similar effects for norm internalization, decreasing inequality.

1.8 For the next step towards a better understanding of norm internalization, we aimed to disentangle conceptually distinct constructs and incorporate them separately. We did so by endorsing a psychological view that regards personal norms as conceptually distinct from habits. This enabled us to study norm internalization independently of other normative and non-normative factors driving decision-making. Based on this, we had two aims. First, we accounted for inter-individual differences that affect norm internalization to investigate individually specific facilitating conditions. Second, we allowed for more than one personal norm to be internalized and influence decision-making to examine the effects of norm internalization given multiple, potentially conflicting internalized norms. On this basis, we aimed to address the two following research questions and test the corresponding hypotheses:

- What are the mechanisms and conditions that facilitate norm internalization?
 - Cooperative agents tend to internalize that it is appropriate to cooperate; defective agents tend to internalize that it is appropriate to defect.
 - Cooperative settings facilitate internalizing the appropriateness of cooperation; defective settings facilitate internalizing the appropriateness of defection.
- What are the effects of norm internalization?
 - Stronger norm internalization is associated with stronger norm compliance.
 - Stronger norm internalization is associated with more behavioural inflexibility.
 - Stronger norm internalization is associated with less payoff inequality.

1.9 The paper is structured as follows: First, we present the theoretical framework, starting with our definitions of norms and proceeding with our conceptual model of norm-based decision-making. Subsequently, we describe the game-theoretical scenario the conceptual model was applied to. We then introduce the implemented agent-based DINO model – *Dynamics of Internalization and Dissemination of Norms*. Next, we document the simulation results addressing our research questions and hypotheses. Finally, we discuss the results and conclude.

● Theoretical Framework

Taxonomy of norms

2.1 We define a norm as a behavioural rule for a specific situation (Dannenberg et al. 2023). Social norms are shared norms between several individuals, and they can have different qualities (Cialdini et al. 1990). A social descriptive norm contains information regarding the observable regularity of a behaviour in a certain situation (i.e., “what most others do”) and is expected to affect behaviour through conformism. It “motivates by providing evidence as to what will likely be effective and adaptive action” (Cialdini et al. 1990, p. 1015). A social injunctive norm refers to the (in)appropriateness of a behaviour in a certain situation (i.e., “what most others consider (in)appropriate”; Dannenberg et al. 2023). An important mechanism explaining their influence is social (dis)approval (Ajzen 1991; Jacobson et al. 2011). Apart from social norms, personal norms describe the norms that an individual holds (Cialdini et al. 1990; Farrow et al. 2017). Personal norms are of an injunctive quality, defining an individual’s belief about (in)appropriate behaviour. They are associated with feelings of moral obligation as well as guilt and shame when violated (Schwartz 1977a; Schwartz & Howard 1981, 1982), which is why they are sometimes also referred to as “moral norms” (Bicchieri & Dimant 2019; Nyborg 2018; Thøgersen 1999). The process of how personal norms develop and change, we call norm internalization (Hoffman 2000; Kohlberg 1984).

2.2 Two more considerations are important for our research. First, we consider a norm of any type as *behaviour-specific*. For example, we assume that there is one personal norm relating to the (in)appropriateness of riding the bike in a specific situation and another personal norm relating to the (in)appropriateness of driving the car, whereas these two personal norms may develop independently from each other. Second, we consider each norm to vary along the dimension of encouragement to discouragement of a behaviour. This relates to prescriptive and proscriptive norms (Bendor & Swistak 2001; Bicchieri 2006; Ullmann-Margalit 1977). Hence, we suggest that, for instance, the personal norm of riding the bike can be approved of (representing a belief of appropriateness) or disapproved of (representing a belief of inappropriateness).

A conceptual model of norm-based decision-making

2.3 To investigate the dynamics of norms and their influence on behaviour, one needs theoretical assumptions on how behavioural decisions are made. Based on psychological literature, we developed a conceptual model of norm-based decision-making. Here, we first present relevant literature and then our conceptual model.

2.4 One of the most influential and best-validated theories in psychological literature is the *theory of reasoned action* (TRA hereafter, Fishbein & Ajzen 1975). The TRA was later expanded to the *theory of planned behaviour* (Ajzen 1991), explaining actions that are not under volitional control. In the present decision scenario that the conceptual model was applied to, we regard behavioural control as given and therefore applied the TRA. Both theories have been extremely successful in explaining behaviour (e.g., Sheeran 2002; Steg & Vlek 2009; Webb & Sheeran 2006). They are based on an *expectancy-value model* (Ajzen 1991; Atkinson 1957), consisting of expectancies and values. Expectancies are situationally adapted anticipations of behavioural outcomes. Values are a person's individual importance of motivational factors. A multi-attribute utility calculation of expectancy and value factors determines a person's *intention*. The intention is the only direct determinant of behaviour (Fishbein & Ajzen 1975, 1981). In the TRA, the authors considered two motivational factors influencing the intention: the *attitude* as a person's inner strivings and the *subjective norm* as outer, social influence.

2.5 The full range of a person's inner strivings evaluating behavioural outcomes in conflictual or mixed-motive situations is represented in Deutsch's *social-value orientations*: individualistic, cooperative, and competitive (cf. Deutsch 1958; Messick & McClintock 1968; Murphy & Ackermann 2014; Murphy et al. 2011). While there are numerous theories describing basic drivers of behaviour under the terms of goals (e.g., Lindenberg & Steg 2007) or needs (e.g., Maslow 1943), two terms often used similarly in social simulation (Kangur et al. 2017; Schlüter et al. 2017), above mentioned social-value orientations are particularly well researched in socially interdependent decision-making situations. The TRA's subjective norm corresponds to the above given definition of a social injunctive norm. Research has shown that social descriptive norms as well as habits influence decision-making over and above the TRA constructs (for social descriptive norms: Ravis & Sheeran 2003; White et al. 2009; for habits: Bamberg & Schmidt 2003; Perugini & Bagozzi 2001; Whitmarsh & O'Neill 2010). Investigating the additional influence of personal norms, some studies have shown (e.g., Conner & Armitage 1998; Harland et al. 1999; Shin et al. 2018) and some have failed to show their independent effects, raising the question whether their effects are separable from those of other behavioural influences (Bamberg & Schmidt 2003; Kaiser & Scheutle 2003). Theorists provided a possible explanation for the inconclusive empirical results, assuming norm internalization to be a higher-level process, qualitatively different from the lower-level processes of social norm imitation (Kohlberg & Hersh 1977; Piaget 1970; Vygotsky 1930).

2.6 Figure 1 depicts our conceptual model of norm-based decision-making, being an adaptation and extension of the TRA, including all factors presented above. Referring to the TRA's attitude, we assume *goals* to represent a person's basic strivings. Like Deutsch (1958), we consider three goals, driving decision-making: the individualistic, cooperative, and competitive goal. The *individualistic goal* is defined as pure self-interest, not caring about the benefit or loss of others. The *cooperative goal* describes a combined interest in the well-being of the self and others. The *competitive goal* depicts concern for the self, while improving the relative gain compared to others. Analogous to the TRA's subjective norm, we consider social influences in *social norms*. Along with the above presented taxonomy of norms, we differentiate social descriptive norm ("what most others regularly do") and social injunctive norms ("what most others consider (in)appropriate"). Furthermore, our decision-making model includes *habits*, being a person's usual behaviour. All the motivational factors (depicted on the left of Figure 1) are comprised of situational *expectations* and a personal *value* factor. We assume that *personal norms* affect decision-making on a higher level, influencing the importance of motivational factors. The *intention* is an action-specific multi-attribute utility calculation, determining the behavioural choice. The behaviour changes the situation.

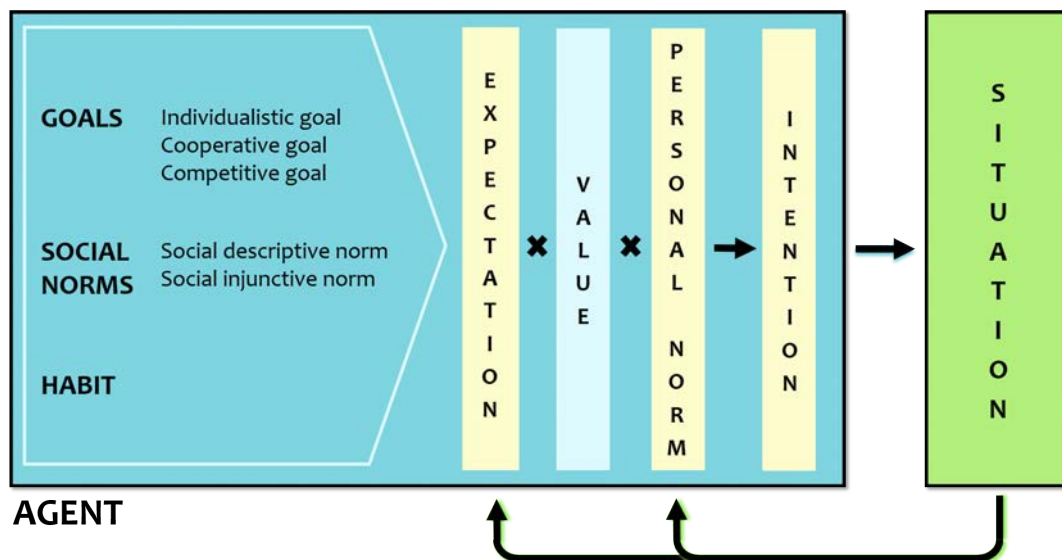


Figure 1: A conceptual model of norm-based decision-making. All motivational factors (on the left) are represented by expectation and value factors, which, shaped by the personal norm, lead to the intention. Based on the intention, an agent chooses an action that influences the situation. Learning processes triggered by a changed situation influence an agent's dynamic expectation and personal norm (indicated by yellow colour). Light blue colour indicates static factors.

- 2.7** Situational consequences serve as feedback to the agent for different learning processes. We assume that learning processes differ in their speeds. On the one hand, we expect situational expectations to change in a *fast* adaptation process, representing an individual's subjective perception of the situation. On the other hand, we expect the norm internalization process to be of a *slow* adaptation speed, storing and abstracting part of the situational learning. We assume that personal norm adaptations are the in-between of quick situational adaptations and rather stable personal values. Values represent the individual's importance of motivational factors manifested over a long time, accounting for personality differences, which we consider static for our purposes. In the following, we present the specifics on fast adaptation of expectations, slow adaptation of personal norms (i.e., norm internalization) and personality differences in values.

Fast adaptation of expectations

- 2.8** Expectancies are situationally adapted outcome anticipations of a specific behaviour (Bandura 2001). According to learning theory, expectations rise through repeated observations, being influenced by own experiences (i.e., *reinforcement learning*, Postman 1947; Sutton & Barto 2018) as well as experiences of others (i.e., *observational learning*, Bandura 1971, 1999). The fit of behaviour and outcome shows in stronger expectations similar to the logic of need satisfaction (Jager 2000; Kangur et al. 2017). Social descriptive norms are influenced by what most others do (Bendor & Swistak 2001; Cialdini et al. 1990; Hollander & Wu 2011). The perception of social injunctive norms is based on more complex, also societal dynamics (Andrighetto et al. 2010a; Dannenberg et al. 2023) and will not be further discussed in this paper. Habits are affected by a person's own behaviour (e.g., Ouellette & Wood 1998; Perugini & Bagozzi 2001) and have therefore, similar to social descriptive norms, also been conceptualized as personal descriptive norms (Batzke & Ernst 2022; Dannenberg et al. 2023).
- 2.9** Here, we assume that goal expectations are learned through reinforcement and observational learning. This implies that one can realize quickly whether a behaviour is useful for a specific goal. A match of goal and behaviour increases goal expectations. Social descriptive norm expectations are adapted through observing behaviour of others and habit expectations through perceiving own behaviour. We suggest that these observations are made rather quickly. Whereas social injunctive norms in a group or a society may change rapidly at times, we focus on the long periods of stability in between events of change. Therefore, we regard social injunctive norm expectations as static.

Slow adaptation: Norm internalization

- 2.10** We now address (1) how norm internalization proceeds and (2) how it affects decision-making, building on assumption formulated in Dannenberg et al. (2023). According to dissonance theoretical approaches, norm internalization is considered an internal reasoning process about one's past behaviour (Bem 1967; Festinger 1957; Rozin 1999). Reviewing psychological literature revealed some key factors affecting the normative evaluation of appropriateness or inappropriateness of a chosen action. First, the process is influenced by a person's perception of (in)appropriate behaviours in the social world (e.g., Bandura 2001; Kohlberg 1964), being represented within their social injunctive norms. Second, observations of other people's behaviours influence internalization (e.g., Miller & Dollard 1941; Sherif & Sherif 1953), pointing to the importance of people's social descriptive norms. Third, a person's habitual choices affect their perception of appropriateness (cf. *self-perception theory*, Bem 1967, 1972). Fourth and last, the internalization process is influenced by a person's goals, serving as feedback to the individual about how it performs in the environment, regarding its personal preferences (cf. *self-determination theory*, Deci & Ryan 1985; Ryan & Deci 2017).
- 2.11** Here, all presented factors influence norm internalization, namely: both social norms, habits, and goals, with internalization depending on their situational expectations and personal values. This makes norm internalization a more abstract process, representing an individual's learning effect, abstracting and storing part of the situational learning. We therefore assume it to be of a slow speed. What is enough support for a behavioural decision to approve of the respective behavioural norm? We suggest that if there is more support for a behavioural decision (rather than against it), the respective personal norm is approved of (i.e., internalized as a belief of appropriateness). Otherwise, it is disapproved of (i.e., internalized as a belief of inappropriateness).
- 2.12** Regarding the effects of internalized norms on decision-making, *motivated reasoning* approaches predict that once an individual has acquired a certain belief, it searches for reasons that support it, trying to maintain an "illusion of objectivity" (Pyszczynski & Greenberg 1987, p. 302), see also Kunda (1990) and Markus & Kunda (1986). Hence, people are generally inclined to confirm a conclusion rather than to disconfirm it (cf. *confirmation bias*, Snyder 1984). *Dissonance theory* states that these justification processes reduce psychologically uncomfortable cognitive dissonance, which arises when a new normative belief is formed that conflicts with existing norms or goals (Festinger 1957; Voisin & Fointiat 2013). Highlighting arguments that favour the belief or trivializing those that oppose it, reduces dissonance (Simon et al. 1995).
- 2.13** Here, we assume that personal norms affect decision-making by influencing the importance of motivational factors, namely social norms, goals, and habits. We suggest that a personal norm of approval highlights the importance of other norm-consistent motivational factors. Conversely, a personal norm of disapproval trivializes norm-dissonant motivational factors, making norm internalization a motivational source for dissonance reduction phenomena.

Personality differences in values

- 2.14** Whereas norm internalization is a universal process, it is assumed to be influenced by a person's personality (Ryan & Deci 2017; Vygotsky 2004). Personality differences arise through personal values of motivational factors (Ajzen 1991). Research suggests that personality variables are rather stable, fundamental patterns of people (Costa & McCrae 1986; Harris et al. 2016; Terracciano et al. 2010). Although they may change over long periods of time, we consider personal values as static, since our research focus lies on norm internalization and personality's influence on it rather than personality change. To reduce possible combinations of different personal values for the model, we developed seven psychologically plausible personality types. The types are ordered along their cooperativeness, ranging from strong cooperators (types 1 and 2) to strong defectors (types 6 and 7), with more internally conflicted and socially sensible conditional cooperators in between (types 3 to 5). Relevant psychological literature for developing these types and their description is presented in the Appendix.

● The Scenario: A Social Dilemma Game

- 3.1** Generally, we assume that norm internalization occurs in every situation. However, many societal challenges arise from people acting selfishly in interdependent and mixed-motive situations, resulting in collective costs. Numerous authors have addressed the issues of how and when individuals act pro-socially and cooperation collectively emerges (e.g., Axelrod 1986; Nowak et al. 2004). One of the most promising factors for long-term behavioural change is motivational change (Otto & Kaiser 2014). Therefore, we were particularly interested

in investigating the role of norm internalization in those situations, in which a selfish action has to be refrained from in favour of a collectively advantageous action. What makes decisions in these situations so difficult is that they are made against the backdrop of a perceived conflict, putting individual's and collective's interests at odds and interdependent of each other (Dawes 1980). The most basic game-theoretical model that reflects this arguably very common social conflict is the prisoner's dilemma game (PDG, Luce & Raiffa 1957). Thus, each player has only two behavioural options: cooperation and defection. The conflict is described within the properties that a person receives a higher payoff for a socially defecting choice; however, all individuals in the society are better off, if all cooperate rather than defect (Axelrod 1984).

- 3.2** As previously mentioned, we here focused on agent dynamics. We aimed to minimize dynamics based on interactions between scenario and agents by limiting complexity in the decision scenario. This allowed us to attribute resulting dynamics to agents. Therefore, we chose one of the simplest game-theoretical scenarios for a first implementation of our conceptual model: an iterated 3-person PDG. In the 3-person game, we are already studying a group, which employs a different logic than the 2-person game (Dawes 1980). In future model extensions, the number of agents can easily be increased, and the same logic still applies. In the game, each time step an agent $i \in I = \{1, 2, 3\}$ chooses between two behavioural actions $a_i \in \{0, 1\}$: cooperation $a_i = 1$, representing the pro-social choice, and defection $a_i = 0$, representing the egoistic choice. Cooperation is beneficial to every individual. Hence, everyone receives benefits b when individual i cooperates. However, cooperation comes with a cost c for the cooperating individual. The individual's payoff (P_i) is a function of its action and the actions of the others given by:

$$P_i = 1 + (a_1 + a_2 + a_3)b - a_1c \text{ with } 3b > c > b > 0 \quad (1)$$

This features the typical characteristics of a PDG. We used a payoff matrix with $b = 1$ and $c = 2$, depicted in Table 1.

Number of cooperators	Payoff to defectors	Payoff to cooperators	Collective payoff
3	-	2	6
2	3	1	5
1	2	0	4
0	1	-	3

Table 1: Payoff matrix of the 3-person prisoner's dilemma game.

● The Agent-Based DINO Model

- 4.1** In the following, the agent-based DINO model, *Dynamics of Internalization and Dissemination of Norms*, is presented. The model was programmed in NetLogo (Wilensky 1999) version 6.2.02.

Agent's decision-making

- 4.2** Figure 2 illustrates agents' decision-making. Motivational factors are depicted in the rows. Each motivational factor is represented by an expectation and a value factor. Agents' dynamic expectations express the situational fit of a motivational factor and a specific action. An exception is the social injunctive norm expectation, which we consider as static. Agents' static values ascribe importance to a motivational factor. Values are set depending on agent type. Personal norms are dynamically internalized, influencing the importance of motivational factors. Agents use a weighted multi-attribute subjective utility matrix to calculate the intention to show an action as described in function (2). Agents perform the action with the highest intention. The model is purely deterministic.

$$\begin{aligned}
 I_{i,t} = & IND_{i,t} \times v_{IND,i} \times PN_{i,t} + COOP_{i,t} \times v_{COOP,i} \times PN_{i,t} + \\
 & COMP_{i,t} \times v_{COMP,i} \times PN_{i,t} + SDN_{i,t} \times v_{SDN,i} \times PN_{i,t} + \\
 & SIN_{i,t} \times v_{SIN,i} \times PN_{i,t} + HA_{i,t} \times v_{HA,i} \times PN_{i,t}
 \end{aligned} \quad (2)$$

In the following, we show the specifics on (1) agents' perception, knowledge, and memory, (2) the adaptation of expectations, (3) the adaptation of personal norms and their influence on decision-making, and (4) the implementation of agent heterogeneity in values through agent types. At the end of this chapter, model parametrization, initialization, and execution are described.

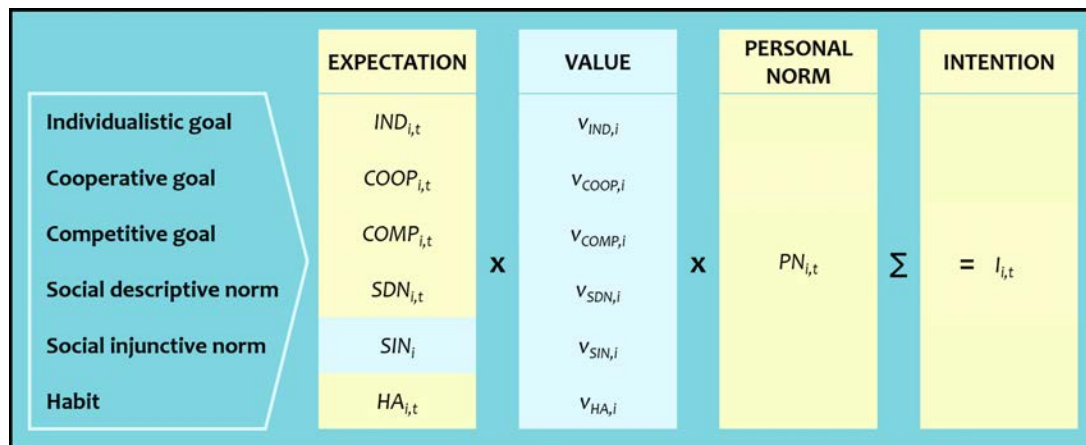


Figure 2: Agents' decision-making in the DINO model. Yellow colour illustrates dynamic parameters, changing over time (additionally indicated by t). Light blue colour indicates static parameters. All parameters are specific for an agent i . Explanations are provided in the text.

Perception, knowledge, and memory

- 4.3** As can be expected in the 3-person PDG, we assumed that agents have knowledge about the payoff matrix and thus know the minimum and maximum achievable individual and collective payoffs. Moreover, agents can observe the other agents' actions and remember their own and the others' payoffs from the previous round.

Adaptation of expectations

- 4.4** We now present the adaptation processes underlying expectations. Expectations are action-specific: one motivational factor is represented in two expectations, one for each action, that is, cooperation or defection.

Goal expectations

- 4.5** Goal expectations determine how promising an agent considers an action in order to achieve a specific goal. Agents evaluate their goal expectation in two successive processes: reinforcement and observational learning. In reinforcement learning, only the goal expectations of the action an agent has taken before are positively or negatively reinforced. After each action an agent evaluates whether it was useful for goal achievement given the circumstances. For that, an agent uses its knowledge about the game. Goal expectations are adapted through subtracting or adding a relative amount of the *expectation-change-rate* (see section on parametrization). The three goals are evaluated according to the following rules.
- An agent considers the individualistic goal as achieved, if either the individual payoff peaked ($P_{i,t} = [\text{maximum individual payoff}]$) or improved compared to the last round ($P_{i,t} > P_{i,t-1}$). If either of the two conditions apply, the individualistic goal expectation of the last action increases; otherwise, it decreases. The degree of strengthening or weakening is relative to the goal (non-)achievement: it depends on the achieved amount of the individual payoff in relation to the maximum achievable individual payoff (i.e., change in individualistic goal expectation = $P_{i,t} / [\text{maximum individual payoff}] \times \text{expectation-change-rate}$).
 - An agent considers the cooperative goal as met, if either the collective payoff peaked ($P_{c,t} = [\text{maximum collective payoff}]$) or improved compared to the last round ($P_{c,t} > P_{c,t-1}$). Again, the degree of strengthening or weakening depends on the achieved collective payoff in relation to the maximum achievable collective payoff (i.e., change in cooperative goal expectation = $P_{c,t} / [\text{maximum collective payoff}] \times \text{expectation-change-rate}$).
 - The competitive goal answers to the question whether the agent's individual payoff is higher than the individual payoff of at least one other agent j ($[P_{i,t} > P_{j,t}] \geq 1$). In case the condition is met, the relative increase of the competitive goal expectation is defined as the number of outperformed others (i.e.,

increase in competitive goal expectation = $[P_{i,t} > P_{j,t}] / [\text{total number of other agents}] \times \text{expectation} - \text{change} - \text{rate}$). In case the condition is not met, the relative decrease depends on the number of others, having outperformed the agent (i.e., decrease in competitive goal expectation = $[P_{i,t} \leq P_{j,t}] / [\text{total number of other agents}] \times \text{expectation} - \text{change} - \text{rate}$).

- 4.6** Observational learning is based on observing the other agents' actions and adapting goal expectations according to their behavioural consequences. Analogous to the rules of reinforcement learning, an agent evaluates the other agents' actions and their consequences and adapts its goal expectations accordingly. Compared to reinforcement learning, observational learning is of a minor importance, with adaptations being one fifth as large (Ernst 2003).

Social descriptive norm expectations

- 4.7** *Social descriptive norm expectations* convey information regarding how unequivocally a behaviour matches a social descriptive norm. Due to having only two behavioural alternatives in the PDG, we assumed that in this specific scenario the two expectations of social descriptive norms are negatively associated. Thus, if an agent observes one behaviour, it also notices the absence of the other. Hence, agents update both expectations at each time step¹. Whereas this seems redundant in the PDG, the DINO model was designed to also fit more complex interdependence structures, where norm information regarding one behaviour cannot necessarily be derived from other behaviours. Expectations were adapted according to the following condition:

- If the majority of agents (≥ 2) cooperated, the cooperative social descriptive norm expectation increases by the *expectation - change - rate* and the defective counterpart decreases accordingly (vice versa, if the majority of agents defected).

Habit expectations

- 4.8** *Habit expectations* express the fit of an action with a behavioural habit. Analogous to social descriptive norms, the two habit expectations are interdependent and adapted according to the following condition:
- If the agent cooperated, the cooperative habit expectation increases by the *expectation - change - rate* and the defective counterpart decreases accordingly (vice versa, if the agent defected).

Adaptation of personal norms and their influence on decision-making

- 4.9** As presented in our norm taxonomy, we assumed that agents have two personal norms, one for each action. Norm internalization is an agent's normative judgement regarding the (in)appropriateness of the chosen action and the following approval or disapproval of the according personal norm. To make the judgement, an agent takes its three goals, descriptive and injunctive social norms, and habits into account. All these motivational factors are collected in agents' intentions. Therefore, agents evaluate the strength of their intention of the last action through dividing it by the maximum intention that could be achieved regarding their personal values. The maximum achievable intention equals the sum of agents' values. The multiplication of values by expectations is unnecessary since their maximum is 1. If the reasons in favour of a behaviour outweigh the ones against, the personal norm corresponding to the last action is approved of and increases by the *internalization - change - rate* (see section on parametrization). Otherwise, it is disapproved of and decreases as described in the following adaptation rule:

$$\begin{aligned} \text{if } [I_{i,t} / (v_{IND,i} + v_{COOP,i} + v_{COMP,I} + v_{SDN,i} + v_{SIN,i} + v_{HA,i})] > 0.5 \\ \text{then } PN_{i,t} + \text{internalization} - \text{change} - \text{rate} \\ \text{otherwise } PN_{i,t} - \text{internalization} - \text{change} - \text{rate} \end{aligned} \quad (3)$$

- 4.10** Personal norms are a multiplier for each motivational factor in the decision-making calculation. That way, personal norms reinforce norm-consistent and inhibit norm-inconsistent motivational factors. More precisely, agents check which action a motivational factor supports best, through comparing the two action-specific expectations of a motivational factor. The personal norm, which supports the same action, is multiplied by value and expectation of a motivational factor. A personal norm of approval ($PN_{i,t} > 1$) strengthens a motivational factor that favours a behaviour. Conversely, a personal norm of disapproval ($PN_{i,t} < 1$) weakens a motivational

factor that favours a behaviour. Taking the individualistic goal as an example, personal norms affect the influence of all motivational factors in decision-making according to the following rule:

$$\begin{aligned}
 & \text{if } IND_{C,i,t} > IND_{D,i,t} \\
 & \text{then } IND_{i,t} \times v_{IND,i} \times PN_{C,i,t} \\
 & \text{otherwise } IND_{i,t} \times v_{IND,i} \times PN_{D,i,t}
 \end{aligned} \tag{4}$$

Agent heterogeneity in values: Seven agent types

4.11 Table 2 illustrates the implementation of the seven agent types, which is based on (1) the presented literature (see the Appendix), (2) the authors' psychological expertise, and (3), whenever (1) and (2) did not give a clear decision, the matching of the resulting agent's behaviour with the personality type. As presented in the conceptual model, DINO agents possess six motivational factors, shown in the rows of Table 2. Agent types differed in their values, shown in the cells, defining the importance of motivational factors.

	Agent Types						
	Cooperators		Conditional cooperators			Defectors	
	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
individualistic goal	0	1	1	2	2	3	1
cooperative goal	3	3	2	2	1	0	0
competitive goal	0	0	1	2	2	1	3
social descriptive norm	0	2	2	2	2	1	0
social injunctive norm	1	3	2	1	3	1	0
habit	3	2	0	0	0	1	3

Table 2: Implementation of seven agent types by differing values of motivational factors. Agent types (depicted in the columns) are defined by the values they ascribe to the single motivational factors (depicted in the rows). Values may ascribe high (3), medium (2), low (1) or no (0) importance to a motivational factor.

Parametrization, initialization, and model execution

- 4.12** All expectations varied between [0-1]. Higher values indicate a better fit of a motivational factor and a specific action. They were initialized to a neutral midpoint of 0.5. Agents' personal values ranged between [0-3] and were initialized depending on agent type (see Table 2). Personal norms ranged from [0-2] and were initialized with 1. Social injunctive norm expectations were static, being set to slightly support cooperative behaviour with 0.6 for the cooperative and 0.4 for the defective social injunctive norm expectation.
- 4.13** As presented in our conceptual model of norm-based decision-making, the different adaptation processes differed in their speeds. First, we assumed that goal expectations, social descriptive norm expectations and habit expectations are adapted in a fast adaptation process. The according *expectation – change – rate* was set to 0.2, making expectations adaptable from one extreme to the other within about five rounds of the game. Second, we assumed that personal norms are adapted in a slow adaptation process. We conducted a sensitivity analysis for the *internalization – change – rate*, keeping the *expectation – change – rate* stable at 0.2 (see Appendix). The model showed relatively stable results for change rates between 0.01 and 0.07. We set the *internalization – change – rate* to 0.02, allowing personal norms to change from full norm approval to full norm disapproval within 100 rounds of the game.
- 4.14** In each time step, agents first calculate their intentions for each action and then act on the strongest one. Agents observe the actions of others and receive their payoff. Next, agents adapt their goal expectations through first reinforcement learning and second observational learning. Then, social descriptive norm and habit expectations are adapted. Lastly, agents adapt their personal norms through norm internalization and update their memory concerning payoffs. Model execution was terminated after 200 time steps, as this is a period in which lock-in phenomena have occurred and agents' internal dynamics have stabilized in most model runs. The first five time steps were excluded from the analyses.

Experiments and Results

- 5.1** We now address our two research questions. First, we investigated which conditions facilitate norm internalization, analysing which agents internalize what norm under which circumstances. Second, we examined the effects of norm internalization by manipulating agents' personal norms. In the present work, we studied norm internalization as an additional influence in decision-making over and above the ones of other normative and non-normative behavioural drivers. As a proof-of-concept simulation, we additionally tested its independent effects by comparing the model *with* norm internalization to the model *without*. Results are presented and discussed in Appendix.
- 5.2** Personal norms to cooperate and to defect are hereafter called C-PN and D-PN, both of which agents internalize (i.e., approve of or disapprove of). A model run relates to a specific group composition of three agents playing the 3-person PDG. Since the model is purely deterministic, no repetitions of the same model runs were conducted. Generally, we analysed 84 different group compositions, because the seven agent types can form 84 different groups of three.

Which conditions facilitate norm internalization?

- 5.3** Figure 3 illustrates agent dynamics of norm internalization, depending on agent type, type of norm (C-PN or D-PN), and group composition. Norm internalization is shown as the absolute value of the personal norm of one agent at a time step (0-200), ranging from norm approval (in turquoise), across indifference (in white) to norm disapproval (in purple). Group compositions are shown as three-digit numbers, indicating the three agent types in the group (e.g., group composition "123" = agent types 1, 2, and 3), ordered along group cooperativeness.

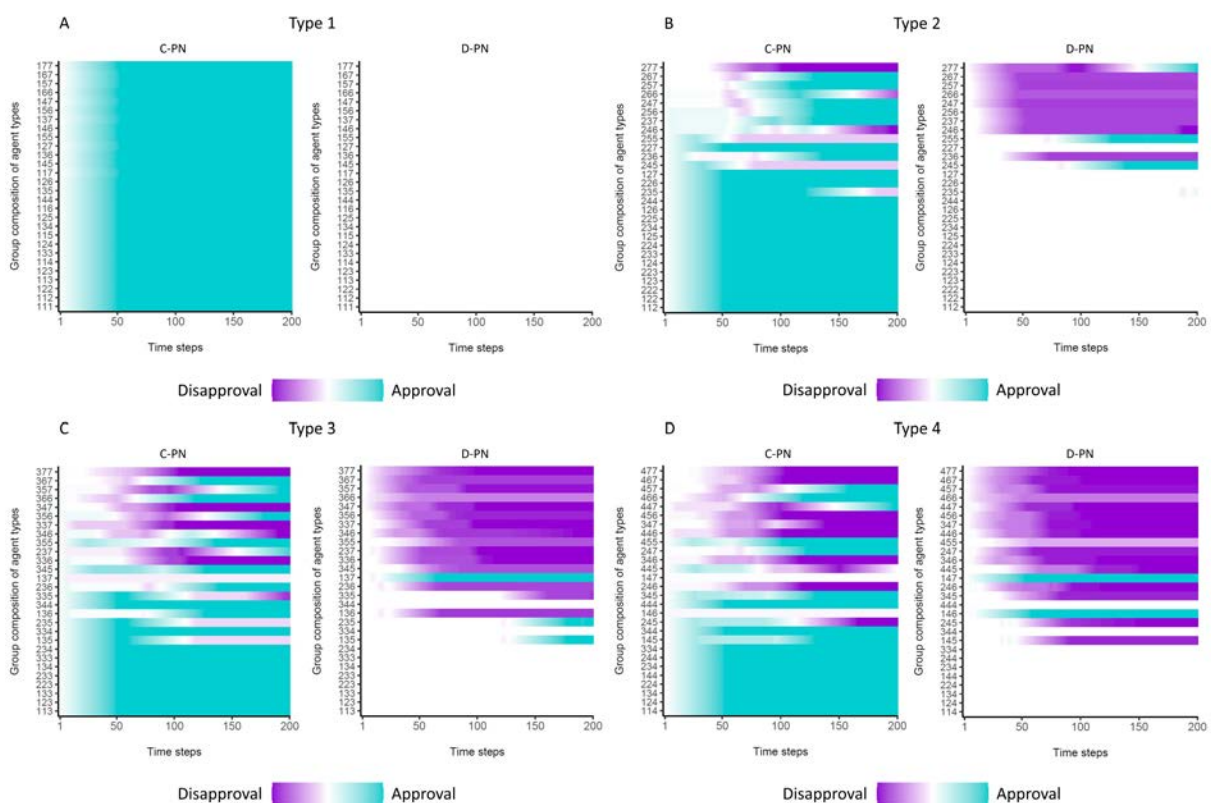


Figure 3: Agent types' internalization of the personal norm to cooperate (C-PN, figures on the left) and the personal norm to defect (D-PN, figures on the right) over 200 time steps, depending on the group composition of agents. Norm internalization ranges from disapproval (in purple) to approval (in turquoise) of a personal norm, while white colour indicates indifference towards a norm. Group compositions are shown as three-digit numbers, indicating the three agent types in the group. Groups are ordered along group cooperativeness (i.e., digit sum and largest single digit) in ascending order (from bottom to top).

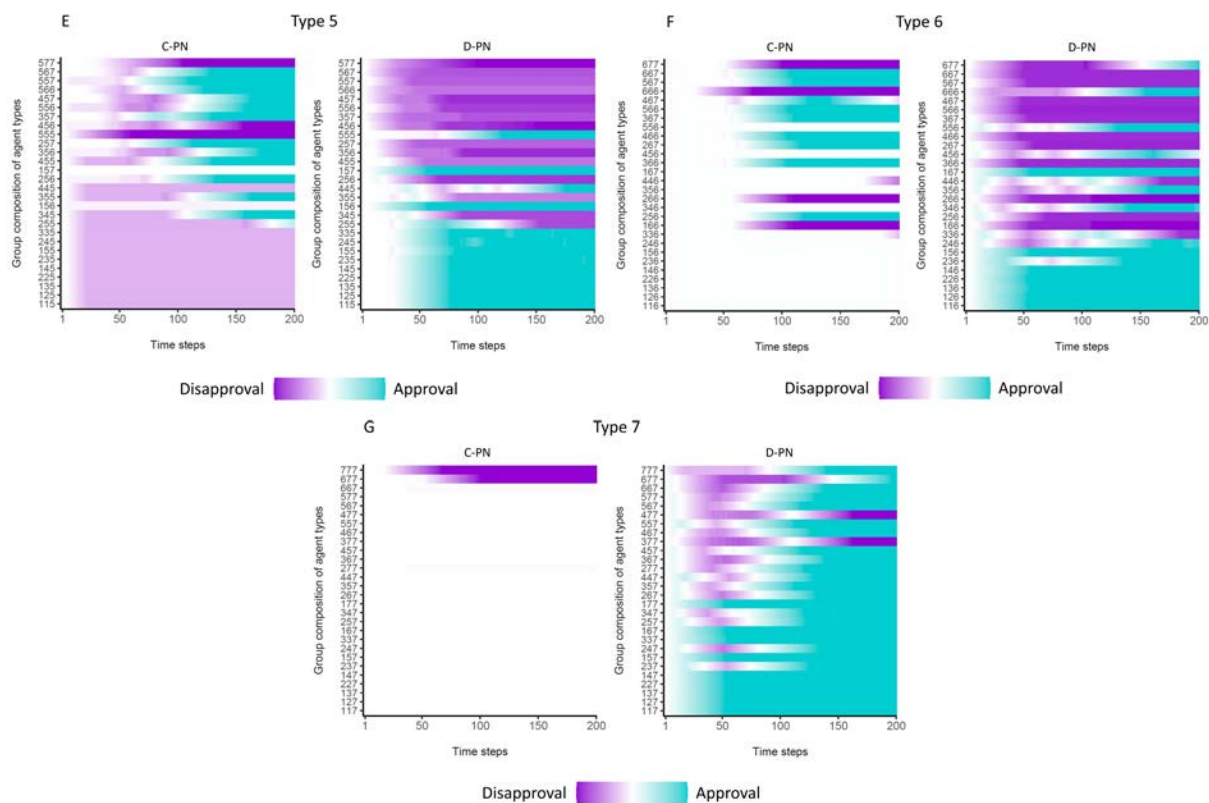


Figure 4: Agent types' internalization of the personal norm to cooperate (C-PN, figures on the left) and the personal norm to defect (D-PN, figures on the right) over 200 time steps, depending on the group composition of agents. Norm internalization ranges from disapproval (in purple) to approval (in turquoise) of a personal norm, while white colour indicates indifference towards a norm. Group compositions are shown as three-digit numbers, indicating the three agent types in the group. Groups are ordered along group cooperativeness (i.e., digit sum and largest single digit) in ascending order (from bottom to top).

- 5.4** Generally, all agent types might at least develop one personal norm of approval. Highly intrinsically driven agents (i.e., types 1 and 7) approved of the norm according to the behaviour they inherently prefer under (almost) all circumstances (see Figures 3A and 4G). The other agent types (i.e., types 2 – 6) might temporarily approve and disapprove of both norms, depending on the group composition (see Figures 3B-D and 4E-F). Agent types 2 – 4 approved of the C-PN in the majority of group compositions (see Figures 3B-D), types 5 and 6 approved of the C-PN or D-PN in roughly equal parts (see Figures 4E-F), and agent type 7 mostly approved of the D-PN (see Figure 4G). This supports Hypothesis 1a that cooperative agents tend to approve of the C-PN and defective agents tend to approve of the D-PN.
- 5.5** For norm approval, a cooperative setting was generally beneficial. It allowed rather cooperative agents to achieve their cooperative goal when cooperating. Interestingly, it also allowed defecting agents to be better than others (satisfying their competitive goal) or to accumulate payoff (individualistic goal) through defection. Achieving goals and complying with norms leads DINO agents to approve of a norm. More defective settings made the norm internalization process longer, in which one or both personal norm(s) were approved or disapproved of before a stable state was reached. This especially applied to conditional cooperators.
- 5.6** For example, in highly heterogeneous settings consisting of a cooperator and a defector, conditional cooperators tended to (partly) disapprove of a one norm before approving of the other (see Figures 3C-D and 4E). This often ended in D-PN approval. While this phenomenon was especially typical to conditional cooperators, it was observable in many agent types (2 – 7). For conditional cooperator types 3 and 4, playing with another (similar or more cooperative) conditional cooperator and a defector resulted in disapproval of both norms (see Figures 3C-D). For types 5 and 6, however, norm internalization in these settings resulted in C-PN approval (see Figures 4E-F). Too cooperative settings led agent types 5 and 6 to approve of the D-PN and types 3 and 4 of the C-PN. Highly defective settings led all conditional cooperators to disapprove of both C-PN and D-PN. Hence, Hypothesis 1b was only partly supported. While cooperative settings did facilitate approval of the C-PN, they facilitated approval of any norm, and defective settings did not facilitate approval of the D-PN.

What are the effects of norm internalization?

- 5.7** To examine the effects of norm internalization, we conducted two series of experiments. In both, we varied the degree to which agents have internalized personal norms. First, we manipulated one norm, keeping the other constant. This shows the effects of having internalized a certain norm depending on the group composition. Second, we varied both norms independently from each other, representing aggregated results across group compositions, showing the effects of internalizing multiple, potentially conflicting norms. To address our hypotheses, we looked at three different outcome variables: cooperation, behavioural changes, and payoff inequality.
- 5.8** Figure 5 shows the effects of agents having internalized a certain norm, the C-PN, to varying degrees, depending on group composition of agent types. Group compositions are again ordered along group cooperativeness, defined by the cooperativeness of the single agents. In the beginning of a model run, we once manipulated agents' personal norms within their boundary values ranging from full disapproval to full approval. In between full approval and disapproval manipulations, manipulation strengths decrease towards the framed "Baseline", showing the standard model runs in which no manipulations were conducted. Outcomes were aggregated across agents and time.

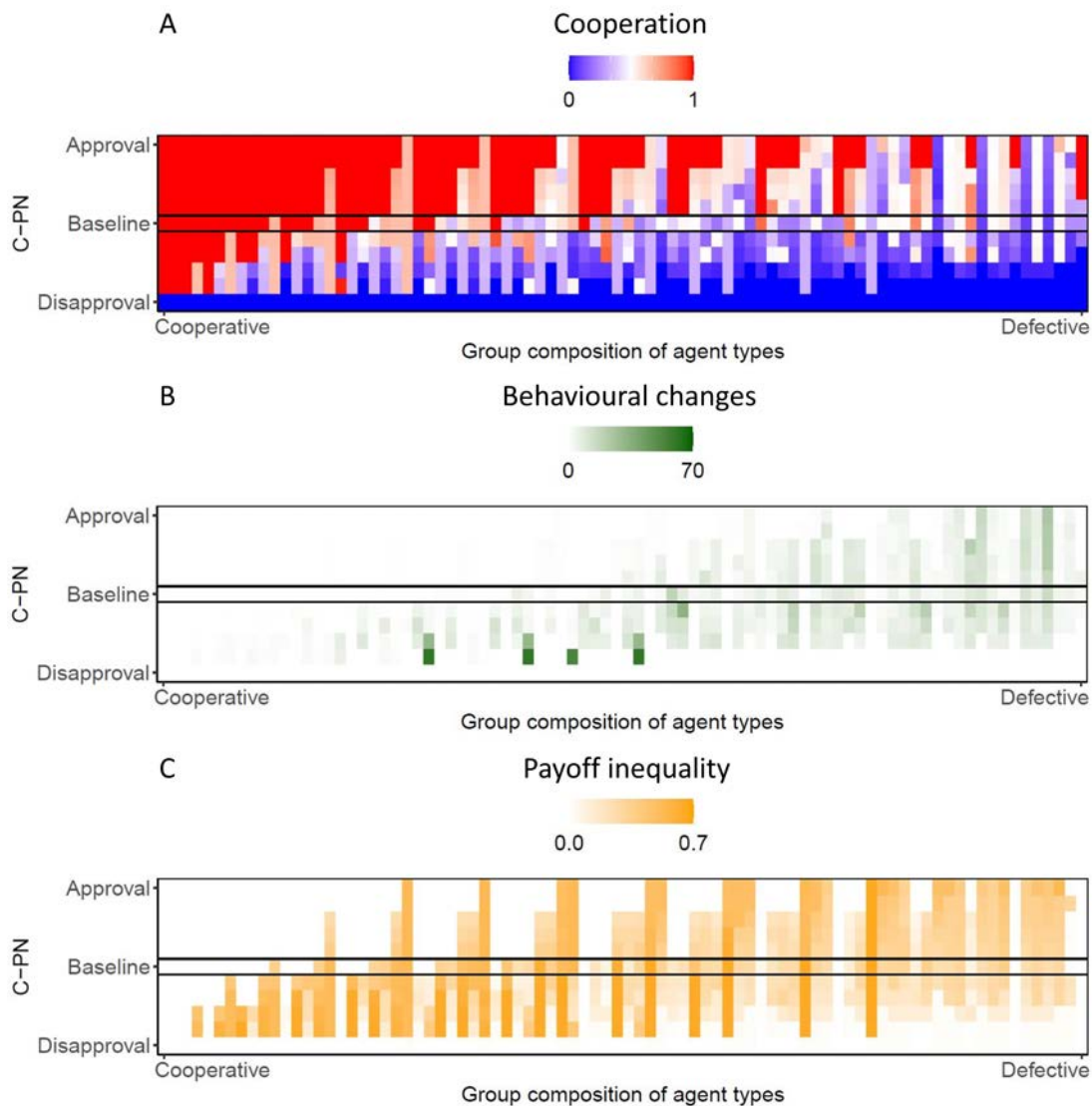


Figure 5: Manipulation of the personal norm to cooperate (C-PN) depending on group composition of agent types. The C-PN was varied from full disapproval to full approval. No manipulation was conducted in baseline model runs. Left to right shows agent group compositions, ordered along cooperativeness. Group compositions are defined by three-digit numbers, indicating the three agent types in the group, ordered by digit sum and largest single digit. Effects are shown regarding (A) cooperation (ranging between $[0,1]$, averaged across agents and time), (B) absolute number of behavioural changes, and (C) inequality between agents' individual payoffs (ranging between $[0,1]$, averaged across agents and time). Duration of model runs: 200 time steps.

5.9 Figure 5A shows that agents' cooperation increased with stronger approval of the C-PN and decreased with its disapproval compared to the baseline model runs, supporting Hypothesis 2a. In rather cooperative group compositions, slight manipulations towards approval of the C-PN led to pure cooperation of all agents. Strong manipulations achieved the same result even in more defective group compositions. Although approval of the C-PN increased average cooperation in all group compositions, some groups of agents were not tipped over to pure cooperation. Conversely, full disapproval of the C-PN led to pure defection in all group compositions, even those consisting of solely cooperators. Figure 5B indicates that high numbers of behavioural changes were generally associated with defective group compositions as well as (partial) disapproval of the C-PN. A similar pattern was found in Figure 5C regarding payoff inequality, whereas inequality increased especially with C-PN disapproval in mixed, heterogeneous groups. In case of full disapproval of the C-PN, behavioural changes disappeared, and equality was established due to pure defection of all agents, supporting Hypotheses 2b and 2c. The effects of full norm approval partly supported the hypotheses as well, but effects were less univocal.

5.10 To investigate the effects of having internalized both, potentially conflicting personal norms, we varied C-PN and D-PN independently. Figure 6 shows the effects of full disapproval to full approval of both norms regard-

ing cooperation, behavioural changes, and payoff inequality. Again, disapproval and approval manipulations decrease towards “Baseline” model runs without manipulation. Results were aggregated across agents, time, and group compositions.

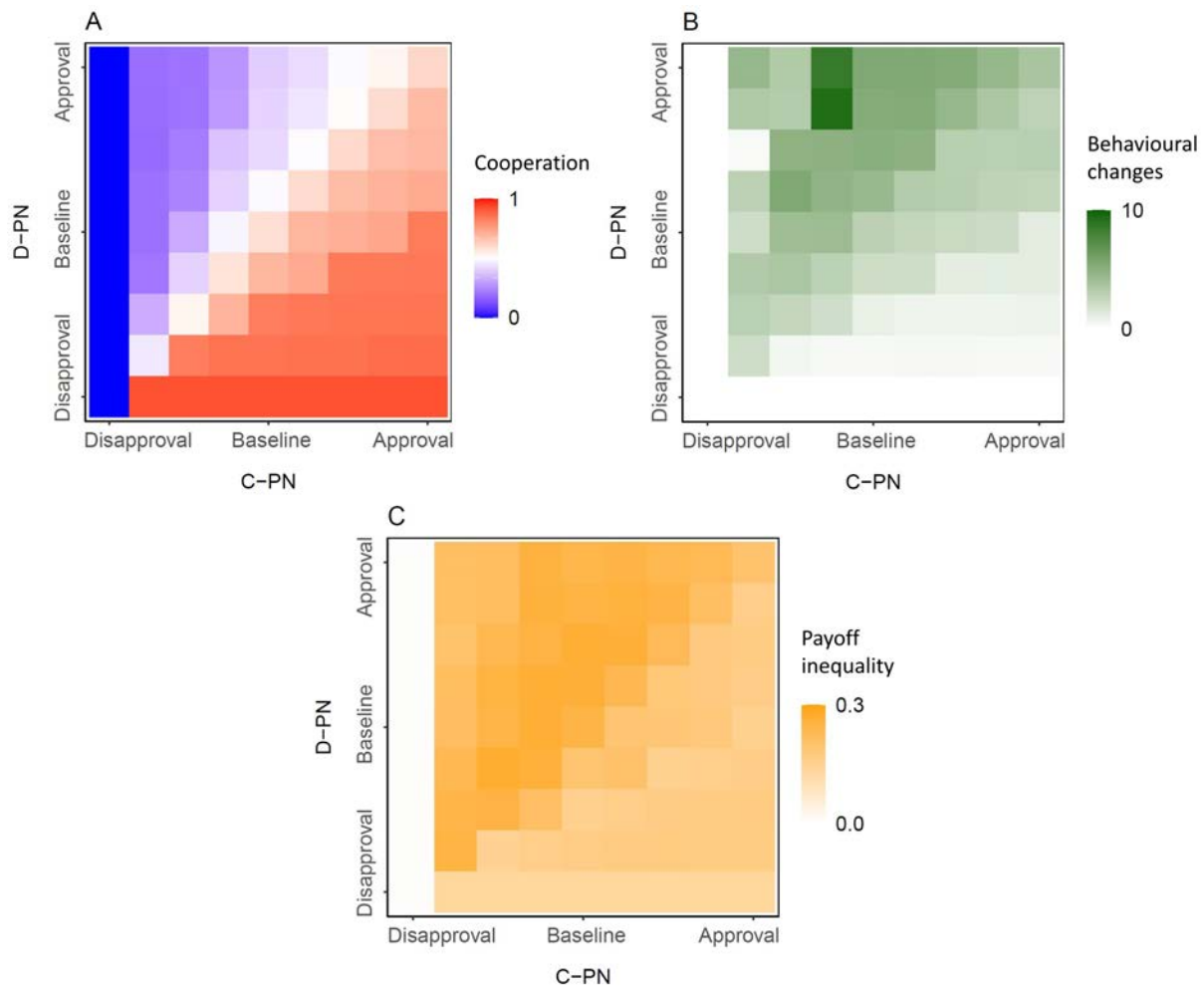


Figure 6: Manipulation of personal norms to cooperate (C-PN) and to defect (D-PN). Personal norms were varied from full disapproval to full approval. Effects are shown regarding (A) cooperation (ranging between [0,1], averaged across agents, time, and group compositions), (B) absolute number of behavioural changes (averaged across group compositions), and (C) inequality between agents' individual payoffs (ranging between [0,1], averaged across agents, time, and group compositions). Results were averaged across 84 group compositions of agent types. Duration of model runs: 200 time steps

5.11 Figure 6A shows that norm approval and disapproval were generally behaviourally effective, which again supported Hypothesis 2a. However, it also shows that a personal norm of disapproval had even stronger behavioural effects than a norm of approval. Full disapproval of the C-PN led to pure defection. As a result, behavioural changes and payoff inequality disappeared (see Figures 6B and 6C), supporting Hypotheses 2b and 2c. Full disapproval of the D-PN led to cooperation in large parts, however, not pure cooperation. The diagonal from the lower left corner (full disapproval of both personal norms) to the upper right corner (full approval of both personal norms) indicates that equally strong norm disapproval was associated with defection. Equally strong norm approval did not affect cooperation, behavioural changes (see Figure 6B), and payoff inequality (see Figure 6C). Dominance of the C-PN over the D-PN tended to decrease behavioural changes and payoff inequality, which supports Hypotheses 2b and 2c. Contrary to the hypotheses, dominance of the D-PN tended to increase behavioural changes and payoff inequality.

● Discussion

6.1 The aim of our research was to provide the next step towards understanding the dynamics of norm internalization better. Based on psychological literature, we developed a conceptual model of norm-based decision-making, made assumptions regarding the mechanisms of norm internalization, and implemented both into the agent-based DINO model. In the conceptual model, personal norms, being the product of norm internalization, are regarded as an additional motivational force in decision-making over and above the situational expectations and personal values of goals, habits, and social norms. Thus, the model accounts for inter-individual differences in norm internalization. The conceptualization of personal norms introduces a new level of complexity, being behaviour-specific rules. This implies the existence of a personal norm for each behavioural alternative. In the DINO model, they may develop independently from each other, while each personal norm may be approved of, representing a normative belief of appropriateness, or disapproved of, representing a normative belief of inappropriateness. In the present work, we tested the model regarding facilitating conditions and independent effects of norm internalization.

Conditions of norm internalization

6.2 A cooperative setting was generally beneficial to develop a personal norm of approval, which partly supports Hypothesis 1b. Cooperative settings allowed cooperators and conditional cooperators to approve of the norm to cooperate (as predicted). However, they also allowed defectors to approve of the defection norm. In highly defective settings, agents tended to disapprove of both norms. Hence, for norm approval most agents were dependent on the cooperation of others. In line with Hypothesis 1a, cooperators mostly approved of the cooperation norm and defectors of the defection norm. These highly intrinsically driven agents mostly approved of the norm they inherently preferred even in defective settings.

6.3 Conditional cooperators were more flexible regarding which norm they internalized. Empirical studies showed that conditional cooperators are highly influenced by the social setting (e.g., Burlando & Guala 2005; Kurzban & Houser 2005), increasing cooperation in cooperative settings (Fischbacher & Gächter 2010) and decreasing it in the presence of defectors (de Oliveira et al. 2015). This strongly relates to the DINO agent types 3 and 4. They approved of the cooperation norm in cooperative settings, which was associated with cooperation, and disapproved of both norms in more defective settings, associated with defection (see the Appendix). The model also replicated the empirical finding of conditional cooperators siding with the defector in highly heterogeneous settings (Hartig et al. 2015; Lucas et al. 2014), showing that this behaviour goes along with approval of the defection norm.

6.4 Relating to more individualistic and competitive DINO agent types 5 and 6, Fischbacher and colleagues (2001) found a behavioural pattern in 14% of participants characterized by conditional cooperation and a decay of cooperation above a certain contribution level of other players; a pattern they called “hump-shaped”. In the DINO model, these types approved of the defection norm in more cooperative settings. However, when playing with a defector and a conditional cooperator (rather than a pure cooperator), they eventually started cooperating and internalizing accordingly. It prevented them from following through on their individualistic and competitive goals, making it difficult to exploit each other. Similarly, Gächter & Thöni (2005) reported that their less cooperative subjects cooperated more when playing with similar others. They concluded that when “there are no cooperators around to free ride on [...] they understand that they need to cooperate among themselves if they want to earn money.” (pp. 310–311). The DINO model showed that defective conditional cooperators’ approval of the cooperation norm is facilitated by other conditional cooperators when defectors are present. On the contrary, this facilitated disapproval of both norms in more cooperative conditional cooperators (types 3 and 4). These types were dependent on a cooperative setting to approve of the cooperation norm.

6.5 Interestingly, conditional cooperators as well as type 6 generally approved of the cooperation norm in many group compositions. Regarding, for instance, the balanced goal structure of type 4 or the strong individualistic goal of type 6, this result was rather surprising. It suggests that approval of the cooperation norm is achievable for many types with different goal structures. This relates to empirical evidence that cooperative behaviour, such as pro-environmental behaviour, may result from different motivations including environmental and economic concerns (Brandon & Lewis 1999; Thøgersen 2003). Moreover, experimental research showed that even people that are predominantly motivated by gain rarely act completely egoistically (Camerer 2003).

6.6 The DINO model also suggests that the relative importance of personal norms matters. In ambiguous settings and generally in conditional cooperators, agents’ motivations were often too ambiguous for developing any norm of approval. As norm internalization was modelled as a motivational source for dissonance reduction

phenomena, disapproval of one norm reduced agents' internal conflicts and consequently could facilitate approving of another norm. This finding relates to Lindenberg & Steg (2007) goal-framing theory, wherein the authors argued for the importance of the relation between multiple goals. Since in the DINO model personal norms influence goal importance, one may assume that not only the relative importance of goals but also of personal norms is of significance.

Effects of norm internalization

- 6.7** In the DINO model, agents' behaviour was generally associated with approval of the according norm, supporting the idea formulated in Hypothesis 2a that norm internalization increases norm compliance (Andrighetto et al. 2010b; Axelrod 1986; Deci & Ryan 2000; Gintis 2004). The model also showed that norm approval is only behaviourally effective in those agents that are not overly disinclined towards the behaviour. For agents that are highly intrinsically driven, norm *disapproval* was more effective in achieving behavioural change. Hence, the motivational strengthening via norm approval of an inherently undesirable action did not exceed the motivational support that an inherently desirable action had in these agents. For example, to make defectors cooperate, disapproving of the defection norm was more effective than approving of the cooperation norm. This is an interesting point, relating to *prospect theory* (Tversky & Kahneman 1992). Therein the authors suggest that people give more weight to avoiding something undesirable than to achieving something desirable. Similarly, recent socio-political developments suggest the power of norm disapproval, such as the 2017 emerging Swedish "flight shame" movement to reduce flying drastically lowered the number of aircraft passengers in Sweden.
- 6.8** However, norm-based intervention studies predominantly focus on norm approval, often not testing the effects of norm disapproval (e.g., Hamann et al. 2015; Terrier & Marfaing 2015). As has been proposed before (Schwartz & Fleishman 1982), our results suggest that the effects of norm disapproval are potentially underestimated and might be worth further investigating. In case of conflict between the two personal norms, the model showed that disapproval of both tends to be associated with defection. This seems plausible with defection being the dominant strategy in the PDG (Dawes 1980) Equally strong norms of approval did not affect overall cooperation, supporting again to the idea that the relative importance of personal norms matters for behavioural decisions (Lindenberg & Steg 2007).
- 6.9** The DINO model presented mixed results for Hypotheses 2b and 2c, showing that norm internalization may increase or decrease behavioural inflexibility and payoff equality. The hypothesized effects of norm internalization increasing inflexibility and payoff equality held true for collective approval of the cooperation norm and disapproval of the defection norm. Hence, the model supports the idea that internalization has the potential to make agents more persistent (Andrighetto et al. 2010b). Moreover, it illustrated that persistence is limited to the behavioural option that is collectively beneficial. In these cases, norm internalization resulted in agents developing strong habits, making the two difficult to differentiate on a behavioural level, which relates to Epstein (2006) assumed connection of internalization and habit formation.
- 6.10** However, approval of the defection norm increased behavioural flexibility and payoff inequality. As Bendor & Swistak (2001) have famously demonstrated, pure defection is an unstable state. The DINO model suggests that the instability of defection is related to collective approval of the defection norm or disapproval of the cooperation norm. Norm disapproval may have counterintuitive effects on agents, making them more flexible and adaptive (see the Appendix), which shows their dissimilarity from habits. Very strong norm disapproval again effectively promoted behavioural consensus and thus stability and equality. Similarly, research has shown that social norms may have diverse and contradictory effects (see the Appendix), also regarding social inequality (Conte & Castelfranchi 1995; Saam & Harrer 1999; Ullmann-Margalit 1977). The DINO model showed that equality can be fostered through collective internalization of the same norm except for approval of the defection norm.

Limitations and future research

- 6.11** This work has a number of limitations, offering various leverage points for future research. The DINO model was implemented within the framework of an iterated 3-person PDG. Real-world situations are often characterized by more than two behavioural alternatives, larger and dynamic groups, changing payoff matrices, outcome uncertainty, et cetera. The focus of our work was to advance the cognitive representation of internalized norms in agent-based models. We considered it important to account for multiple personal norms, effects of their

approval and disapproval as well as personality differences, in order to better understand norm internalization processes.

- 6.12** For an initial implementation of these aspects, we tried to limit complexity in the situational framework. Our choice fell on the PDG, describing the core of a conflict that is common to many situations (Dawes 1980) and that we consider particularly interesting for studying norm internalization. Compared to the 2-person game, the 3-person game entails a different logic (Dawes 1980), allowing to easily increase the agent number later-on. In the 3-person game, emerging norms are specific for the small group context. In larger groups, one would expect that subgroups may emerge and develop their own social norms – a process that is not represented so far.
- 6.13** In developing our conceptual model, we aimed for principles and structures that also fit a more complex game with more behavioural alternatives, such as a *commons dilemma* (Ernst 2010; Hardin 1968; Ostrom et al. 2002). Applying the conceptual model to different behavioural domains may require different/additional goals as Jager (2000) similarly described for needs. It remains for future research to investigate norm internalization in larger, dynamic groups, test the effects of multiple group memberships, and take outcome uncertainty into account. Moreover, effects of situational influences through norm interventions, changes in the payoff matrix, et cetera, are promising approaches for influencing norm internalization and should further be investigated.
- 6.14** Whereas the developed conceptual model of norm-based decision-making is an extension of the frequently used theory of reasoned action, it still simplifies in various aspects. For instance, it does not consider all psychological factors that have been proven to significantly influence people's behaviour, such as values in the sense of higher-order beliefs (Schwartz 1977b). Psychological constructs sometimes lack a clear definition and differentiation with concepts even being used interchangeably. To improve (dynamic) theories of decision-making, further research is needed investigating the distinct mechanisms and independent effects of motivational factors – something social simulation can greatly contribute to (Jager 2017; Jager & Ernst 2017).
- 6.15** Furthermore, the model so far considers only one process in decision-making, being a subjective utility-maximizing decision calculation in the sense of deliberation. Scholars have argued for the existence of two distinct processes (cf. *dual-process theories*, Kahneman 2003; Petty & Cacioppo 1986) with the second being a fast and intuitive decision process. Implementing the conceptual model into a more complex game with more situational information will facilitate including heuristic shortcuts that agents use when specific situational cues appear (Gigerenzer 2001). Additionally, we simplified by considering social injunctive norms and personal values as static. To better understand the implications of norm internalization on the dynamics of social norms, it is an important step to endogenize social injunctive norms. Including personal value change would enable the study of norm internalization in the larger context of personality development.
- 6.16** There is a long tradition in studying normative agent-based systems, focusing on various aspects that are relevant to norms (Neumann 2010a). Whereas the DINO model does by far not represent all important mechanisms, missing for example social enforcement (Andrighetto et al. 2010b), we aimed at advancing the study of norms in agent-based systems on different levels. First, our representation of norms as (1) behaviour-specific and (2) ranging from approval to disapproval introduces a new level of complexity. The DINO model suggests that these aspects matter. Second, by implementing the norm internalization process in a generic decision-making context, it is applicable to a variety of decision-making models – also those not primarily focusing on norms. Prerequisite for applying the DINO internalization process is an expectancy-value model of decision-making, such as approaches based on the theory of planned behaviour which include models of voting behaviour (Kotona & Pahl-Wostl 2004), recycling behaviour (Scalco et al. 2017), resource use (Briegel et al. 2012; Nerb et al. 1997), and innovation diffusion (Schwarz & Ernst 2009). The theory of planned behaviour offers an integrative decision framework that can easily be expanded, and micro theories applied (Jager 2000).
- 6.17** The DINO model is a theoretical model, developed on the grounds of theoretical and empirical research. Some model results are well relatable to existing literature. Some assumptions, such as on the adaptation speed of different norms, remain free parameters and thus to be empirically tested and validated. Presently, there is very little empirical time-series data on internalization processes, while first shining examples show that they can be assessed (Szekely et al. 2021). In any case, there is much value in a formalized, coherent, and dynamic psychological theory (Jager 2017; Neumann 2010b). Simulated theories mean they can be experimented on (Dowling 1999; Troitzsch 2017). They allow us to investigate the relations between psychological phenomena on the micro level, such as norm internalization, and social/sociological ones on the macro level (Lorenz et al. 2021). They allow for the induction of hypotheses that may then be tested empirically and – in a recursive process – for improvement of theory, simulation, and data. The DINO model presents various hypotheses that may stimulate future empirical research. For instance, it suggests that norm approval is facilitated by a cooperative setting. This could be a starting point for investigating conditions of norm internalization empirically.

The model also implies that fostering norm disapproval can be a powerful means to achieve norm-consistent behaviour.

● Conclusion

- 7.1** Norm internalization is considered a fundamental principle in human development (Vygotsky 1930) and an essential element of socialization (Hoffman 1977). A better understanding of norm internalization seems crucial (Neumann 2010b). In this work, we presented an agent architecture for studying the conditions and independent effects of norm internalization within a decision-making framework. The DINO model demonstrated several main theoretical assumptions and empirical findings. It complements existing research by providing insights into the interactions of internal and external processes that underlie conditions of cooperation and highlights the potential of social simulation to provide causal explanations for empirically observed phenomena. Furthermore, it provides useful conceptualizations for advancing normative agent-based systems and promising theoretical conclusions that may serve as hypotheses for future research.

● Acknowledgements

The present work was developed within the *ZumWert* project. We acknowledge the funding of the University of Kassel in their profile building initiative. We thank three anonymous reviewers for their valuable and detailed comments. We also thank Georg von Wangenheim and Fabian Mankat for insightful discussions on model development.

● Model Documentation

The model is implemented in NetLogo version 6.2.2. The code and ODD protocol are available at: <https://www.comses.net/codebases/1bb193d4-6c9e-4f19-84d1-b95e2780e9ed/releases/1.0.0/>.

● Appendix

A1: Personality differences in values

First, we now briefly present relevant literature to developing the seven personality types and second introduce them. People's goal structure, relating to their social value orientation, translates into their *willingness to cooperate* (Murphy & Ackermann 2013, 2014; Murphy et al. 2011). Several authors categorized people along that continuum. Frank (1988) suggested two types: cooperative individuals, who always decide to maximize joint payoff, and defecting individuals, who strive to maximize their own payoff. Fischbacher et al. (2001) added a third type: conditional cooperators (see also Fehr & Fischbacher 2002). They have been described as a melting pot for all kinds of motivations such as "sucker aversion", 'conformity' or 'miming' [...] (Burlando & Guala 2005, p.36). Their motivation "depends directly on how others behave or are believed to behave" (Fischbacher & Gächter 2010, p.542), being social descriptive norms. Being prone to social influence is connected to the personality trait openness (McCrae 2000). Openness is linked to flexibility (Deyoung et al. 2002), relating to a low importance of relying on habits.

We adopted the classification along the willingness to cooperate continuum with people at the extremes inherently favouring either cooperation or defection. Conditional cooperators in between the extremes are inherently more conflicted and sensible to their social environment. Based on the literature, we assumed that highly cooperative and highly defective individuals score lower on openness than conditional cooperators. Hence, they are less affected by situational changes, such as social influences, and rely more on habits in decision-making. To represent more complexity than the conventional three categories of cooperators, conditional cooperators, and defectors, we distinguished seven personality types along their motivations. While the number of seven is in principle arbitrary, it allowed us a more fine-grained differentiation within each category similar

to a Likert scale. Moreover, each type possesses stereotypical characteristics relevant in the context of norm internalization. Of course, types do not represent the empirical reality of personalities, which one could consider as being continuous rather than categorical. However, for our purpose types were expedient, reducing complexity.

We now introduce our seven types, ordered from most to the least cooperative. Cooperator type 1 is intrinsically highly willing to cooperate and thus to sacrifice the own well-being for the group. It is strongly guided by its goals, inflexible, and barely influenced by social norms. Cooperator type 2 is similarly high motivated to contribute to the benefits of the group. However, type 2 also has some self-interest in mind and derives its motivation from its goals as well as from social norms. Empirically, cooperators represent between 1% and 18% of the population (1-4% in Noosey et al. 2020; 18% in Burlando & Guala 2005).

The three conditional cooperators (types 3 to 5) are more conflicted, having contradictory motivations. Type 3 is still highly willing to cooperate. However, it has equally strong other goals and is highly susceptible to social pressures. This makes type 3 inconsistent and malleable by its social surroundings. Type 4 is highly driven by what others do. Having a rather balanced goal structure, it is highly flexible and can jump on any bandwagon. Type 5 is worried about getting one's share from the cake (through individualistic and competitive goals), which is why cooperation seems too risky. At the same time, it is highly susceptible to social norms. Empirically, conditional cooperators are the largest fraction, constituting between 35% and 63% of the population (35% in Burlando & Guala 2005; 63% in Kurzban & Houser 2005).

Defector type 6 is a typical individualist, considering above all its own advantage, generally paying little attention to social norms. Defector type 7 represents the exact opposite to type 1 along the willingness to cooperate continuum, which means being highly competitive. Similar to type 1, type 7 is highly motivated by its goals, and little influenced by social pressures, resulting in a highly inflexible defector. Empirically, defectors represent between 4% and 33% of the population (33% in Fischbacher et al. 2001; 4-25% in Martinsson et al. 2009).

A2: Sensitivity analysis of the *internalization – change – rate*

We conducted a sensitivity analysis for the adaptation parameter of personal norms, namely the *internalization – change – rate*. We assumed that the adaptation speed of the personal norms was lower than the one of expectations. We set the *expectation – change – rate* to 0.2, assuming that expectations are adapted quickly within few rounds of the game, and varied the *internalization – change – rate* from 0.01 to 0.2. Figure 7 shows the effects on (A) cooperation, (B) behavioural changes, (C) payoff inequality, and (D & E) internalization of both personal norms. Figure 7A shows that agents' cooperation is more similar in several group compositions for *internalization – change – rates* between 0.01 and 0.07. Several groups were tipped over to pure cooperation at higher change rates. Lower change rates were associated with more behavioural changes (Figure 7B) and slightly higher payoff inequality (Figure 7C). Figures 7D and 7E show that internalization dynamics tended to be more similar with low change rates.

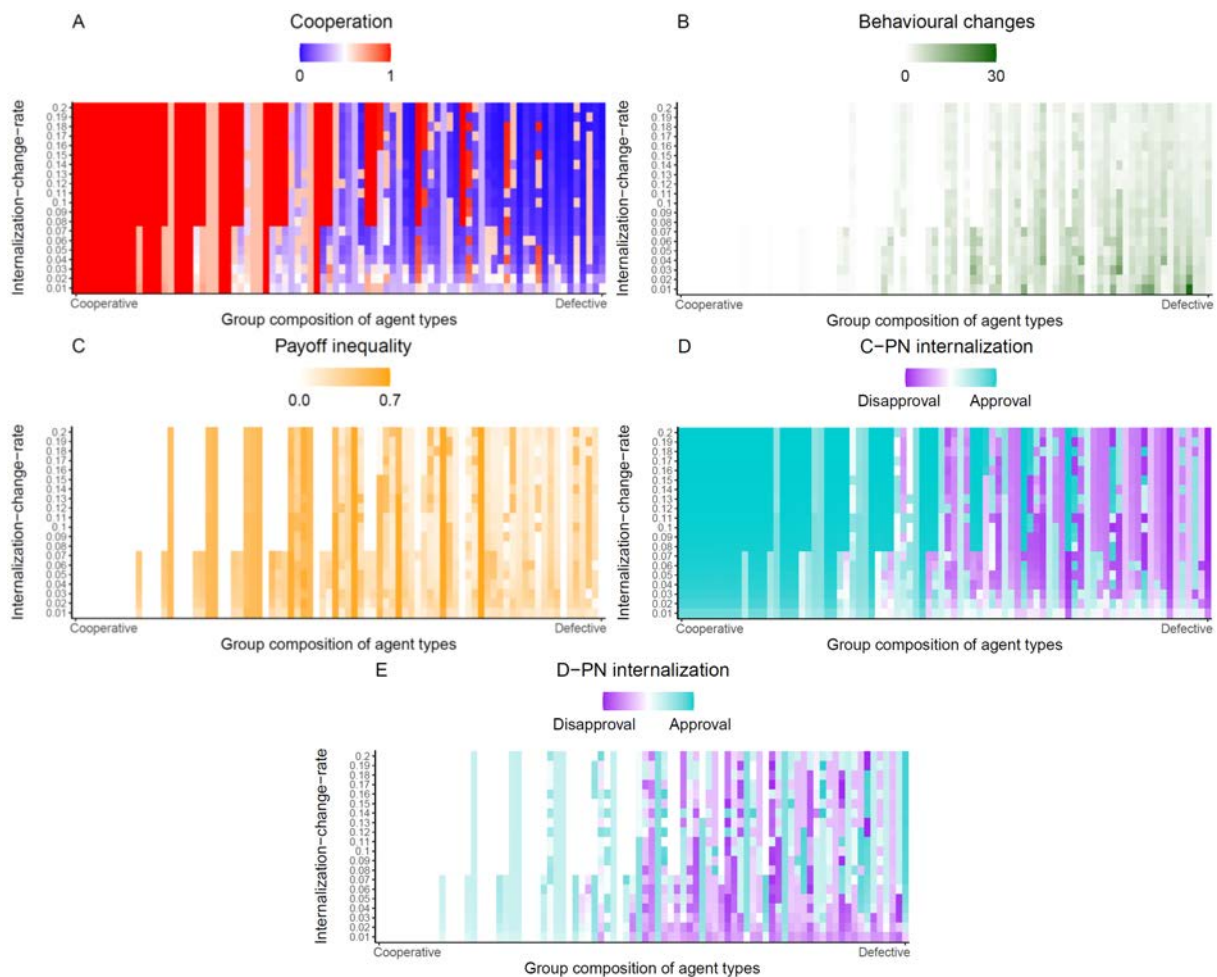


Figure 7: Variation of the adaptation speed of personal norms (i.e., *internalization – change – rate*) depending on group composition of agent types. Left to right shows agent group compositions, ordered along group cooperativeness. Group compositions are defined by three-digit numbers, indicating the three agent types in the group, ordered by digit sum and largest single digit. Effects are shown regarding (A) cooperation (ranging between [0,1], averaged across agents and time), (B) absolute number of behavioural changes, (C) inequality between agents' individual payoffs (ranging between [0,1], averaged across agents and time), and (D E) internalization of the personal norm to cooperate (C-PN) and the personal norm to defect (D-PN; both ranging from disapproval to approval, averaged across agents and time). Duration of model runs: 200 time steps.

A3: Proof-of-concept simulation: What are the independent effects of norm internalization apart from other normative and non-normative behavioural influences?

Results

To examine the independent effects, we compared the model *with* norm internalization to the one *without*. As guidance for this exploratory analysis, we used the same outcome variables as in the other experiments, namely: cooperation, behavioural changes, and payoff inequality. Outcome variables were averaged across time and group compositions. Table 3 shows agents' cooperation in the model with and without norm internalization, depending on agent type. Most agent types cooperated more in the model with norm internalization, which also caused a general increase in cooperation (see Table 4). Did norm internalization lead to norm-consistent behaviour? Whereas approval of the C-PN was strongly associated with cooperation ($p_{pb} = 0.93$), the D-PN was not correlated with defection ($p_{pb} = -0.05$), resulting from the fact that defection could result from D-PN approval as well as disapproval of both norms (see Figure 8)². Similarly, correlational data showed that approval of the C-PN was associated to greater differences in the two behavioural intentions favouring cooperation ($p_{pb} = 0.81$), whereas approval of the D-PN was close to uncorrelated with differences in intentions favouring defection ($p_{pb} = -0.11$).

Cooperation

	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
Without norm internalization	0.98	0.67	0.56	0.52	0.53	0.00	0.00
With norm internalization	0.98	0.84	0.72	0.65	0.40	0.28	0.02

Table 3: Agent types' cooperation in the model without and with norm internalization. Cooperation ranges between [0,1], being averaged across 84 group compositions of agent types and 200 time steps per model run.

	Without norm internalization	With norm internalization
Cooperation	0.51	0.59
Behavioural changes	0.12	2.93
Round of an agent's last behavioural change	1.20	54.63
Payoff inequality	0.13	0.25

Table 4: Differences between the model without and with norm internalization. Models were compared regarding cooperation (ranging between [0,1], averaged across agents, time, and group compositions), number of behavioural changes (averaged across group compositions), round of an agent's last behavioural change (averaged across group compositions), and inequality between agents' individual payoffs (ranging between [0,1], averaged across agents, time, and group compositions). Results were averaged across 84 group compositions of agent types. Duration of model runs: 200 time steps.

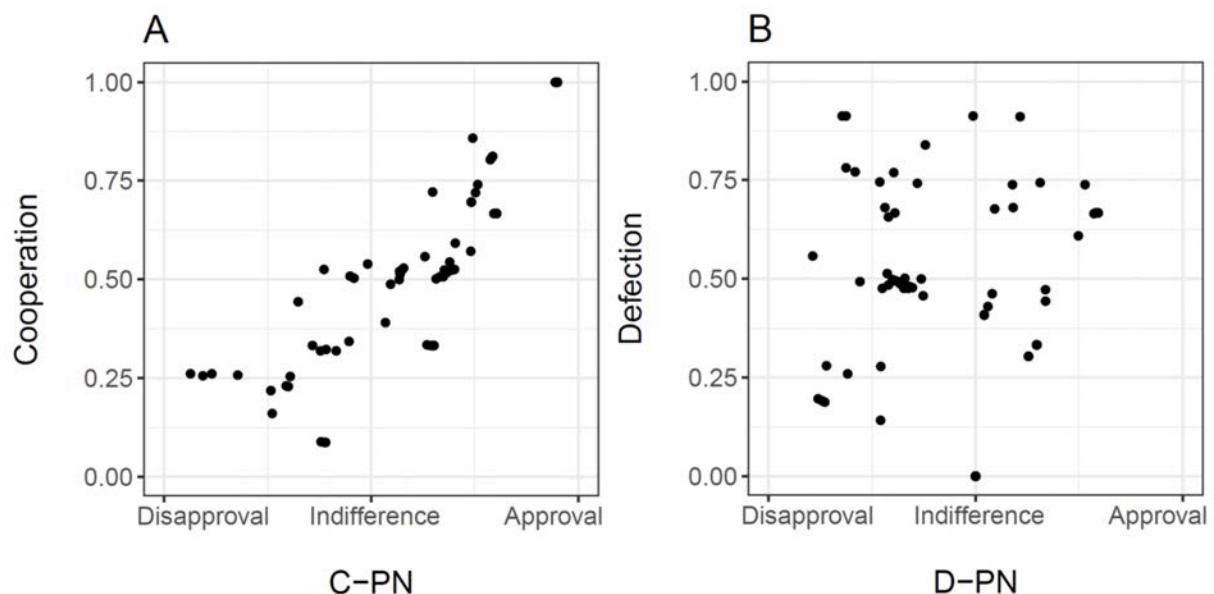


Figure 8: Association of cooperation and the personal norm to cooperate (C-PN; Figure A) as well as defection and the personal norm to defect (D-PN; Figure B). Each data point represents one agent group composition (out of 84), averaged across agents and time. Duration of model runs: 200 time steps.

Table 4 shows that agents changed behaviour more often when they internalized norms. Moreover, in the model with norm internalization behaviour stabilized later and lock-in phenomena occurred later in the game. Without norm internalization, model runs were strongly determined by agents' first few actions and their consequences, while norm internalization made agents' behaviour more volatile. Norm internalization could be a longer process depending on agent type and social setting. Especially those model runs in which agents' norm internalization was a longer process differed more strongly between the model without and with norm internalization. Norm internalization was associated with behavioural volatility. Correlations showed that particularly disapproval of a norm was associated with more behavioural changes ($p_{pb,C-PN} = -0.61$ and $p_{pb,D-PN} = -0.68$) and later occurrence of lock-in phenomena ($p_{pb,C-PN} = -0.83$ and $p_{pb,D-PN} = -0.56$).

Payoff inequality emerged when agents behaved differently from one another. Norm internalization changed payoff allocations between agents (see Table 4). Although norm internalization is influenced by social norms, personal norms do not encourage behavioural conformity and hence payoff equality, but rather increased inequality. Payoff inequality was slightly associated with approval of the D-PN ($p_{pb} = 0.25$) and disapproval of the C-PN ($p_{pb} = -0.37$), whereas the latter relation is better described by an inverted U-shape (see Figure 9A).

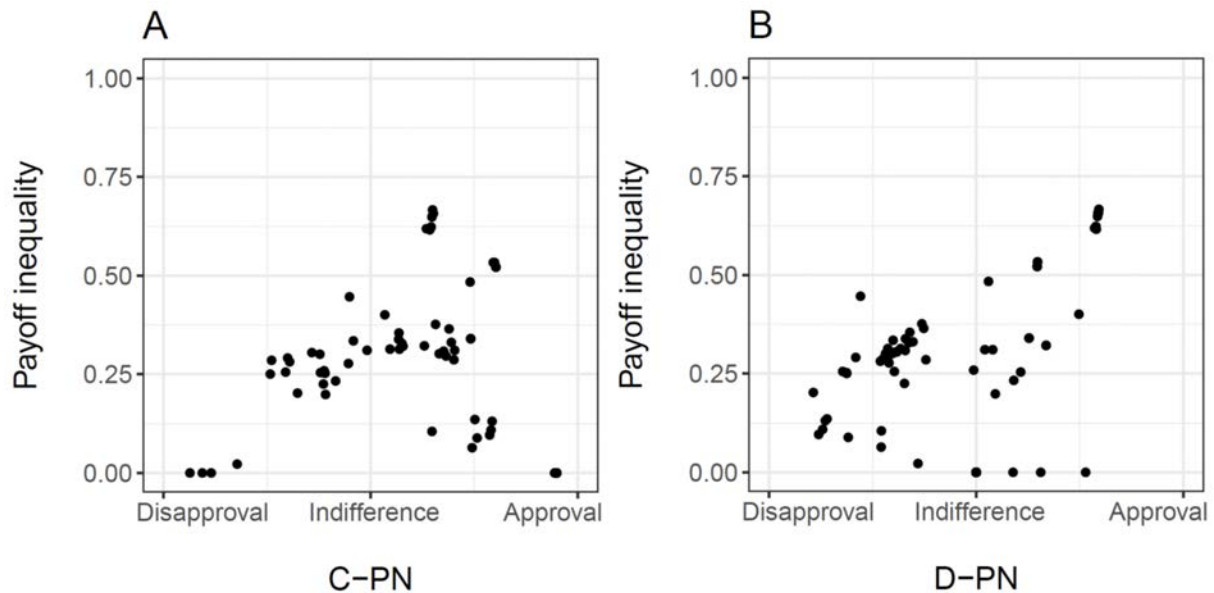


Figure 9: Association of payoff inequality and the personal norm to cooperate (C-PN; Figure A) and to defect (D-PN; Figure B). Each data point represents one agent group composition (out of 84), averaged across agents and time. Duration of model runs: 200 time steps.

Discussion

In our model, norm internalization overall promoted the emergence of cooperation. Regarding the vast amount of literature stemming from developmental psychology, moral psychology, economics, etc., closely connecting norm internalization to morality, this result may not seem surprising (Bicchieri & Dimant 2019; Haidt 2001; Hoffman 2000; Nyborg 2018; Schwartz 1977a; Thøgersen 1999). While we drew on these approaches to formulate assumptions on internalization mechanisms, our conceptualization differs significantly. We did not employ an ethical conceptualization in the sense of linking personal norms to “moral” or “prosocial” behaviour, but rather described the mechanisms for making a judgment on appropriateness or inappropriateness. Interestingly, the model showed that our conceptualization of norm internalization as a slow adaptation process, storing and abstracting part of the situational learning, promotes cooperation – a result that strongly relates to the above-mentioned literature.

Norm internalization increased the intention to show a behaviour with agents still being able to behave against their internalized norms. This result is highly consistent with empirical research (e.g., Bamberg & Schmidt 2003) and psychological theories (e.g., Schwartz 1977a; Bandura 2001). The DINO model represents, to our knowledge, a first approach towards understanding these empirical norm-behaviour relations. It sheds not only light onto the question when norm internalization does lead to norm-consistent behaviour but also when it does not.

Norm internalization increased behavioural volatility and later occurrence of lock-in phenomena. Especially the ability to disapprove of a norm displays displayed agents’ multi-dimensional motivational structure, leading to counterintuitive effects of a conditional cooperator (type 5) defecting more and a defector (type 6) cooperating more. In the model, personal norms influence the importance of motivational factors, making individual preferences malleable by the situation and hence agents more adaptive. That way, agents are (slowly) transformed by the situation, making the norm internalization a transformative process of the individual as has been theoretically assumed (Piaget 1970; Vygotsky 1930). Similarly, Gintis (2004) argued that “internalization of norms

is thus adaptive because it facilitates the transformation of drives, needs, desires and pleasures” and “altering agents’ goals” (p. 62).

Whereas introducing norm internalization increased cooperation, it decreased payoff equality. The rise in cooperation tended to be driven by single agents approving of the cooperation norm, making agents more persistent cooperators. As a result, they cooperated more even in groups with defecting others. This led conditional cooperators and defectors in mixed groups to approve of the defection norm more easily and defect increasingly. Hence, introducing norm internalization had polarizing effects on motivation as well as payoff distributions. Social norms may have different, sometimes contradictory macro level social effects, such as reducing social friction and improving coordination (Sen & Airiau 2007; Shoham & Tennenholtz 1992), maintaining or undermining social cohesion (Taylor & Davis 2018), and solving or enforcing social inequality (Conte & Castelfranchi 1995; Saam & Harrer 1999; Ullmann-Margalit 1977). Regarding norm internalization, there is generally little knowledge about its macro level effects. Lorenz et al. (2021) have shown that motivated cognition causes societal attitude polarization, which relates to the DINO norm internalization process, building on motivated reasoning literature (Festinger 1957; Kunda 1990; Rozin 1999). It seems plausible to assume that in social groups with multiple norms, internalization of different norms may lead to increased inequality.

Notes

¹As the two social descriptive norm expectations are assumed negatively dependent in the present decision scenario, we refrained from representing them as ranging from appropriateness to inappropriateness for simplicity and consistency with other expectations.

²Based on the nature of the data, we calculated robust correlation coefficients, i.e., the percentage bend correlation p_{pb} , since in standard correlation measures such as Pearson’s correlation outliers and normality violations have strong impacts.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211
- Andrighetto, G., Campenni, M., Cecconi, F. & Conte, R. (2010a). The complex loop of norm emergence: A simulation model. In K. Takadama, C. Cioffi-Revilla & G. Deffuant (Eds.), *Simulating Interacting Agents and Social Phenomena*, (pp. 19–35). Berlin Heidelberg: Springer
- Andrighetto, G., Villatoro, D. & Conte, R. (2010b). Norm internalization in artificial societies. *AI Communications*, 23(4), 325–339
- Atkinson, J. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, 64, 359–372
- Axelrod, R. (1984). *The Evolution of Cooperation*. New York, NY: Basic Books
- Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review*, 80(4), 1095–1111
- Bamberg, S., Hunecke, M. & Blöbaum, A. (2007). Social context, personal norms and the use of public transportation: Two field studies. *Journal of Environmental Psychology*, 27(3), 190–203
- Bamberg, S. & Schmidt, P. (2003). Incentives, morality, or habit? Predicting students’ car use for university routes with the models of Ajzen, Schwartz, and Triandis. *Environment and Behavior*, 35(2), 264–285
- Bandura, A. (1971). *Social Learning Theory*. New York, NY: General Learning Press
- Bandura, A. (1999). Social cognitive theory: An agentic perspective. *Asian Journal of Social Psychology*, 2(1), 21–41
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52(1), 1–26
- Batzke, M. C. L. & Ernst, A. (2022). Explaining and resolving norm-behavior inconsistencies – A theoretical agent-based model. In M. Czupryna & B. Kamiński (Eds.), *Advances in Social Simulation*, (pp. 41–52). Berlin Heidelberg: Springer

- Bem, D. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 74(3), 183
- Bem, D. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, vol. 6, (pp. 1–62). Cambridge, MA: Academic Press
- Bendor, J. & Swistak, P. (2001). The evolution of norms. *American Journal of Sociology*, 106(6), 1493–1545
- Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press
- Bicchieri, C. & Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public Choice*, 191, 1–22
- Brandon, G. & Lewis, A. (1999). Reducing household energy consumption: A qualitative and quantitative field study. *Journal of Environmental Psychology*, 19, 75–85
- Briegel, R., Ernst, A., Holzhauer, S., Klemm, D., Krebs, F. & Martínez Piñáñez, A. (2012). Social-ecological modelling with LARA: A psychologically well-founded lightweight agent architecture. International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet. Sixth Biennial Meeting, Leipzig, Germany. Available at: <http://www.iemss.org/society/index.php/iemss-2012-proceedings>
- Broersen, J., Dastani, M., Hulstijn, J., Huang, Z. & Torre, L. (2001). The BOID architecture: Conflicts between beliefs, obligations, intentions and desires. Agents '01: Proceedings of the fifth international conference on Autonomous agents. Association for Computing Machinery
- Burlando, R. & Guala, F. (2005). Heterogeneous agents in public goods experiments. *Experimental Economics*, 8(1), 35–54
- Camerer, C. (2003). *Behavioral Game Theory*. New York, NY: Russell Sage
- Castelfranchi, C., Dignum, F., Jonker, C. M. & Treur, J. (2000). Deliberative normative agents: Principles and architecture. In N. R. Jennings & Y. Lesperance (Eds.), *Intelligent Agents VI. Agent Theories, Architectures, and Languages*, vol. 1757, (pp. 364–378). Berlin Heidelberg: Springer
- Cialdini, R. B., Reno, R. R. & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015–1026
- Conner, M. & Armitage, C. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology*, 28(15), 1429–1464
- Conte, R., Andrighetto, G. & Campenni, M. (2010). Internalizing norms: A cognitive model of (social) norms' internalization. *International Journal of Agent Technologies and Systems*, 2(1), 63–73
- Conte, R. & Castelfranchi, C. (1995). Understanding the functions of norms in social groups through simulation. In N. Gilbert & R. Conte (Eds.), *Artificial Societies: The Computer Simulation of Social Life*, (pp. 213–226). London: Routledge
- Costa, P. T. & McCrae, R. R. (1986). Personality stability and its implications for clinical psychology. *Clinical Psychology Review*, 6(5), 407–423
- Dannenberg, A., Gutsche, G., Batzke, M., Christens, S., Engler, D., Mankat, F., Möller, S., Weingärtner, E., Ernst, A., Lumkowsky, M., Wangenheim, G., Hornung, G. & Ziegler, A. (2023). The effects of norms on environmental behavior. *Review of Environmental Economics and Policy*, in press
- Dawes, R. (1980). Social dilemmas. *Annual Review of Psychology*, 31, 169–193
- de Oliveira, A. C. M., Croson, R. T. A. & Eckel, C. (2015). One bad apple? Heterogeneity and information in public good provision. *Experimental Economics*, 18(1), 116–135
- Deci, E. & Ryan, R. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268
- Deci, E. L. & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, 19(2), 109–134

- Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution*, 2(3), 265–279
- Deyoung, C. G., Peterson, J. B. & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, 33(4), 533–552
- Dowling, D. (1999). Experimenting on theories. *Science in Context*, 12(2), 261–273
- Durkheim, E. (1893). *Über soziale Arbeitsteilung. Studie über die Organisation höherer Gesellschaften [The Division of Labour in Society]*. Frankfurt am Main: Suhrkamp
- Epstein, J. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton: Princeton University Press
- Ernst, A. (2003). Agentenbasierte Modellierung des Handelns in Gemeingutdilemmata. *Jahrbuch Ökologische Ökonomik*, 3, 139–170
- Ernst, A. (2010). Social simulation: A method to investigate environmental change from a social science perspective. In M. Gross & H. Heinrichs (Eds.), *Environmental Sociology*, (pp. 109–122). Berlin Heidelberg: Springer
- Fehr, E. & Fischbacher, U. (2002). Why social preferences matter – The impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal*, 112(478), 1–33
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Palo Alto, CA: Stanford University Press
- Fischbacher, U. & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1), 541–556
- Fischbacher, U., Gächter, S. & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397–404
- Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Boston, MA: Addison-Wesley
- Fishbein, M. & Ajzen, I. (1981). Attitudes and voting behavior: An application of the theory of reasoned action. *Progress in Applied Social Psychology*, 1(1), 253–313
- Frank, R. (1988). *Passions Within Reason: The Strategic Role of the Emotions*. W.W. Norton & Co
- Freud, S. (1932). *Neue Folge der Vorlesungen zur Einführung in die Psychoanalyse [New Introductory Lectures on Psychoanalysis]*. Fischer
- Gächter, S. & Thöni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, 3(2–3), 303–314
- Gigerenzer, G. (2001). The adaptive toolbox. In G. Gigerenzer & R. Selten (Eds.), *Bounded Rationality: The Adaptive Toolbox*, (pp. 37–50). Cambridge, MA: MIT Press
- Gintis, H. (2004). The genetic side of gene-culture coevolution: Internalization of norms and prosocial emotions. *Journal of Economic Behavior and Organization*, 53, 57–67
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814
- Hamann, K. R., Reese, G., Seewald, D. & Loeschinger, D. C. (2015). Affixing the theory of normative conduct (to your mailbox): Injunctive and descriptive norms as predictors of anti-ads sticker use. *Journal of Environmental Psychology*, 44, 1–9
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859), 1243–1248
- Harland, P., Staats, H. & Wilke, H. A. (1999). Explaining proenvironmental intention and behavior by personal norms and the theory of planned behavior. *Journal of Applied Social Psychology*, 29(12), 2505–2528
- Harris, M. A., Brett, C. E., Johnson, W. & Deary, I. J. (2016). Personality stability from age 14 to age 77 years. *Psychology and Aging*, 31(8), 862
- Hartig, B., Irlenbusch, B. & Kölle, F. (2015). Conditioning on what? Heterogeneous contributions and conditional cooperation. *Journal of Behavioral and Experimental Economics*, 55, 48–64

- Hines, J. M., Hungerford, H. R. & Tomera, A. N. (1987). Analysis and synthesis of research on responsible environmental behavior: A meta-analysis. *The Journal of Environmental Education*, 18(2), 1–8
- Hoffman, M. (1977). Moral internalization: Current theory and research. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, vol. 10, (pp. 85–133). New York, NY: Academic Press
- Hoffman, M. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge: Cambridge University Press
- Hollander, C. D. & Wu, A. S. (2011). The current state of normative agent-based systems. *Journal of Artificial Societies and Social Simulation*, 14(2), 6
- Horne, C. (2003). The internal enforcement of norms. *European Sociological Review*, 19(4), 335–343
- Jacobson, R. P., Mortensen, C. R. & Cialdini, R. B. (2011). Bodies obliged and unbound: Differentiated response tendencies for injunctive and descriptive social norms. *Journal of Personality and Social Psychology*, 100(3), 433
- Jager, W. (2000). Modelling consumer behaviour. Available at: <https://research.rug.nl/en/publications/modelling-consumer-behaviour>
- Jager, W. (2017). Enhancing the realism of simulation (EROS): On implementing and developing psychological theory in social simulation. *Journal of Artificial Societies and Social Simulation*, 20(3), 14
- Jager, W. & Ernst, A. (2017). Introduction of the special issue: "Social simulation in environmental psychology". *Journal of Environmental Psychology*, 52, 114–118
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5), 1449–1475
- Kaiser, F. & Scheuthle, H. (2003). Two challenges to a moral extension of the theory of planned behavior: Moral norms and just world beliefs in conservationism. *Personality and Individual Differences*, 35(5), 1033–1048
- Kangur, A., Jager, W., Verbrugge, R. & Bockarjova, M. (2017). An agent-based model for diffusion of electric vehicles. *Journal of Environmental Psychology*, 52, 166–182
- Kohlberg, L. (1964). Development of moral character and moral ideology. *Review of Research in Child Development*, 1, 383–431
- Kohlberg, L. (1984). *Essays on Moral Development: The Psychology of Moral Development*. Skokie, IL: Row Publishers, Inc
- Kohlberg, L. & Hersh, R. H. (1977). Moral development: A review of the theory. *Theory into Practice*, 16(2), 53–59
- Kottonau, J. & Pahl-Wostl, C. (2004). Simulating political attitudes and voting behavior. *Journal of Artificial Societies and Social Simulation*, 7(4), 6
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480
- Kurzban, R. & Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences*, 102(5), 1803–1807
- Lindenberg, S. & Steg, L. (2007). Normative, gain and hedonic goal frames guiding environmental behavior. *Journal of Social Issues*, 63(1), 117–137
- Lorenz, J., Neumann, M. & Schröder, T. (2021). Individual attitude change and societal dynamics: Computational experiments with psychological theories. *Psychological Review*, 128(4), 623
- Lucas, P., Oliveira, A. & Banuri, S. (2014). The effects of group composition and social preference heterogeneity in a public goods game: An agent-based simulation. *Journal of Artificial Societies and Social Simulation*, 17(3), 5
- Luce, R. D. & Raiffa, H. (1957). *Games and Decisions: Introduction and Critical Survey*. Hoboken, NJ: John Wiley & Sons

- Mahmoud, M. A., Ahmad, M. S., Mohd Yusoff, M. Z. & Mustapha, A. (2014). A review of norms and normative multiagent systems. *The Scientific World Journal*, 2014, 684587
- Mahmoud, S., Griffiths, N., Keppens, J. & Luck, M. (2012). Norm emergence: Overcoming hub effects in scale free networks. Proceedings of the AAMAS 2012 workshop on coordination, organizations, institutions and norms
- Markus, H. & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology*, 51, 858–866
- Martinsson, P., Villegas-Palacio, C. & Wollbrant, C. (2009). Conditional cooperation and social group-experimental results from Colombia. Discussion Paper Series, Environment for Development. Available at: https://www.jstor.org/stable/resrep14913#metadata_info_tab_contents
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50, 370–396
- McCrae, R. (2000). *Emotional Intelligence from the Perspective of the Five-Factor Model of Personality*. Hoboken, NJ: John Wiley & Sons
- Messick, D. & McClintock, C. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, 4
- Miller, N. & Dollard, J. (1941). *Social Learning and Imitation*. New Haven, CT: Yale University Press
- Murphy, R. & Ackermann, K. (2013). Explaining behavior in public goods games: How preferences and beliefs affect contribution levels. Available at SSRN: <https://ssrn.com/abstract=2244895>
- Murphy, R. & Ackermann, K. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, 18(1), 13–41
- Murphy, R., Ackermann, K. & Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771–781
- Nerb, J., Spada, H. & Ernst, A. (1997). A cognitive model of agents in a commons dilemma. Proceedings of the 19th annual conference of the Cognitive Science Society
- Neumann, M. (2008). Homo socionicus: A case study of simulation models of norms. *Journal of Artificial Societies and Social Simulation*, 11(4), 6
- Neumann, M. (2010a). A classification of normative architectures. In K. Takadama, C. Cioffi-Revilla & G. Deffuant (Eds.), *Simulating Interacting Agents and Social Phenomena*, (pp. 3–18). Berlin Heidelberg: Springer
- Neumann, M. (2010b). Norm internalisation in human and artificial intelligence. *Journal of Artificial Societies and Social Simulation*, 13(1), 12
- Noosey, L., Isaac, R. M., Norton, D. & Stinn, J. (2020). Cooperation, contributor types, and control questions. *Journal of Behavioral and Experimental Economics*, 85, 101489
- Nowak, M., Sasaki, A., Taylor, C. & Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983), 646–650
- Nyborg, K. (2018). Social norms and the environment. *Annual Review of Resource Economics*, 10, 405–423
- Ostrom, E., Dietz, T., Dolšak, N., Stern, P., Stonich, S. & Weber, E. (2002). *The Drama of the Commons*. Washington, DC: National Academy Press
- Otto, S. & Kaiser, F. (2014). Ecological behavior across the lifespan: Why environmentalism increases as people grow older. *Journal of Environmental Psychology*, 40, 331–338
- Ouellette, J. & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, 124(1), 54–74
- Parsons, T. (1937). *The Structure of Social Action*. New York, NY: Free Press
- Perugini, M. & Bagozzi, R. (2001). The role of desires and anticipated emotions in goal-directed behaviours: Broadening and deepening the theory of planned behaviour. *British Journal of Social Psychology*, 40(1), 79–98

- Petty, R. & Cacioppo, J. (1986). The elaboration likelihood model of persuasion. In R. Petty & J. Cacioppo (Eds.), *Communication and Persuasion*, (pp. 1–24). Berlin Heidelberg: Springer
- Piaget, J. (1970). Piaget's theory. In P. Mussen (Ed.), *Carmichaels' Manual of Child Psychology*, vol. 1, (pp. 703–732). Hoboken, NJ: John Wiley & Sons
- Postman, L. (1947). The history and present status of the law of effect. *Psychological Bulletin*, 44(6), 489–563
- Pyszczynski, T. & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. *Advances in Experimental Social Psychology*, 20, 297–340
- Rivis, A. & Sheeran, P. (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. *Current Psychology*, 22(3), 218–233
- Rozin, P. (1999). The process of moralization. *Psychological Science*, 10(3), 218–221
- Ryan, R. & Deci, E. (2017). *Self-Determination Theory. Basic Psychological Needs in Motivation, Development, and Wellness*. New York, NY: The Guilford Press
- Saam, N. & Harrer, A. (1999). Simulating norms, social inequality, and functional change in artificial societies. *Journal of Artificial Societies and Social Simulation*, 2(1), 2
- savarimuthu, B., Cranefield, S., Purvis, M. & Purvis, M. (2007). Role model based mechanism for norm emergence in artificial agent societies. In J. Sichman, J. Padget, S. Ossowski & P. Noriega (Eds.), *International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*, (pp. 203–217). Berlin Heidelberg: Springer
- Scalco, A., Ceschi, A., Shiboub, I., Sartori, R., Frayret, J. M. & Dickert, S. (2017). The implementation of the theory of planned behavior in an agent-based model for waste recycling: A review and a proposal. In A. Alonso-Betanzos, N. Sánchez-Marroño, O. Fontenla-Romero, J. Polhill, T. Craig, J. Bajo & J. Corchado (Eds.), *Agent-Based Modeling of Sustainable Behaviors*, (pp. 77–97). Berlin Heidelberg: Springer
- Schlüter, M., Baeza, A., Dressler, G., Frank, K., Groeneveld, J., Jager, W., Janssen, M., McAllister, R., Müller, B., Orach, K., Schwarz, N. & Wijermans, N. (2017). A framework for mapping and comparing behavioural theories in models of social-ecological systems. *Ecological Economics*, 131, 21–35
- Schönbach, P. (1990). *Account Episodes: The Management or Escalation of Conflict*. Cambridge: Cambridge University Press
- Schwartz, S. (1977a). Normative influences on altruism. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, vol. 10, (pp. 221–279). Cambridge, MA: Academic Press
- Schwartz, S. (1977b). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in Experimental Social Psychology*, vol. 25, (pp. 1–65). Cambridge, MA: Academic Press
- Schwartz, S. & Fleishman, J. (1982). Effects of negative personal norms on helping behavior. *Personality and Social Psychology Bulletin*, 8(1), 81–86
- Schwartz, S. & Howard, J. (1981). A normative decision-making model of altruism. In J. Rushton (Ed.), *Altruism and Helping Behaviour: Social, Personality and Developmental Perspectives*, (pp. 189–211). Lawrence Erlbaum Associates Inc
- Schwartz, S. & Howard, J. (1982). Helping and cooperation: A self-based motivational model. In V. Derlega & J. Grzelak (Eds.), *Cooperation and Helping Behavior: Theories and Research*, (p. 327–353). Cambridge, MA: Academic Press
- Schwarz, N. & Ernst, A. (2009). Agent-based modeling of the diffusion of environmental innovations – An empirical approach. *Technological Forecasting and Social Change*, 76(4), 497–511
- Sen, S. & Airiau, S. (2007). Emergence of norms through social learning. Proceedings of the Twentieth International Joint Conference on Artificial Intelligence

- Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. *European Review of Social Psychology, 12*(1), 1–36
- Sherif, M. & Sherif, C. (1953). *Groups in Harmony and Tension*. New York, NY: Harper & Brothers
- Shin, Y., Im, J., Jung, S. & Severt, K. (2018). The theory of planned behavior and the norm activation model approach to consumer behavior regarding organic menus. *International Journal of Hospitality Management, 69*, 21–29
- Shoham, Y. & Tennenholtz, M. (1992). On the synthesis of useful social laws for artificial agent societies. Proceedings of the AAAI Conference, Stanford, CA
- Shoham, Y. & Tennenholtz, M. (1995). On social laws for artificial agent societies: Off-line design. *Artificial Intelligence, 73*(1–2), 231–252
- Simon, L., Greenberg, J. & Brehm, J. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology, 68*(2), 247
- Steg, L. & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology, 29*(3), 309–317
- Sutton, R. & Barto, A. (2018). *Reinforcement Learning: An Introduction*. MIT Press
- Snyder, M. (1984). When belief creates reality. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, vol. 18, (pp. 247–305). Cambridge, MA: Academic Press
- Szekely, A., Lipari, F., Antonioni, A., Paolucci, M., Sánchez, A., Tummolini, L. & Andrighetto, G. (2021). Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications, 12*(1), 1–7
- Taylor, J. & Davis, A. (2018). Social cohesion. *The International Encyclopedia of Anthropology*
- Terracciano, A., McCrae, R. R. & Costa Jr, P. T. (2010). Intra-individual change in personality stability and age. *Journal of Research in Personality, 44*(1), 31–37
- Terrier, L. & Marfaing, B. (2015). Using social norms and commitment to promote pro-environmental behavior among hotel guests. *Journal of Environmental Psychology, 44*, 10–15
- Thøgersen, J. (1999). The ethical consumer: Moral norms and packaging choice. *Journal of Consumer Policy, 22*(4), 439–460
- Thøgersen, J. (2003). Monetary incentives and recycling: Behavioural and psychological reactions to a performance-dependent garbage fee. *Journal of Consumer Policy, 26*, 197–228
- Troitzsch, K. (2017). Axiomatic theory and simulation. A philosophy of science perspective on Schelling's segregation model. *Journal of Artificial Societies and Social Simulation, 20*(1), 10
- Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323
- Ullmann-Margalit, E. (1977). *The Emergence of Norms*. Oxford: Clarendon Press
- Verhagen, H. (2001). Simulation of the learning of norms. *Social Science Computer Review, 19*(3), 296–306
- Villatoro, D., Andrighetto, G., Conte, R. & Sabater-Mir, J. (2015). Self-policing through norm internalization: A cognitive solution to the tragedy of the digital commons in social networks. *Journal of Artificial Societies and Social Simulation, 18*(2), 2
- Voisin, D. & Fointiat, V. (2013). Reduction in cognitive dissonance according to normative standards in the induced compliance paradigm. *Social Psychology, 44*(3), 191–195
- Vygotsky, L. (1930). The genesis of higher mental functions. He concept of activity in Soviet psychology
- Vygotsky, L. (2004). Analysis of sign operations of the child. In R. Rieber & D. Robinson (Eds.), *The Essential Vygotsky*, (pp. 557–569). Kluwer Academic/Plenum Press

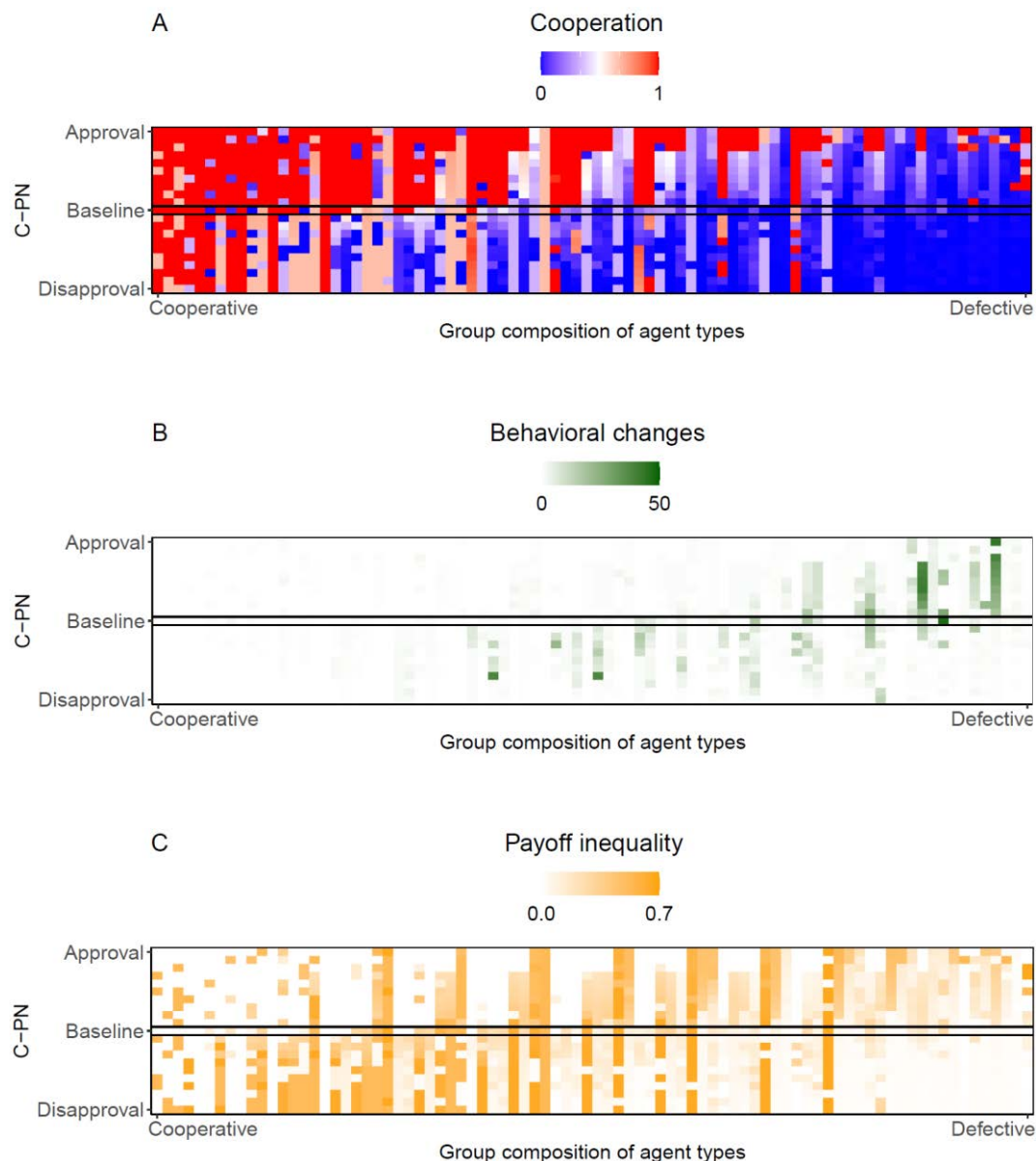
- Webb, T. & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, 132(2), 249
- White, K., Smith, J., Terry, D., Greenslade, J. & McKimmie, B. (2009). Social influence in the theory of planned behaviour: The role of descriptive, injunctive, and in-group norms. *British Journal of Social Psychology*, 48(1), 135–158
- Whitmarsh, L. & O’Neill, S. (2010). Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. *Journal of Environmental Psychology*, 30(3), 305–314
- Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston, IL

Appendix C

Results from DINO Model 1.1

Figure C1

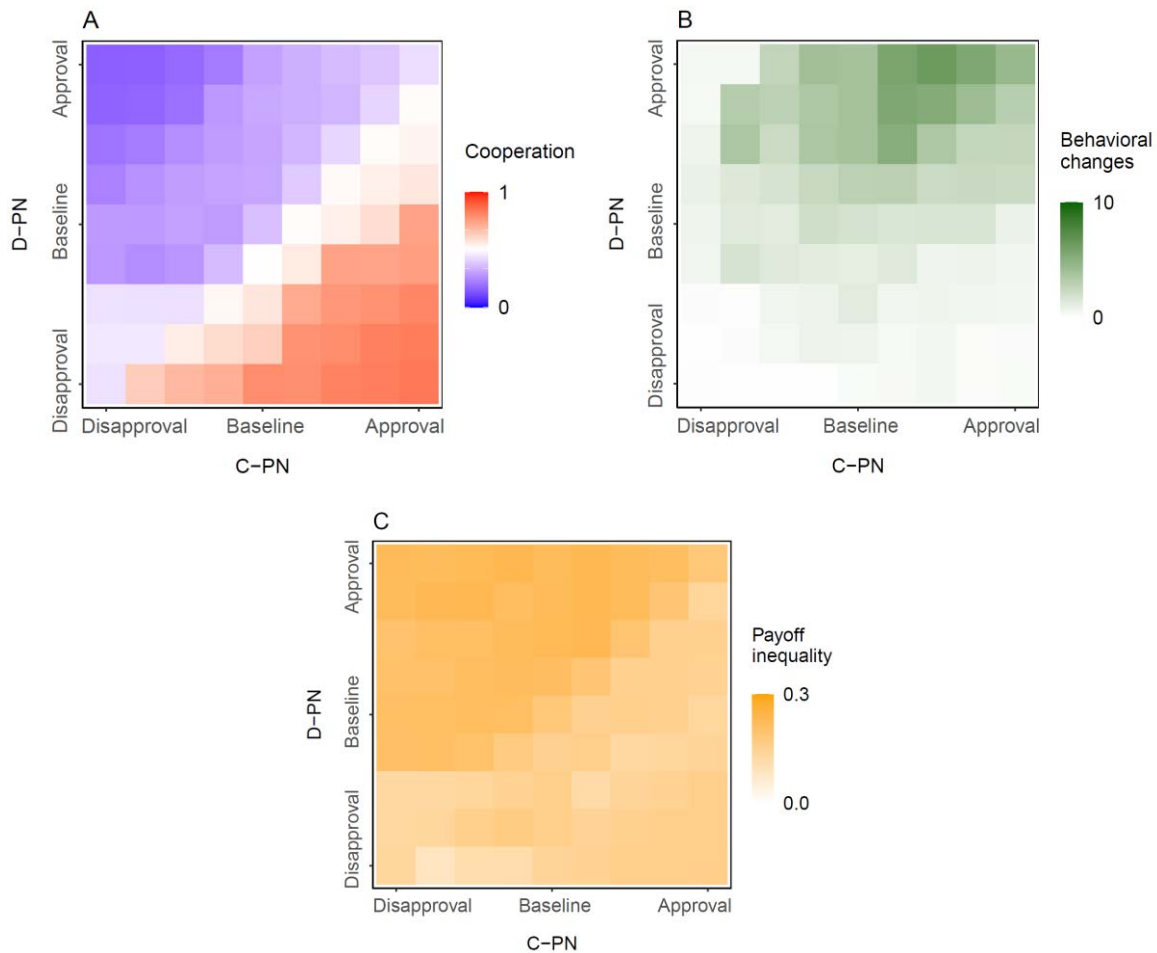
Manipulation of personal norms to cooperate (C-PN) with noise in expectation start values



Note. The personal norm to cooperate (C-PN) was varied from full disapproval to full approval. No manipulation was conducted in baseline model runs. Expectation start values, per default set to neutral midpoints of 0.5, were randomly added with a noise factor (a value drawn from a normal distribution $\mathcal{N}(0,0.1)$). Each model run was repeated 10 times. Left to right shows agent group compositions, ordered along cooperativeness. Group compositions are defined by three-digit numbers, indicating the three agent types in the group, ordered by digit sum and largest single digit. Effects are shown regarding (A) cooperation (ranging between [0,1], averaged across agents and time), (B) absolute number of behavioral changes, and (C) inequality between agents' individual payoffs (ranging between [0,1], averaged across agents and time). Duration of model runs: 200 time steps.

Figure C2

Manipulation of personal norms to cooperate (C-PN) and to defect (D-PN) with noise in expectation start values



Note. Personal norms (C-PN and D-PN) were varied from full disapproval to full approval. Expectation start values, per default set to neutral midpoints of 0.5, were randomly added with a noise factor (a value drawn from a normal distribution $N(0,0.1)$). Each model run was repeated 10 times. Effects are shown regarding (A) cooperation (ranging between [0,1], averaged across agents, time, and group compositions), (B) absolute number of behavioral changes (averaged across group compositions), and (C) inequality between agents' individual payoffs (ranging between [0,1], averaged across agents, time, and group compositions). Results were averaged across 84 group compositions of agent types. Duration of model runs: 200 time steps.

Appendix D

Paper III: Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation

The following manuscript was submitted for publication.

Batzke, M. C. L., & Ernst, A. (2023b). *Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation* [Manuscript submitted for publication]. Center for Environmental Systems Research, University of Kassel.

Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation

Marlene C. L. Batzke^{1*} [0000-0001-5882-9813] and Andreas Ernst¹ [0000-0001-5773-4441]

¹ Center for Environmental Systems Research, University of Kassel, Germany

* Corresponding author

Abstract

Psychological research on norms has shown that norms are highly relevant for individuals' decision-making. Yet, there is so far little understanding of how norms change over time. Knowledge about how norms change may help better understanding their potential for as well as limitations in guiding decision-making and changing behavior. The present work investigated change in individuals' cooperation norms. As an indicator of different underlying processes of norm change, the temporal dynamics of different types of norms were examined. It was assumed that participants' social norms are adapted quickly whenever the social situation changes, while personal norms change more slowly and gradually, abstracting part of the situational learning in interaction with one's personality. In an experimental study, participants played a repeated social dilemma game with artificial co-players representing a predominantly cooperative or uncooperative social setting, depending on the experimental condition. The condition was expected to affect slow learning of personal norms. Additionally, the cooperativeness of the social setting was varied repeatedly *within* conditions, expected to result in fast changes in social norms. Participants' personal and social norms were assessed throughout the game. As predicted, the temporal dynamics differed between norms with social norms changing quickly and personal norms more slowly. Personal norms strongly predicted behavioral decision-making and were predicted by situational and personality factors. Potential qualitative differences of the underlying norm change processes are discussed.

Keywords: Social norms, personal norms, norm change, cooperation, social dilemma game, decision-making

1. Introduction

Humans are social beings. We attend to what others do, what others believe, and form assumptions about what that is others believe and do. We then use these assumptions to make accurate and appropriate behavioral decisions. One of the key features in how social influence affects behavior is described in the concept of social norms. Social norms describe what many people consider appropriate or normal behavior and change over time – sometimes rapidly, sometimes strikingly slowly. While individuals perceive and learn social norms, they also develop their own personal norms, which may well differ from social and societal norms, being influenced by the individuals' experiences, their social network, and so forth. Over the past decades, psychological research has well documented the power of social norms on individuals' behavioral decision-making. However, there is still little understanding of how norms dynamically unfold over time (van Kleef et al., 2019). How do norms develop and change? Knowledge about how norms change may help better understanding their potential for as well as limitations in guiding decision-making and changing behavior (Andrighetto & Vriens, 2022). In the present work, differences in the temporal dynamics of norm change are explored, which may be considered a proxy for different underlying processes of norm change.

1.1 Norms of cooperation

Norms can be defined as behavioral rules for a specific situation (Dannenbergh et al., in press). While norms may regulate any type of behavior, norms of cooperation have received a great deal of attention, being a key feature in the well-functioning of social life and the emergence of collective cooperation (Axelrod, 1986; Bicchieri et al., 2018; Biel & Thøgersen, 2007). As every person tends to act on his or her self-interest, collective cooperation is a fragile state (Hardin, 1968). Many social and societal crisis result from selfish actions. Cooperation is the favored solution to many possibly conflicting situations in small groups and society at large, such as living in a shared apartment or protecting the environment. Cooperation is particularly difficult to achieve in so-called social dilemma situations. In a social dilemma game, a player can choose each round between only two actions: cooperation, being the collective beneficial action, and non-cooperation or defection, being the choice maximizing the own benefit. A player receives a higher payoff for the defecting choice; however, all players are better off if all cooperate rather than defect (Axelrod, 1984). Hence, the individual's self-interest is at odds with the collective interest (Dawes, 1980). Repeated social dilemmas existing of several rounds allow for the development of social norms (Peysakhovich & Rand, 2015). Norms have been

shown to motivate cooperative behavior in social dilemmas, influencing small group cooperation (Bicchieri et al., 2023; Ostrom, 2000) and creating tipping points for large-scale transformations (Nyborg et al., 2016; Otto et al., 2020). The present work therefore addresses cooperative norms and cooperative decision-making in a social dilemma situation.

1.2 Types of norms

Social situations are informed by a multitude of social norms, being what many people consider appropriate or normal behavior. The power of social norms has long been known in psychology (Asch, 1956; Deutsch & Gerard, 1955; Sherif, 1936). People are highly sensible to their social surrounding, yielding to social norm pressure and conforming to the group to gain social approval and avoid social sanctions (Cialdini et al., 1990; McDonald & Crandall, 2015; Nyborg et al., 2016). Providing people with information about social norms affects their behavior (Goldstein et al., 2008; Keizer et al., 2008; Miller & Prentice, 2016; Schultz et al., 2007, 2008) – even without them consciously knowing about it (e.g., Nolan et al., 2008). People tend to assimilate with social norms in the sense of imitation and conformism (Cialdini & Goldstein, 2004). Cialdini et al. (1990) demonstrated the distinct importance of two different qualities of social norms: the injunctive and descriptive quality. *(Social) injunctive norms* contain information about the (in)appropriateness of a behavior in a specific situation (i.e., what most others consider (in)appropriate), motivating through the promise of social (dis)approval. *(Social) descriptive norms* refer to the observable regularity/normality of a behavior in a specific situation (i.e., what most others do), motivating by “what will likely be effective and adaptive” (p. 1015).¹

Moreover, norms have been stated to function at different levels, meaning the social/societal and the individual level (Bicchieri et al., 2018; Cialdini et al., 1990; Dannenberg et al., in press; Farrow et al. 2017). Hence, individuals not only perceive and tend to conform to social norms, but also develop their own personal norms, a concept most notably known by the work of Schwartz (1977).² *Personal norms* can be defined as individuals’ beliefs about (in)appropriate behavior in a specific situation, being of an injunctive quality (Batzke & Ernst, 2023). As personal norms are considered more internal to the individual than social norms

¹ Social injunctive and descriptive norms have also been referred to as *normative* and *empirical expectations* (e.g., Bicchieri et al., 2018, 2022; Bicchieri & Xiao, 2014; Szekely et al., 2023).

² Personal norms have also been referred to as *personal normative beliefs* (e.g., Andrighetto et al., 2015; Bicchieri & Xiao, 2014) or *moral norms* (Bicchieri & Dimant, 2019; Nyborg, 2018; Thøgersen, 1999).

(Thøgersen, 2006), they tend to be associated with feelings of moral obligation as well as guilt and shame when violated (Schwartz 1977; Schwartz & Howard 1981, 1982). Research has shown that individuals' personal norms strongly predict behavioral decisions (Bamberg, 2013; Han, 2014; Hunecke et al., 2001; Klöckner & Matthies, 2004; Onwezen et al., 2013; Szekely et al., 2021) and explain variance in behavioral decisions over and above social norms (Conner & Armitage 1998; Harland et al. 1999; Shin et al. 2018). Although Steg and de Groot (2010) demonstrated that personal norms can be manipulated, there is little experimental evidence on the effects of personal norms, research being mostly survey based. Biel and Thøgersen (2007) stated that “there is a need for more research providing unambiguous research on that topic” (p. 105).

Apart from the quality and subject, norms can be differentiated by their orientation, being self-oriented (i.e., concerning the individual's behavior) or other-oriented (i.e., concerning behavior of others). Social norms are most often implicitly conceptualized as other-oriented (e.g., what others do, rather than what others believe I do). Regarding personal norms, there is ambiguity in the literature. Without specific declaration, they have often been operationalized as self-oriented personal norms (i.e., what I consider (in)appropriate *for myself*, see Bamberg, 2013; Bamberg et al., 2007; Klöckner & Matthies, 2004; Han, 2014; Hunecke et al., 2001; Steg & de Groot, 2010), yet occasionally as other-oriented personal norms (i.e., what I consider (in)appropriate *for others*, see Bicchieri et al., 2014) or a combination (i.e., what I consider (in)appropriate behavior *in general*, see Bicchieri et al., 2022, Szekely et al., 2021).

1.3 Norm change

Psychological research on norms has often focused on showing the causal effect of norms on behavioral decisions. The fact that normative information situationally “nudges” decision-making has been well-documented (for reviews see Bicchieri & Dimant, 2019; Miller & Prentice, 2016). The behavioral adaptation is immediate (Cialdini & Goldstein, 2004; Nolan et al., 2008) and in direction of the presented social norm (e.g., Schultz et al., 2007).

The topic of norm change in individuals has gained far less attention in experimental research (Andrighetto & Vriens, 2022). Theoretically, it is assumed that cognitive processes shape the individually learned norms, which in turn shape the social dynamics (Hawkins et al., 2019). Accordingly, acquiring norms is based on social cognition and social learning (Howard & Renfrow, 2003; Kelly & Davis, 2018; Theriault et al., 2021). It is assumed that social constructs become internalized, meaning part of the individual's identity or self, as described in *self-determination theory* (Ryan & Deci, 2000) or *self-categorization theory* (Turner, 1987). Social

norm change has been shown to affect change behavioral decisions (Gino et al., 2009; Nakashima et al., 2017; Paluck, 2009; Peysakhovich & Rand, 2016).

Traditionally, norm psychology is closely related to moral psychology. The therefore also called *moral norms* (c.f. Bicchieri & Dimant, 2019; Nyborg, 2018; Thøgersen, 1999), relating to personal norms, have been assumed to be predominantly acquired in early childhood (Nucci, 2001; Turiel, 1983). The foundation was laid in classic developmental psychological studies, showing how children develop moral principles of right and wrong (Kohlberg, 1964; Piaget, 1970). More recent literature takes a learning perspective on morality, ranging from intuitionist, emotional (e.g., Haidt, 2001) to rational, reward-maximizing approaches (e.g., Crockett, 2013; Cushman, 2013). Hence, it can be assumed that also personal norms are acquired and change throughout the lifetime. Yet, one of the key questions remains how those “moral rules” change (Cushman et al., 2017, p. 1).

While one might assume that some norms remain relatively stable throughout one’s life, such as an individual’s personal norm about brushing the teeth before going to bed, it has also been assumed that norms may continuously change over time (Chudek & Henrich, 2011; Kelly & Davis, 2018; McDonald & Crandall, 2015). People experience new situations every day, challenging them to read social cues to behave in a socially appropriate way. Through repetition (Prentice & Paluck, 2020), social enforcement (Schultz et al., 2007), and internal feedback (Schwartz & Howard, 1981), new norms may develop and existing ones may change. Yet, processes of norm change unfolding over time are little understood (Anderson & Dunning, 2014; Dannels & Miller, 2017; Dannels et al., 2022; van Kleef et al., 2019). Addressing questions regarding norm change is a step towards understanding the mechanisms of norm change, which is “is crucial for identifying interventions which could lead to large-scale behavioral change” (Andrighetto & Vriens, 2022, p. 4) and fostering cooperative decision-making (Cushman et al., 2017).

1.4 Different norm change processes in different types of norms

There are yet few approaches investigating norm change experimentally. In the following, some experimental approaches are presented, focusing on those that also explicitly assessed participants’ social and personal norms. In Szekely et al. (2021), participants were confronted with different collective risks in a long-term social dilemma experiment. It was shown that social norms are adapted along the within-subjects (high vs. low) risk variation. Effects on personal normative beliefs (i.e., personal norms) were not analyzed; descriptive data seemed inconclusive (see Szekely et al., 2021, Supplementary Figure 5). Similarly, Bicchieri et al.

(2022) found that observing norm violations led participants to adapt their social norms, but not their personal normative beliefs. This might lead to believe that personal norms do in fact not change. Tverskoi and colleagues (2023) showed that they can change. Participants played an online common pool resources game for 35 days either with or without messaging. Personal norms as well as normative expectations and empirical expectations (relating to social injunctive and social descriptive norms) all became less cooperative over the course of the game (indicating higher resource extraction).

Norm change processes have also been investigated via modeling and simulation methods. While norms were predominantly conceptualized as social level phenomena of behavioral convergence (e.g., Axelrod, 1986; Sen & Airiau, 2007), some researchers have also focused on individual norm processes, representing norms as mental objects that artificial agents can deliberate upon (Conte & Castelfranchi, 1995; Castelfranchi et al., 2000; Dignum, 1999). Villatoro et al. (2015) characterized agents' internalization of norms as a multi-step process which depends on (1) the salience of the social norm and (2) a cost-benefit-calculation. Batzke and Ernst (2023) introduced the idea of different norm learning processes possessing different temporal dynamics, meaning that they occur at different rates of change (see also Batzke & Ernst, 2022). Whereas social norm learning is presumably merely based on observation, it was assumed a fast adaptation process with a high rate of change. Personal norm learning, however, was supposed to be slower with a low rate of change, being influenced by situational as well as personal factors, abstracting part of the situational learning across situations. It was assumed to manifest subjective experiences of social norms and their individual evaluation over time. Hence, personal norm change was stated to be qualitatively differently, individually specific, and evolving over a longer period of time (i.e., to be slower).

1.5 Research questions

In the present work, the assumption of temporal differences between social and personal norm processes was adopted, using the temporal dynamics as a proxy for potentially different underlying processes. The purpose of the present work is to investigate social and personal norm change experimentally in a social dilemma game, addressing the following questions.

1. Do social and personal norms differ in their temporal dynamics, i.e., their rates of change?
2. Can personal factors explain personal norm change over and above situational factors, while social norms are solely predicted by situational factors?

3. Do personal norms have an independent effect on behavioral decisions (over and above social norms)?
4. Can personal norms be influenced towards more cooperativeness?

2. Materials and methods

2.1 Hypotheses

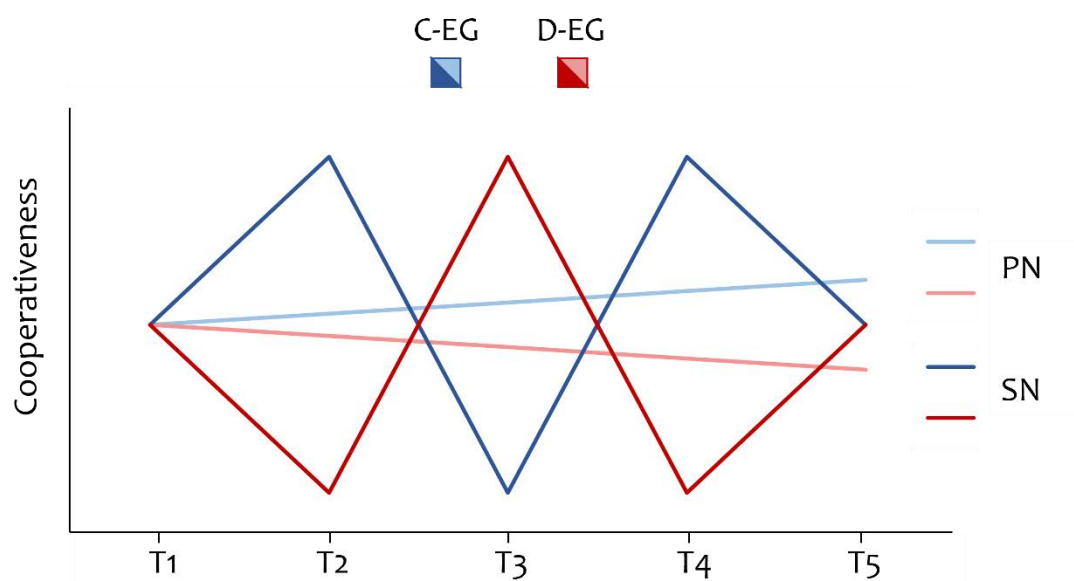
During the social dilemma game, participants were confronted with changing social settings indicating different social norms. The effects of the fast adaptation process were assumed to show in changing social norms according to the social setting. The effects of the slow adaptation process of personal norms were expected to show in differences between groups after the game, due to the different amount of cooperation between groups. Although it seemed unlikely that participants would fully learn and develop personal norms within the brevity of the game, it was assumed that temporal differences in the processes already show in shorter periods of time. No assumptions on how self- and other-oriented personal norms might differ were formed. Thus, hypotheses refer to both types of personal norms. Similarly, the term personal norms will be used for self- and other-oriented personal norms in the present work, if not specified otherwise. Hypotheses on social norms refer to their descriptive and injunctive quality, but only to other-oriented social norms. Similarly, the term “social norms” will only refer to other-oriented social descriptive and injunctive norms in the following. Self-oriented social norms were assessed for exploratory purposes. Addressing the research questions, the following hypotheses were developed:

1. Social norms have a higher rate of change than personal norms. Changes in social norms are in line with the social setting and personal norms develop gradually across the game (see Figure 1).
 - 1.1 The average difference between single consecutive measurements is greater in social norms than in personal norms.
 - 1.2 Social norms are repeatedly reversed during the game in line with the social setting (in the sense of seasonal changes).
 - 1.3 Personal norms change gradually during the game (in the sense of a linear trend).
2. Personal norms are predicted by person and situation factors, social norms only by the situation.

- 2.1 Social norms are solely predicted by the experimental group.
- 2.2 Personal norms are predicted by the experimental group, trait cooperativeness, and their interaction.
3. Personal norms have independent effects on behavioral decisions.
 - 3.1 In addition to social norms, personal norms explain variance in the follow-up behavior.
4. Personal norms may be influenced towards more cooperativeness, which reflects in more cooperative behavioral decision-making.
 - 4.1 After the game, participants in the cooperative group show higher levels of personal norms of cooperation than those in the defective group.
 - 4.2 Participants in the cooperative group cooperate more in the follow-up behavior than those in the defective group.

Figure 1

Hypothesized change in social and personal norms depending on the experimental group



Note. T1 – T5 = measurement time points; C-EG = cooperative group; D-EG = defective group; PN = personal norms; SN = social norms.

2.2 Design and manipulation

$N = 365$ participants (see sample description in Section 2.3) played an online repeated 3-person social dilemma game over 17 rounds. The two co-players were pre-defined behavioral

sequences. The experimental design was a 2 (between-subjects factor *group*: cooperative vs. defective) x 5 (within-subjects factor *time*: T1 – T5) mixed design. The between-subjects factor experimental group varied the overall degree of cooperative actions of the artificial players. In the cooperative experimental group, the majority of the artificial players' actions was cooperative, in the defective experimental group, the majority of actions was defective (see Figure 2). Participants were randomly assigned to groups.

Within each experimental group, the situational social setting was varied repeatedly. Participants experienced different phases of the game, being each several rounds. In the cooperative group, the phases were cooperative-defective-cooperative (C-D-C): five rounds in a cooperative setting, three rounds in a defective setting, and four rounds in a cooperative setting. In the defective group, the phases were reversed to defective-cooperative-defective (D-C-D, see Figure 2). In between two phases, there was each one round of a mixed setting with one of the artificial players cooperating and the other defecting, to make transitions more realistic. After these three phases, a distraction phase followed in both groups, being three rounds of a mixed setting (for the exact setup of the 17 rounds of the game, see Appendix A). The distraction phase was supposed to set the last setting social in each group to a neutral level, before testing for behavioral differences between groups after the game. At in total five measurement time points, participants were asked to state their norms: Before (T1) and (roughly) after each phase (T2 – T5).

Figure 2

Operationalization of experimental groups

	Phase			
	1	2	3	Distraction
C – EG	C	D	C	Mixed
D – EG	D	C	D	Mixed
	T1	T2	T3	T4
				T5

Note. The social dilemma game differed between the cooperative (C-EG) and defective experimental group (D-EG) in the order of their phases. Each experimental group consisted of three phases, characterized by either a cooperative setting (C), in which the artificial players cooperated, or a defective setting (D), in which they defected. The distraction phase was characterized by a mixed setting (Mixed), in which one of the artificial players cooperated and the other defected. Social and personal norms were assessed before the game (T1) and roughly after each phase at T2 – T5. The exact game setup of the 17 rounds is shown in Appendix A.

2.3 Sample

Using the online tool Glimmpse (<https://glimmpse.samplesizeshop.org/>), a target sample size for obtaining a power of 0.95 at a .05 familywise alpha error probability for the interaction effect of the contrast analyses between time (within-subjects factor) and experimental group (between-subjects factor), formulated in Hypotheses 1.2 and 1.3, of $N = 10$ for social norms and $N = 386$ for personal norms was calculated (see preregistration for the specifics, <https://osf.io/xgucf>). In total, $N = 440$ participants were recruited in March/June 2022 and compensated via the survey institute Bilendi, assuming that some data had to be excluded according to the predefined exclusion criteria (see preregistration at <https://osf.io/xgucf>). Full-aged, German-speaking individuals were permitted to participate in the study. The sample was assessed according to age, gender, and income statistics for Germany.

From the initial pool, two participants were excluded due to being underage and eight participants due to low proficiency in the German language. Another 31 participants were excluded for pausing during the social dilemma game for more than three minutes. Participants were directly instructed not to take breaks during the game, as breaks potentially exposed the cover story of participants playing an online game with real other participants. Three participants were excluded due to highly conspicuous item response patterns. Thirty-one participants were removed due to perceiving the game as extremely unreal (being statistical outliers) or, in the open questions at the end of the study, expressing serious doubts concerning the realness of the artificial players or correctly stating the goal of the study.

The final sample comprised $N = 365$ participants with an average age of 46 years ($SD = 16.03$). 47% of the participants categorized themselves as female. 19% stated having a bachelor's or higher educational degree. Further sample characteristics are presented in Appendix B.

2.4 Procedure

The study was conducted as an online study via the platform SoSci Survey (Leiner, 2019). Participants were recruited via the survey institute Bilendi in year 2022. The study was announced as an online game, called the “Concert Game”, as part of a study. To strengthen the cover story of a real-time online game, the link to participate in the study was only active between 8 am and 10 pm so that online matching of participants would seem likely. Completing the study took participants on average 21 minutes ($SD = 6.60$).

Before the game. In the beginning, participants were welcomed and gave their informed consent. Next, trait cooperativeness was assessed. Then, the 3-person social dilemma game, called “The Concert Game”, was explained. Participants received full information on the payoff matrix of the game and were presented with an example round. To ensure that participants read and understood the game instructions, five previously announced, multiple choice comprehension questions followed (see Appendix C). If more than one question was answered incorrectly, game instructions were presented again. If for a second time, more than one of the same comprehension questions was answered incorrectly, participation was terminated. That was the case for 92 participants. The other participants were then asked to rate their social descriptive, social injunctive and personal norms for a first time (T1). Before the game started, to convey the impression of an online game, participants were supposedly matched in a group with two other participants. Participants were presented with a progress bar and told that matching took on average four minutes. After one minute, two more participants supposedly had joined the group and participants were able to start the game.

The Concert Game. The game was explained as follows:

Please imagine the following scenario: You are a passionate musician. Your instrument is the piano. In a while you have an important performance; you play your first big concert. This concert is decisive for your future career as a pianist. To prepare for the concert, you have rented a practice room with a piano identical to the one you will play at the concert for 3 hours per day.

Unfortunately, your practice room is located in a triangle with two other practice rooms, which are used by two people who are also preparing for public performances. The walls of the rooms are very thin, so that you can hear each other practicing. If you play the piano loudly, you disturb the others while practicing. Likewise, if the others play loudly, you will be disturbed. Therefore, all pianos have the option of being played electronically via headphones. This way, others are not disturbed, but the sound production is not the same when playing electronically via headphones, which prevents you from practicing certain subtleties. So via headphones your learning achievement is somewhat limited, but nowhere near as much as when disturbed from the loud practicing of others in the neighboring practice rooms.

You will have to practice with the same two people in the time to come. Every day you may decide anew whether you want to practice with headphones or loudly.

Participants played 17 rounds of the Concert Game. Each round consisted of a decision page, a holding page, and a feedback page. On the decision page, participants were presented with the

choice of practicing with headphones versus practicing loudly. For all decisions, the information on the payoff matrix was presented. To simulate semblance with an online game, after each decision, participants were directed to a holding page, having to wait between 0 and 10 seconds for the artificial players before being able to continue (with longer waiting times in the beginning of the game and after questionnaires). The feedback page showed a table with all players' practice behaviors of the current round, the practice points for the current round and the total collected points. Participants were given no information on the number of rounds played or upcoming. During the game, at T2 – T4 social norms, personal norms, and a manipulation check were assessed.

After the game. After the last round of the game, which participants were not aware of, the same questionnaire as during the game was presented, assessing all types of norms and the manipulation check (T5). Afterwards, participants were told that the game would proceed, and a follow-up behavior was assessed, representing a more aggregated form of behavioral decision (see Section 2.4). Subsequently, the following variables were assessed in the presented order: perceived realness of the game scenario, supposed goal of the study, credibility of the cover story, and demographics. Finally, participants were informed that they had successfully practiced for the concert, debriefed, and dismissed.

2.5 Measures

All materials are shown in Appendix C. If not indicated otherwise, items were presented with a response slider ranging from 1 “not agree at all” to 101 “absolutely agree”. The norm scales as well as the manipulation check scale were created by taking the mean of two items, one for each behavioral option in the game (i.e., playing the piano via headphones vs. playing loudly). All items directed towards defectivity (i.e., playing loudly) were reversed beforehand. Hence, higher values indicate stronger cooperativeness. Internal consistencies for the scales at each measurement time point are presented in Appendix D.

Trait cooperativeness (at T1). The slider measure of social value orientation (Murphy et al., 2011) was applied for the trait cooperativeness measure. Participants were asked to allocate hypothetical money to themselves and to another unknown person. The money is allocated using a single slider for both allocations. The amounts of money received by the person herself and the other person are displayed above and underneath a slider, changing dynamically when the slider is moved. Participants were presented with an example and then asked to make six money allocations. The trait cooperativeness scale was created according to

the instructions given by Murphy and colleagues (2011). Higher values indicate a stronger motivation to cooperate.

Social norms (at T1 – T5). (Other-oriented) social descriptive norms were assessed via two items, such as “The others mostly play the piano via headphones.” (Other-oriented) social injunctive norms were assessed via: “The others believe that they should play the piano via headphones.” and the inverted item for the non-cooperative option. For exploratory purposes, self-oriented social norms were additionally assessed. Self-oriented social descriptive norms were assessed via two items, for instance: “The others believe that I mostly play the piano via headphones.” Self-oriented social injunctive norms were assessed via: “The others believe that I should play the piano via headphones.” and its inversion. Items for social descriptive norms at T1 slightly differed, assessing expectations (see Appendix C).

Personal norms (at T1 – T5). Other-oriented personal norms were assessed via: “I am deeply convinced that the others should play the piano via headphones.” and its inversion. Self-oriented personal norms were assessed via: “I am deeply convinced that I should play the piano via headphones.” and its inversion.

Manipulation check (at T2 – T5). Two items were used to indicate a successful manipulation, such as: “In the past two days, the others have mostly played the piano via headphones.”

Follow-up behavior (at T5). After the game, a follow-up behavior was assessed, measuring an aggregated behavior across multiple decisions, via the item: “Please decide now on how you will practice the next five days.” Participants were asked to indicate their cooperative behavior (i.e., practicing via headphones) from 0 to 5 days.

Perceived realness of the game scenario (at T5). Three items assessed how well participants could imagine themselves being a pianist practicing for a concert. An example item is: “During the game, the scenario felt very real to me.”

Supposed goal of the study and credibility of the cover story (at T5). In an open question, participants were asked to express their thoughts on the study’s goals. Credibility of the cover story was first assessed via an open question on perceiving anything as odd during the game.

Demographics (at T5). Age, gender, German language proficiency, education, occupation, income, and political orientation on the left-right spectrum were assessed.

3. Results

Multiple comparisons were accounted for by correcting alpha error rates with Bonferroni-adjustments by the number of tests conducted on the same null hypothesis (i.e., $\alpha = .05 / [\text{number of tests}]$, Rubin, 2017). The analysis plan was preregistered (see <https://osf.io/xgucf>); any deviations are mentioned, and additional analyses are presented in a separate paragraph in each section and introduced as exploratory. Data were analyzed using R, version 4.1.3. The data that support the findings of this study are openly available in Open Science Framework at <https://doi.org/10.17605/OSF.IO/4CZ2B>.

3.1 Manipulation check

To examine whether the experimental manipulation was successful, and participants experienced the different phases within the game as significantly differently depending on the group, a mixed analysis of variance on the manipulation check was conducted first. The interaction effect of the within-subject factor time (T2 vs. T3 vs. T4 vs. T5) and the between-subject factor group (C-EG vs. D-EG) was significant, $F(3,1089) = 564.81, p < .001, \eta_p^2 = .61$. Second, through a set of planned contrasts, it was investigated whether each phase within the game was perceived as significantly differently to the one before and after, depending on the group. Three repeated measures contrasts were defined, comparing T2 to T3, T3 to T4, and T4 to T5. All contrast analyses yielded significant interaction effects between contrast and group ($ps < .001$). Hence, participants perceived each phase as significantly differently to the one before and after, depending on their group.

3.2 Differences in rates of change

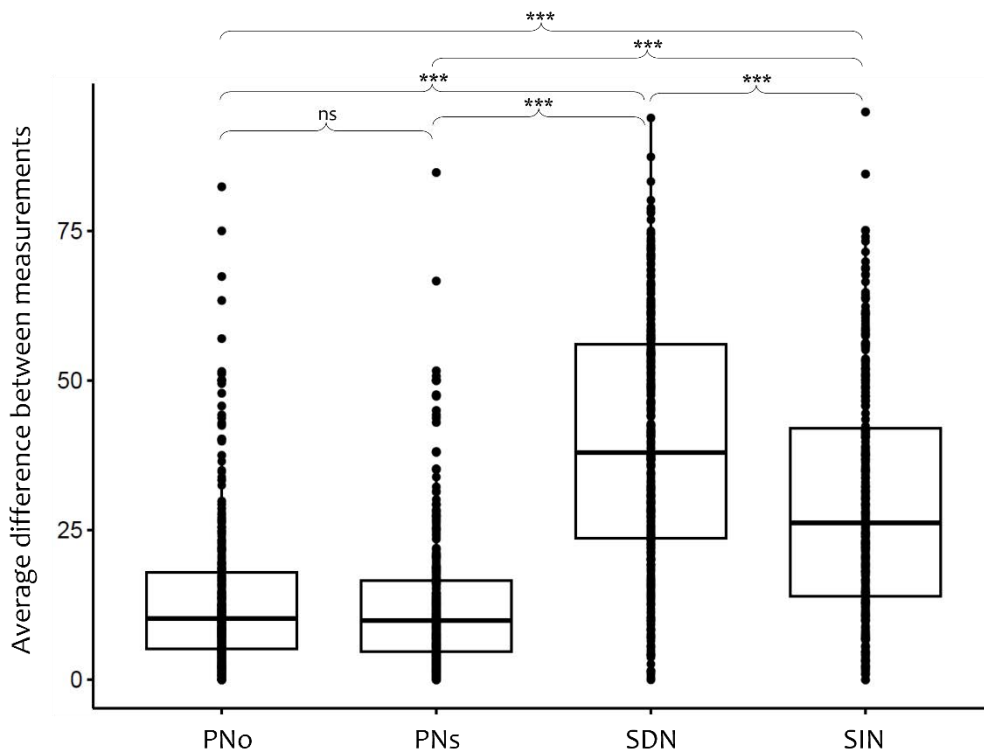
To test for different rates of change between social and personal norms, difference variables for each norm (i.e., self-oriented personal norms, other-oriented personal norms, social descriptive norms, and social injunctive norms) between single consecutive measurement time points were calculated and aggregated across time. Using a multilevel model approach, a repeated measures orthogonal contrast [1 1 -1 -1] for the type of norm was defined, comparing both social norms to both personal norms regarding their differences between measurements. The orthogonal contrast revealed that social norms changed significantly more between measurements than

personal norms ($B = 10.70$, $t(1094) = 28.23$, $p < .001$, $r = .65$), supporting Hypothesis 1.1.³ Hence, social norms showed a higher rate of change than personal norms.

Exploratory, pairwise comparisons (t -tests with Bonferroni adjustment) between the difference variables of all four types of norms were conducted, illustrated in Figure 3. Apart from self-oriented and other-oriented personal norms ($p = 1$), all post hoc tests resulted to be significant ($ps < .001$) with social descriptive norms changing the fastest.

Figure 3

Average differences between measurements in personal and social norms



Note. Difference variables are averaged across differences between consecutive measurements (i.e., T1 to T2, T2 to T3, T3 to T4, and T4 to T5). PNo = other-oriented personal norm; PNs = self-oriented personal norm; SDN = social descriptive norm; SIN = social injunctive norm.

*** $p < .001$.

3.3 Seasonal change in social norms vs. linear change in personal norms

Next, seasonal change in social norms and linear change in personal norms was investigated, using a multilevel model approach. For each mixed model, two factors were defined: (1) the

³ Upon further consideration, the preregistered first step in the analysis, meaning the analysis of variance testing for a main effect of type of norm, functioning as an omnibus test, was dropped due to no additional explanatory value.

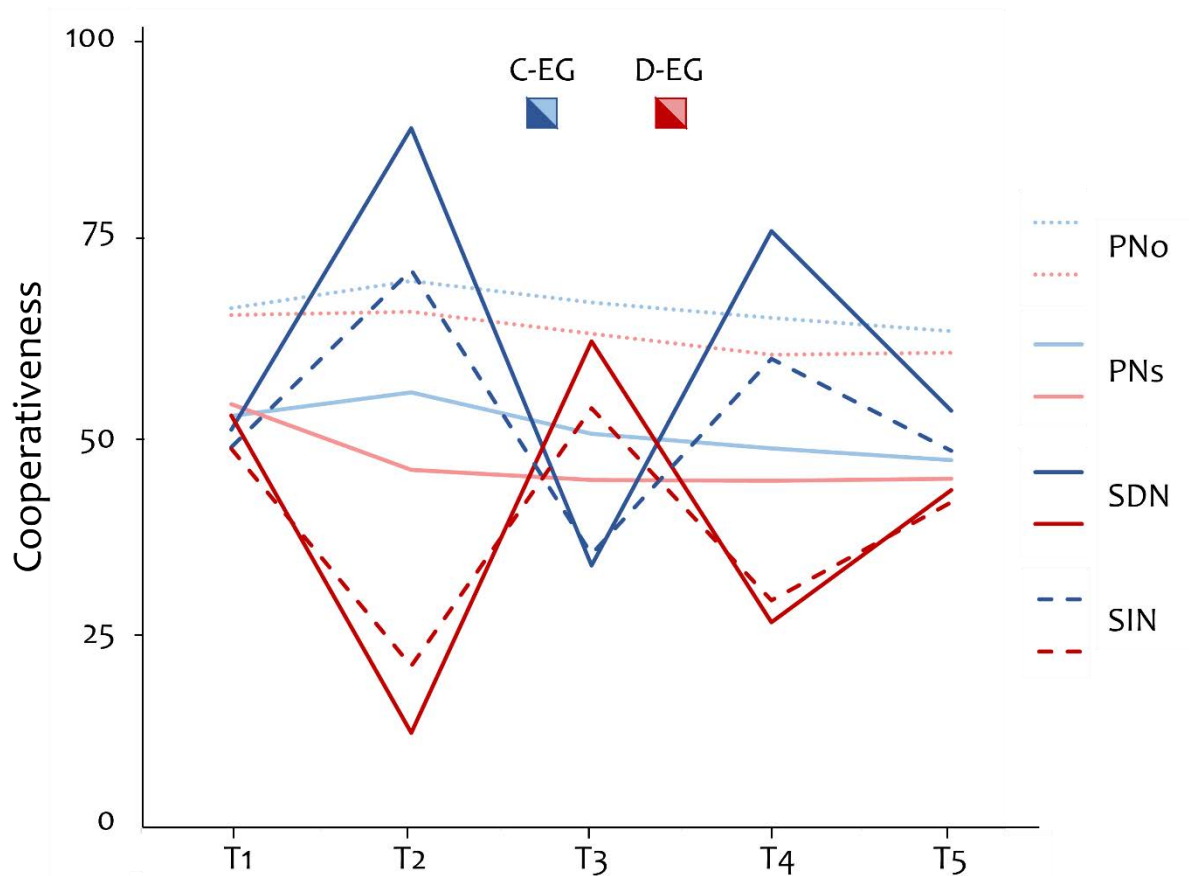
between-subject factor group (C-EG vs. D-EG) and (2) a contrast for the within-subject factor time (T1 vs. T2 vs. T3 vs. T4 vs. T5), describing the assumed development.

First, it was tested whether social norms changed seasonally [-1 3 -4 3 -1] with the different phases of the game depending on the group, by defining separate models for each social descriptive and social injunctive norms. The interaction effect of seasonal contrast and group showed to be significant for both social descriptive norms ($B = -6.63$, $t(1458) = -36.02$, $p < .001$, $r = -.69$) and social injunctive norms ($B = -4.24$, $t(1458) = -22.87$, $p < .001$, $r = -.51$), confirming Hypothesis 1.2 that social norms change seasonally with the social setting (see Appendix E1).

For change in personal norms, a linear trend contrast was set for the factor time [-2 -1 0 1 2]. Again, two separate models for the dependent variables self-oriented and other-oriented personal norms were defined. The interaction effect of the linear trend contrast and the group showed to be non-significant for both self-oriented personal norms ($B = 0.10$, $t(1458) = 0.37$, $p = .713$, $r = .01$) and other-oriented personal norms ($B = 0.22$, $t(1458) = 0.79$, $p = .428$, $r = .02$). However, there was a significant main effect of the linear trend in both self-oriented personal norms ($B = -1.90$, $t(1458) = -7.17$, $p < .001$, $r = -.18$) and other-oriented personal norms ($B = -1.26$, $t(1458) = -4.57$, $p < .001$, $r = -.12$). Results are presented in Appendix E2. Hence, the assumed linear trend in personal norms depending on the group was not confirmed (Hypothesis 1.3). Personal norms rather decreased linearly in both groups. Figure 4 shows the change over time in all four types of norms.

Figure 4

Changes in the cooperativeness of social and personal norms depending on the group



Note. T1 – T5 = measurement time points; C-EG = cooperative group; D-EG = defective group; PNo = other-oriented personal norms; PNs = self-oriented personal norms; SDN = social descriptive norms; SIN = social injunctive norms.

Exploratory, paired t-tests (two-sided) were used to investigate differences between T2 and T4 in both social norms within each group. All four tests resulted to be significant, showing a decay in the amplitude from the second to fourth measurement time point ($ps < .001$). Moreover, group differences in social norms at T5 were examined (two-sided t-tests). Although the preceding distraction phase, was supposed to “neutralize” social norms, both social norms still differed at T5 ($ps < .001$).

3.4 Predictors of social and personal norms

Using multiple regression analysis, it was investigated whether social norms were solely explained by the experimental group (see Appendix F1), whereas personal norms were additionally explained by trait cooperativeness and the interaction term (see Appendix F2). As predicted in Hypothesis 2.1, social injunctive and social descriptive norms were solely predicted

by the experimental group ($ps < .001$). Contrary to Hypothesis 2.2, self- and other-oriented personal norms were neither explained by the group, nor trait cooperativeness or the interaction.

Exploratory, regression analyses on self-oriented and other-oriented personal norms with trait cooperativeness, the experimental group, self-oriented, and other-oriented social norms as predictors were conducted (see Appendix F3). By including social norms as predictors into the analysis, both trait cooperativeness and the group in addition to all social norms showed to significantly predict self-oriented personal norms. Self-oriented social descriptive norms were the strongest predictor of self-oriented personal norms. Other-oriented personal norms were solely predicted by self-oriented social injunctive norms.

3.5 Predictors of behavior

Testing the influence of personal norms on the follow-up behavior (Hypothesis 3.1), a regression analysis with the social and personal norms as predictors was conducted (see Table 1).⁴ Solely self-oriented personal norms significantly predicted behavior after the game.

Table 1

Regression of follow-up behavior on social and personal norms

	R^2_{adj}	B	β	t	F	p
Model	.35				50.09	< .001***
Social descriptive norm		0.01	.07	1.35		.178
Social injunctive norm		-0.00	-.00	-0.06		.950
Self-oriented personal norm		0.04	.58	12.20		< .001***
Other-oriented personal norm		0.00	.02	0.53		.593

Note. $N = 365$.

* $p < .05$. *** $p < .001$.

3.6 Group differences in personal norms

Addressing Hypothesis 4.1 that personal norms are more cooperative in the cooperative than defective group after the game, personal norms at T5 were compared between groups. Descriptively, self-oriented personal norms at T5 were slightly more cooperative in the cooperative ($M_{C-EG} = 46.05$, $SD_{C-EG} = 26.25$) than the defective group ($M_{D-EG} = 43.70$, SD_{D-EG}

⁴ Although preregistered, the interaction terms were not included in the analyses to prevent overfitting.

= 26.68), while the difference did not result to be of significance in a one-tailed Welch's t-test ($t(362) = 0.85, p = .199, \delta = 0.09$). Similarly, the descriptive difference between groups regarding other-oriented personal norms ($M_{C-EG} = 62.16, SD_{C-EG} = 24.23; M_{D-EG} = 59.47, SD_{D-EG} = 26.82$) was not significant ($t(363) = 1.01, p = .157, \delta = 0.11$).

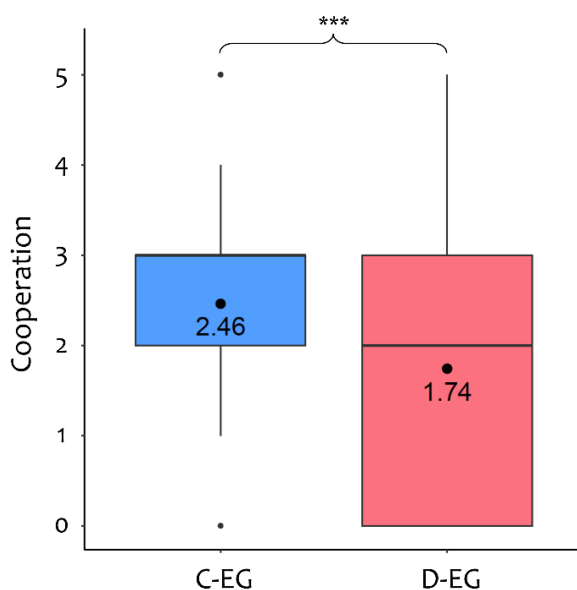
Exploratory, group differences in personal norms throughout the game were investigated, using one-tailed t-tests (see Appendix G). Self-oriented personal norms were significantly more cooperative in the cooperative than defective group at T2 ($p < .001, \delta = 0.33$). Other-oriented personal norms did not differ between groups at any point in time. Additionally, differences in personal norms within each group before and after the game (i.e., between T1 and T5) were explored in paired, two-sided t-tests. After correcting for multiple comparisons, self-oriented personal norms decreased throughout the game in both groups ($ps < .01$), while other-oriented personal norms missed the corrected alpha level ($p_{C-EG} = .065, p_{D-EG} = .021$).

3.7 Group differences in behavioral decisions

As predicted in Hypothesis 4.2, participants in the cooperative group cooperated significantly more in the follow-up behavior after the game than those in the defective group, shown in a significant one-tailed Welch's t-test ($t(362) = 4.53, p < .001, \delta = 0.47$), as indicated in Figure 5.

Figure 5

Cooperation in the follow-up behavior depending on the group



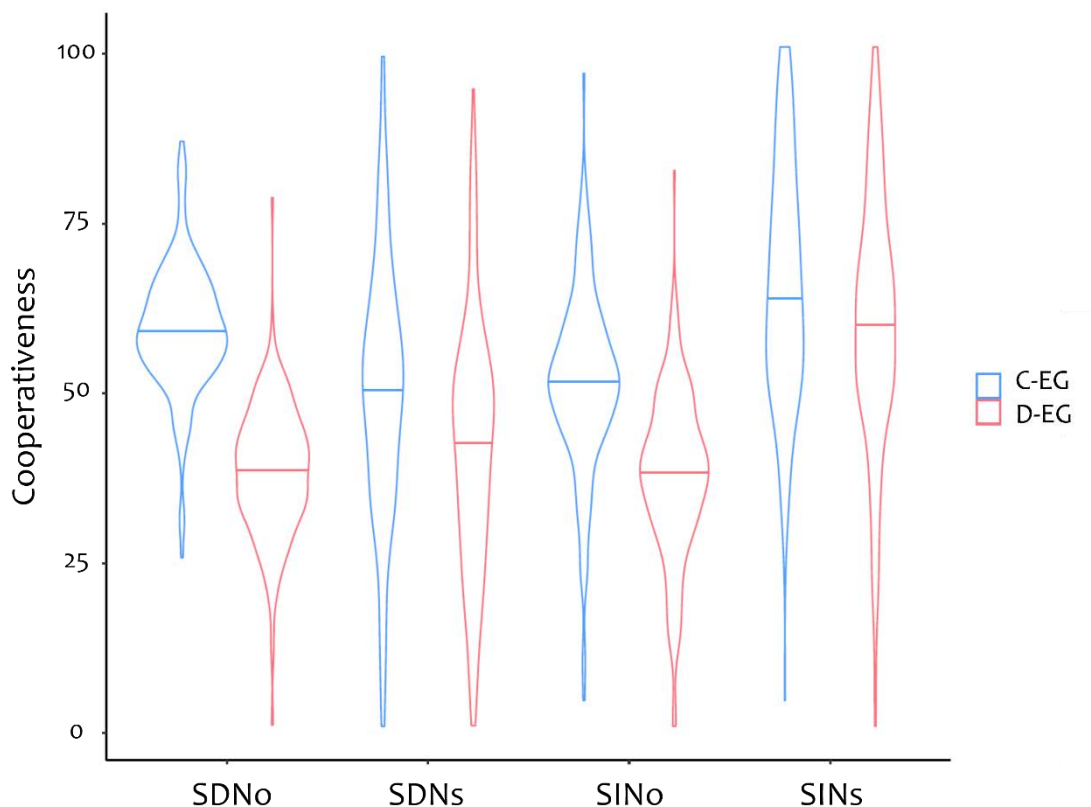
Note. Group medians are indicated by the black lines, means by the black points and their values given underneath. Significance is calculated by a one-tailed Welch's t-test. C-EG = cooperative group; D-EG = defective group. *** $p < .001$.

3.8 Self-oriented vs. other-oriented social norms

Exploratory, differences between self-oriented social norms and other-oriented social norms were analyzed. Figure 6 shows the cooperativeness of different types of social norms, indicating that self-oriented social norms, i.e., what others expect I do or approve of, have a greater variability. Using Levene tests, differences in variances were examined in six pairwise comparisons. Variances of all self-oriented social norms differed significantly from those of other-oriented social norms ($ps < .001$), whereas the variances of the two social norms of the same orientation did not differ. To further investigate the assumption that self-oriented social norms are stronger subject to subjectivity, the different social norms were correlated with trait cooperativeness, resulting to be significant only for self-oriented social descriptive norms ($r = .25, p < .001$).

Figure 6

Cooperativeness of different types of social norms split by experimental group



Note. C-EG = cooperative group, D-EG = defective group. SDNo = other-oriented social descriptive norm, SDNs = self-oriented social descriptive norm, SIno = other-oriented social injunctive norm, SInS = self-oriented social injunctive norm.

4. Discussion

The present work aimed at contributing to a better understanding of change in individuals' norms. Norms have been of great interest to social psychologists and many others; however, dynamic norm processes are yet little understood (Andrighetto & Vriens, 2022; van Kleef et al., 2019). In the present work, an experimental setting that allowed investigating differences in the temporal dynamics of social and personal norm change was introduced and assumptions on differences in norm change were addressed. Social norms were assumed to be adapted quickly whenever the social situation changes, while personal norms were expected to change more slowly and gradually, depending not only on situational but also on personal factors (cf. Batzke & Ernst, 2023). To investigate the assumed differences, participants played a repeated social dilemma game with artificial co-players. Therein, the social situation changed repeatedly *within* each experimental group, which was assumed to result in participants quickly adapting their social norms. Moreover, the overall cooperativity of the social situation differed *between* groups, which was to show in slow adaptations in participants' personal norms. Group differences in personal norms were assumed to affect behavioral decision-making. In the following, results relating to the assumed underlying qualitatively different processes of different norms (Section 4.1), influences on and of personal norms on decision-making (Section 4.2) as well as limitations and future work concerning the questions of whether and how personal norms change (Section 4.3) are discussed. Section 4.4 concludes.

4.1 Qualitatively different processes in social and personal norms and their influence on decision-making

Based on the experimental results, one may assume different temporal dynamics in social and personal norms, potentially indicating qualitatively different processes of norm change. As predicted, social descriptive and social injunctive norms had higher rates of change than personal norms (supporting Hypothesis 1.1). They were adapted repeatedly, according to the social setting showing in seasonal changes (supporting Hypothesis 1.2). Moreover, social norms were solely explained by the experimental group (supporting Hypothesis 2.1). Hence, hypotheses regarding social norms were largely supported, suggesting that changes in the social environment were observed and accounted for immediately (Nolan et al., 2008; Schultz et al., 2007). The presented results additionally led to suggest that social descriptive and social injunctive norms can in fact change repeatedly within a short timeframe, being strongly context dependent. One may thus conclude that social norm adaptation is fast and quickly reversible.

Unlike social norms, personal norms were assumed to change linearly, developing towards cooperativeness or defectivity, depending on the group, which could not be supported (contradicting Hypothesis 1.3). The results rather indicated that personal norms trended towards defectivity in both groups. Similarly, results from Szekely et al. (2021) descriptively showed a trend of decreasing personal contribution norms over time, being 28 days respectively (see also Tverskoi et al., 2023). In the present work, a group difference was only found in self-oriented personal norms at T2 with a small-medium sized effect. While the group difference descriptively remained until the end of the game, significance disappeared after T2 and effect sizes decreased (see Appendix G), not supporting the expected increase in the group difference throughout the game (contradicting Hypothesis 4.1). There are different interpretations for these results that are discussed in Section 4.3.

There is an ongoing debate, whether social norms explain variance in behavioral decisions over and above personal norms (e.g., Biel & Thøgersen, 2007). In line with the present results, studies showed that when including personal norms in the analysis, the influence of social norms diminished (Hopper & Nielsen, 1991; Thøgersen, 1999). Interestingly, the present data also suggested a decrease in the sensitivity of adapting social norms by a decay in the amplitude from the second to fourth measurement time point. Based on that, it may be assumed that people are particularly attentive to social norms when being introduced to a new situation. Their influence might decrease over time with personal norms becoming increasingly important for decision-making.

4.2 Influences on and of personal norms on cooperation

Regarding influencing factors on personal norms, psychological research so far has identified several associated factors such as social norms, ascription of responsibility, problem awareness, guilt, etc. (see Bamberg & Schmidt, 2003; De Groot & Steg, 2009; Stern et al., 1999). In the present work, it was assumed that situational and personal variables as well as their interaction affect personal norm change. Neither the experimental group, nor trait cooperativeness, nor their interaction resulted to be of significance in the preregistered analysis (contradicting Hypothesis 2.2). However, exploratory analyses revealed that the influence of both predictors might have been masked. When adding social norms to the analysis, the group, the personality factor trait cooperativeness as well as different types of social norms explained variance in self-oriented personal norm. This supports the assumption of situational and personal factors influencing personal norm change and the idea of an underlying more complex process (compared to social norms change).

While existing research showed that (other-oriented) social norms and personal norms are associated (e.g., Bamberg, 2013; Bamberg et al., 2007), the present results suggest that variance in personal norms is particularly well predicted by *self-oriented* social norms. Hence, what an individual perceives as expectations of others regarding the *own* behavior (vs. general behaviors of others) is strongly related to the individual's personal norms. Moreover, personal norms (as well as self-oriented social norms) showed to be related to trait cooperativeness. Based on these results personal norms cannot be assumed to merely result from a general cooperative personality and lack “social conditionality”, as stated by Bicchieri and Dimant (2019), defining a “moral rule” as *not* depending “on others doing X [i.e., a behavior] or thinking that you should do X.” (p. 447f, text in square brackets added). Rather the idea finds support that personal norms are learned expectations of others in between purely external social factors (such as other-oriented social descriptive norms, i.e., what one believes others do) and internal personality factors. This relates to the concept of internalization of social norms into personal norms, refined in the extended norm taxonomy by Thøgersen (2006).

Accordingly, self-oriented social norms might be in-between personal and other-oriented social norms as they were less univocally given by the situation than other-oriented social norms with their variances being significantly larger. Thus, self-oriented social norms can be assumed to be more determined by subjectivity as they partly correlated with trait cooperativeness. Since they were strongly related to personal norms, others' expectations regarding the individual (and not people in general) could be a promising leverage point for future norm-based intervention studies. So far, norm-based intervention research has focused on *other-oriented* social norms under the terms of *social norm information and feedback* (Abrahamse & Steg, 2013), *social norms marketing* (Miller & Prentice, 2016) or *social norm nudges* (Sunstein, 2014).

Looking at influences of personal norms on cooperation, the follow-up behavior was predicted by self-oriented personal norms (but not other-oriented personal norms, only partly supporting Hypothesis 3.1). In line with existing research, personal norms showed to be highly relevant for behavioral decisions (cf. Han, 2014; Hunecke et al., 2001; Onwezen et al., 2013). As expected, the present results also showed a group difference in the follow-up behavior (after the game) with participants in the cooperative group cooperating significantly more (supporting Hypothesis 4.2). However, as the group difference in self-oriented personal norms was lost at T5, it remains unclear whether the change in personal norms or in social norms affected the behavior change. The distraction phase at the end of the game was supposed to “neutralize”

social norms to an intermediate level, being identical for both groups. However, social norms still differed at T5.

Throughout all analyses other-oriented personal norms (i.e., an individual's beliefs about what *others* should do) were less affected by other variables and had close to no effects on behavioral decision-making. Contrary, self-oriented personal norms (i.e., an individual's beliefs about what *itself* should do) were strongly predictive of behavior and at least temporarily affected by the experimental manipulation. The missing specification on self-oriented personal norms is one possible explanation, why previous research on norm change assessing personal norms as a self-/other-combination found no effects on personal norms (Bicchieri et al., 2022; Szekely et al., 2021).

4.3 Do personal norms change, and how? – Limitations and future work

While social norms did change with the social setting, the more interesting questions remained inconclusive: Do personal norms change, and how? There are two possible answers to that question. On the one hand, one may assume that personal norms remain rather stable throughout the lifetime with major changes largely happening in childhood and early adulthood (cf. Nucci, 2001; Turiel, 1983). Based on this assumption, the shown group difference in personal norms could be attributed to differences in the activation level, in line with Schwartz' (1977) *norm activation model* (see also Schwartz & Howard, 1981, 1982) and its extension, the *belief-value-norm theory* (Stern et al., 1999; Stern, 2000). Accordingly, self-oriented personal norms might have been activated by the respective social norm in the first phase of the game. The differences showed at T2, with personal norms being activated towards cooperation in the cooperative group that just experienced a cooperative setting and vice versa for the defective group. Furthermore, one could assume that people got frustrated after T2, which deactivated their personal norms, showing in a tendency towards central neutrality, indicating indifference. Being one of the most influential psychological theories on personal norms, the norm activation model has incited numerous studies on influencing factors of personal norm activation (Bamberg et al., 2007; Han, 2014; Hunecke et al., 2001; Klöckner & Matthies, 2004). Yet, to our knowledge none of them addressed personal norm change as the theory is purely static, describing situational activation, not accounting for the possibility of belief change. But does that mean that personal norms are in fact static entities similar to traits? Research has shown that people are highly adaptive, learn throughout their life, change their attitudes, values, and even personality traits (Bardi et al., 2009; Bleidorn et al., 2021; Otto & Kaiser, 2014).

This leads us to the second possible answer: Personal norms do change – presumably over longer periods of time like change in attitudes or values. This supports the presented assumption of a slower adaptation process and calls for long-term studies of personal norms change. However, assuming slow change does still allow for observing an excerpt of personal norm change in shorter periods of time. Although it suggests that assessing change empirically may be challenging, as one is looking for small effects.

There are numerous potential reasons why the present study was unsuccessful in showing lasting change in personal norms. Possibly, the manipulation was too weak. As differences did show descriptively between groups, one could assume that stronger manipulations would increase the effect. In future work, a more existential game scenario than practicing the piano for a concert could be applied. Also, the personal norms measures might not have been sensitive to the induced change. Indicating the own personal norms via self-report requires some level of introspection, and, doing so repeatedly within a short amount of time (game duration was about 20 minutes on average), requires a great amount of compliance. More indirect measures might improve validity, for instance giving participants the option to lie about a coin toss that occurred in complete anonymity and is supposed to decide about the appropriateness of a behavior, as proposed by Bicchieri et al. (2014).

While all the above may be (partly) accurate, why did the cooperative group not show a development towards more cooperativeness in personal norms? Possibly, social dilemma games make learning cooperative personal norms difficult, as contextual variables are limited to a minimum. In the complexity of the real world, early stages of learning cooperative personal norms may be accompanied by attributing behavioral decisions to situational cues before a personal norm is generalized across single situations. Situational cues are however limited in the simplified social dilemma situation. This could also explain the descriptive decrease in the cooperativeness of personal norms found by Szekely et al. (2021), where participants played a collective-risk social dilemma. Dilemma research also showed that there is a general effect of decrease in contribution/cooperation levels over time (e.g., Dal Bó & Dal Bó, 2014), which might mirror erosion of personal norms.

Still, none of these potential reasons may explain the change in personal norms that participants showed in the cooperative group, increasing in cooperativeness to T2 and decreasing thereafter. An interesting and quite plausible explanation relates to *prospect theory* (Tversky & Kahneman, 1992), stating that negative experiences have a stronger impact than positive experiences. Therein, the authors described an asymmetry between losses and gains, stating that “losses loom larger than gains” (p. 298). Accordingly, the present results showed

that personal norms in the cooperative group developed towards defectivity only *after* T2 (i.e., after the first encounter with defection), showing an asymmetrical development in self-oriented personal norms compared to the defective group. Experiencing a defective setting (i.e., having made a negative experience) might have eroded participants' personal norms – qualitatively differently to the positive impact of the prior cooperative setting (i.e., a positive experience). To test that assumption, an experimental group in which participants experience a longer cooperative or even purely cooperative setting would be necessary. So far, it remains unclear whether personal norms change is due to belief change or change in the activation level and how it can be directed towards more cooperativeness in the long run.

4.4 Conclusion

The present paper demonstrated an experimental approach to studying differences in the temporal dynamics of norm change processes. Assumptions were tested concerning the temporal dynamics of social and personal norm change. Results led to assume that social and personal norms change differently – faster and slower. While the fast change in social norms was well predicted by situational changes, slow change in personal norms was multidetermined.

The present work aimed at taking a step towards better understanding norm change. Being able to truly grasp the potentials and limitations of norms in behavioral change, calls for knowledge about how, when, and why norms change. Yet, many questions particularly regarding personal norm change remain still open, among them: When do personal norms change, of what kind are the underlying mechanisms, how could results be used to foster learning cooperative personal norms, among others. These questions may inspire further research. Addressing them seems particularly relevant, as the significance of personal norms for behavioral decisions was demonstrated once again.

References

- Abrahamse, W., & Steg, L. (2013). Social influence approaches to encourage resource conservation: A meta-analysis. *Global Environmental Change, 23*(6), 1773-1785. <https://doi.org/10.1016/j.gloenvcha.2013.07.029>
- Anderson, J. E., & Dunning, D. (2014). Behavioral norms: Variants and their identification. *Social and Personality Psychology Compass, 8*(12), 721-738. <https://doi.org/10.1111/spc3.12146>
- Andrighetto, G., Grieco, D., & Tummolini, L. (2015). Perceived legitimacy of normative expectations motivates compliance with social norms when nobody is watching. *Frontiers in Psychology, 6*, 1413. <https://doi.org/10.3389/fpsyg.2015.01413>
- Andrighetto, G., & Vriens, E. (2022). A research agenda for the study of social norm change. *Philosophical Transactions of the Royal Society A, 380*(2227), 20200411. <https://doi.org/10.1098/rsta.2020.0411>
- Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and Applied, 70*(9), 1. <https://doi.org/10.1037/h0093718>
- Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.
- Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review, 80*(4), 1095-1111. <https://doi.org/10.2307/1960858>
- Bamberg, S. (2013). Changing environmentally harmful behaviors: A stage model of self-regulated behavioral change. *Journal of Environmental Psychology, 34*, 151-159. <https://doi.org/10.1016/j.jenvp.2013.01.002>
- Bamberg, S., Hunecke, M., & Blöbaum, A. (2007). Social context, personal norms and the use of public transportation: Two field studies. *Journal of Environmental Psychology, 27*(3), 190–203. <https://doi.org/10.1016/j.jenvp.2007.04.001>
- Bamberg, S., & Schmidt, P. (2003). Incentives, morality, or habit? Predicting students' car use for university routes with the models of Ajzen, Schwartz, and Triandis. *Environment and Behavior, 35*(2), 264-285.
- Bardi, A., Lee, J. A., Hofmann-Towfigh, N., & Soutar, G. (2009). The structure of intraindividual value change. *Journal of Personality and Social Psychology, 97*(5), 913–929. <https://doi.org/10.1037/a0016617>
- [dataset] Batzke, M. C. L. (2023). Fast and slow adaptation of norms. Open Science Framework. <https://doi.org/10.17605/OSF.IO/4CZ2B>
- Batzke, M. C. L., & Ernst, A. (2022). Explaining and resolving norm-behavior inconsistencies – A theoretical agent-based model. In M. Czupryna & B. Kamiński (Eds.), *Advances in Social Simulation* (pp. 41–52). Springer. https://doi.org/10.1007/978-3-030-92843-8_4

- Batzke, M. C. L., & Ernst, A. (2023). Conditions and Effects of Norm Internalization. *Journal of Artificial Societies and Social Simulation*, 26(1), 1–31. <https://doi.org/10.18564/jasss.5003>
- Bicchieri, C., & Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public Choice*, 1-22. <https://doi.org/10.1007/s11127-019-00684-6>
- Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132, 59-72. <https://doi.org/10.1016/j.geb.2021.11.012>
- Bicchieri, C., Dimant, E., Gelfand, M., & Sonderegger, S. (2023). Social norms and behavior change: The interdisciplinary research frontier. *Journal of Economic Behavior & Organization*, 205, A4-A7. <https://doi.org/10.1016/j.jebo.2022.11.007>
- Bicchieri, C., Lindemans, J. W., & Jiang, T. (2014). A structured approach to a diagnostic of collective practices. *Frontiers in Psychology*, 5, 1418. <https://doi.org/10.3389/fpsyg.2014.01418>
- Bicchieri, C., Muldoon, R., & Sontuoso, A. (2018). Social Norms. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2018/entries/social-norms/>
- Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2), 191-208. <https://doi.org/10.1002/bdm.621>
- Biel, A., & Thøgersen, J. (2007). Activation of social norms in social dilemmas: A review of the evidence and reflections on the implications for environmental behaviour. *Journal of Economic Psychology*, 28(1), 93-112. <https://doi.org/10.1016/j.joep.2006.03.003>
- Bleidorn, W., Hopwood, C. J., Back, M. D., Denissen, J. J. A., Hennecke, M., Hill, P. L., Jokela, M., Kandler, C., Lucas, R. E., Luhmann, M., Orth, U., Roberts, B. W., Wagner, J., Wrzus, C., & Zimmermann, J. (2021). Personality trait stability and change. *Personality Science*, 2, 1-20. <https://doi.org/10.5964/ps.6009>
- Castelfranchi, C., Dignum, F., Jonker, C. M., & Treur, J. (2000). Deliberative normative agents: Principles and architecture. In N. R. Jennings & Y. Lesperance (Eds.), *Intelligent Agents VI. Agent Theories, Architectures, and Languages* (Vol. 1757, pp. 364–378). Springer. https://doi.org/10.1007/10719619_27
- Chudek, M., & Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences*, 15(5), 218-226. <https://doi.org/10.1016/j.tics.2011.03.003>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591-621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>

- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015-1026. <https://doi.org/10.1037/0022-3514.58.6.1015>
- Conner, M., & Armitage, C. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology*, 28(15), 1429–1464. <https://doi.org/10.1111/j.1559-1816.1998.tb01685.x>
- Conte, R., & Castelfranchi, C. (1995). Understanding the functions of norms in social groups through simulation. In N. Gilbert & R. Conte (Eds.), *Artificial Societies: The Computer Simulation of Social Life* (pp. 213–226). Routledge.
- Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, 17(8), 363-366. <https://doi.org/10.1016/j.tics.2013.06.005>
- Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, 17(3), 273-292. <https://doi.org/10.1177/1088868313495594>
- Cushman, F., Kumar, V., & Railton, P. (2017). Moral learning: Psychological and philosophical perspectives. *Cognition*, 167, 1-10. <https://doi.org/10.1016/j.cognition.2017.06.008>
- Dal Bó, E., & Dal Bó, P. (2014). “Do the right thing:” the effects of moral suasion on cooperation. *Journal of Public Economics*, 117, 28-38. <https://doi.org/10.1016/j.jpubeco.2014.05.002>
- Dannals, J. E., & Miller, D. T. (2017). Social norm perception in groups with outliers. *Journal of Experimental Psychology: General*, 146(9), 1342. <https://doi.org/10.1037/xge0000336>
- Dannals, J. E., Halali, E., Kopelman, S., & Halevy, N. (2022). Power, constraint, and cooperation in groups: The role of communication. *Journal of Experimental Social Psychology*, 100, 104283. <https://doi.org/10.1016/j.jesp.2022.104283>
- Dannenberg, A., Gutsche, G., Batzke, M. C. L., Christens, S., Engler, D., Mankat, F., Möller, S., Weingärtner, E., Ernst, A., Lumkowsky, M., von Wangenheim, G., Hornung, G., & Ziegler, A. (in press). The effects and dynamics of norms on environmentally relevant behavior. *Review of Environmental Economics and Policy*.
- Dawes, R.M. (1980). Social dilemmas. *Annual Review of Psychology*, 31(1), 169-193. <https://doi.org/10.1146/annurev.ps.31.020180.001125>
- De Groot, J. I., & Steg, L. (2009). Morality and prosocial behavior: The role of awareness, responsibility, and norms in the norm activation model. *The Journal of Social Psychology*, 149(4), 425-449. <https://doi.org/10.3200/SOCP.149.4.425-449>
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, 51(3), 629. <https://doi.org/10.1037/h0046408>

Dignum, F. (1999). Autonomous agents with norms. *Artificial Intelligence and Law*, 7, 69-79. <https://doi.org/10.1023/A:1008315530323>

Farrow, K., Grolleau, G., & Ibanez, L. (2017). Social norms and pro-environmental behavior: A review of the evidence. *Ecological Economics*, 140, 1-13. <https://doi.org/10.1016/j.ecolecon.2017.04.017>

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*, 20(3), 393-398. <https://doi.org/10.1111/j.1467-9280.2009.02306.x>

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3), 472-482. <https://doi.org/10.1086/586910>

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814. <https://doi.org/10.1037/0033-295X.108.4.814>

Han, H. (2014). The norm activation model and theory-broadening: Individuals' decision-making on environmentally-responsible convention attendance. *Journal of Environmental Psychology*, 40, 462-471. <https://doi.org/10.1016/j.jenvp.2014.10.006>

Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859), 1243-1248. <https://doi.org/10.1126/science.162.3859.1243>

Harland, P., Staats, H., & Wilke, H. A. (1999). Explaining proenvironmental intention and behavior by personal norms and the Theory of Planned Behavior. *Journal of Applied Social Psychology*, 29(12), 2505-2528. <https://doi.org/10.1111/j.1559-1816.1999.tb00123.x>

Hawkins, R. X., Goodman, N. D., & Goldstone, R. L. (2019). The emergence of social norms and conventions. *Trends in Cognitive Sciences*, 23(2), 158-169. <https://doi.org/10.1016/j.tics.2018.11.003>

Hopper, J. R., & Nielsen, J. M. (1991). Recycling as altruistic behavior: Normative and behavioral strategies to expand participation in a community recycling program. *Environment and Behavior*, 23(2), 195-220. <https://doi.org/10.1177/0013916591232004>

Howard, J. A., & Renfrow, D. G. (2003). Social cognition. In J. Delamater (Ed.), *Handbook of Social Psychology* (pp. 259-281). Kluwer Academic/Plenum Publishers.

Hunecke, M., Blöbaum, A., Matthies, E., & Höger, R. (2001). Responsibility and environment: Ecological norm orientation and external factors in the domain of travel mode choice behavior. *Environment and Behavior*, 33(6), 830-852. <https://doi.org/10.1177/00139160121973269>

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science*, 322(5908), 1681-1685. <https://doi.org/10.1126/science.1161405>

Kelly, D., & Davis, T. (2018). Social norms and human normative psychology. *Social Philosophy and Policy*, 35(1), 54-76. <https://doi.org/10.1017/S0265052518000122>

- Klößner, C. A., & Matthies, E. (2004). How habits interfere with norm-directed behaviour: A normative decision-making model for travel mode choice. *Journal of Environmental Psychology, 24*(3), 319-327. <https://doi.org/10.1016/j.jenvp.2004.08.004>
- Kohlberg, L. (1964). Development of moral character and moral ideology. In M. Hoffman & L. W. Hoffman (Eds.), *Review of Research in Child Development* (Vol. 1, pp. 383-431). Russell Sage Foundation.
- Leiner, D. J. (2019). SoSci Survey (Version 3.4.17) [Computer software]. Available at <https://www.soscisurvey.de>
- McDonald, R. I., & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences, 3*, 147-151. <https://doi.org/10.1016/j.cobeha.2015.04.006>
- Miller, D. T., & Prentice, D. A. (2016). Changing norms to change behavior. *Annual Review of Psychology, 67*, 339-361. <https://doi.org/10.1146/annurev-psych-010814-015013>
- Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making, 6*(8), 771-781. <https://doi.org/10.1017/S1930297500004204>
- Nakashima, N. A., Halali, E., & Halevy, N. (2017). Third parties promote cooperative norms in repeated interactions. *Journal of Experimental Social Psychology, 68*, 212-223. <https://doi.org/10.1016/j.jesp.2016.06.007>
- Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin, 34*(7), 913-923. <https://doi.org/10.1177/0146167208316691>
- Nucci, L. P. (2001). *Education in the Moral Domain*. Cambridge University Press.
- Nyborg, K. (2018). Social norms and the environment. *Annual Review of Resource Economics, 10*, 405-423. <https://doi.org/10.1146/annurev-resource-100517-023232>
- Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W.N., Arrow, K. J., Barrett, S., Carpenter, S., Chapin III, F. S., Crépin, A.-S., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S. A., Madsen, O. J., ... De Zeeuw, A. (2016). Social norms as solutions. *Science, 354*(6308), 42-43. <https://doi.org/10.1126/science.aaf8317>
- Onwezen, M. C., Antonides, G., & Bartels, J. (2013). The Norm Activation Model: An exploration of the functions of anticipated pride and guilt in pro-environmental behaviour. *Journal of Economic Psychology, 39*, 141-153. <https://doi.org/10.1016/j.joep.2013.07.005>
- Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives, 14*(3), 137-158. <http://www.jstor.org/stable/2646923>

Otto, I. M., Donges, J. F., Cremades, R., Bhowmik, A., Hewitt, R. J., Lucht, W., Rockström, J., Allerberger, F., McCaffrey, M., Doe, S. S. P., Lenferna, A., Morán, N., van Vuuren, D. P., & Schellnhuber, H. J. (2020). Social tipping dynamics for stabilizing Earth's climate by 2050. *Proceedings of the National Academy of Sciences*, *117*(5), 2354-2365. <https://doi.org/10.1073/pnas.1900577117>

Otto, S., & Kaiser, F. G. (2014). Ecological behavior across the lifespan: Why environmentalism increases as people grow older. *Journal of Environmental Psychology*, *40*, 331-338. <https://doi.org/10.1016/j.jenvp.2014.08.004>

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment. *Journal of Personality and Social Psychology*, *96*, 574-587. <https://doi.org/10.1037/a0011989>

Peysakhovich, A., & Rand, D. G. (2016). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*, *62*(3), 631-647. <https://doi.org/10.1287/mnsc.2015.2168>

Piaget, J. (1970). Piaget's theory. In P. Mussen (Ed.), *Carmichaels' Manual of Child Psychology* (3rd ed., Vol. I, pp. 703-732). Wiley.

Prentice, D., & Paluck, E. L. (2020). Engineering social change using social norms: Lessons from the study of collective action. *Current Opinion in Psychology*, *35*, 138-142. <https://doi.org/10.1016/j.copsyc.2020.06.012>

Rubin, M. (2017). Do p values lose their meaning in exploratory analyses? It depends how you define the familywise error rate. *Review of General Psychology*, *21*(3), 269-275. <https://doi.org/10.1037/gpr0000123>

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54-67. <https://doi.org/10.1006/ceps.1999.1020>

Schultz, W. P., Khazian, A. M., & Zaleski, A. C. (2008). Using normative social influence to promote conservation among hotel guests. *Social Influence*, *3*(1), 4-23. <https://doi.org/10.1080/15534510701755614>

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18*(5), 429-434. <https://doi.org/10.1111/j.1467-9280.2007.01917.x>

Schwartz, S. H. (1977). Normative influences on altruism. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 221-279). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60358-5](https://doi.org/10.1016/S0065-2601(08)60358-5)

Schwartz, S., & Howard, J. (1981). A normative decision-making model of altruism. In J. Rushton (Ed.), *Altruism and Helping Behaviour: Social, Personality and Developmental Perspectives* (pp. 189-211). Lawrence Erlbaum Associates Inc.

- Schwartz, S., & Howard, J. (1982). Helping and cooperation: A self-based motivational model. In V. Derlega & J. Grzelak (Eds.), *Cooperation and Helping Behavior: Theories and Research* (p. 327–353). Academic Press.
- Sen, S., & Airiau, S. (2007). Emergence of norms through social learning. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence* (Vol. 1507, pp. 1507-1512). AAAI Press.
- Sherif, M. (1936). *The psychology of social norms*. Harper.
- Shin, Y. H., Im, J., Jung, S. E., & Severt, K. (2018). The theory of planned behavior and the norm activation model approach to consumer behavior regarding organic menus. *International Journal of Hospitality Management*, *69*, 21-29. <https://doi.org/10.1016/j.ijhm.2017.10.011>
- Steg, L., & De Groot, J. (2010). Explaining prosocial intentions: Testing causal relationships in the norm activation model. *British Journal of Social Psychology*, *49*(4), 725-743. <https://doi.org/10.1348/014466609X477745>
- Stern, P. C. (2000). New environmental theories: Toward a coherent theory of environmentally significant behavior. *Journal of Social Issues*, *56*(3), 407-424. <https://doi.org/10.1111/0022-4537.00175>
- Stern, P. C., Dietz, T., Abel, T., Guagnano, G. A., & Kalof, L. (1999). A value-belief-norm theory of support for social movements: The case of environmentalism. *Human Ecology Review*, *6*(2), 81-97. <https://www.jstor.org/stable/24707060>
- Sunstein, C. R. (2014). Nudging: a very short guide. *Journal of Consumer Policy*, *37*, 583-588. <https://doi.org/10.1007/s10603-014-9273-1>
- Szekely, A., Bruner, D., Steinmo, S., Todor, A., Volintiru, C., & Andrighetto, G. (2023). Preferences for honesty can support cooperation. *Journal of Behavioral Decision Making*, e2328. <https://doi.org/10.1002/bdm.2328>
- Szekely, A., Lipari, F., Antonioni, A., Paolucci, M., Sánchez, A., Tummolini, L., & Andrighetto, G. (2021). Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications*, *12*(1), 1-7. <https://doi.org/10.1038/s41467-021-25734-w>
- Theriault, J. E., Young, L., & Barrett, L. F. (2021). The sense of should: A biologically-based framework for modeling social pressure. *Physics of Life Reviews*, *36*, 100-136. <https://doi.org/10.1016/j.plrev.2020.01.004>
- Thøgersen, J. (1999). The ethical consumer: Moral norms and packaging choice. *Journal of Consumer Policy*, *22*(4), 439-460. <https://doi.org/10.1023/A:1006225711603>
- Thøgersen, J. (2006). Norms for environmentally responsible behaviour: An extended taxonomy. *Journal of Environmental Psychology*, *26*(4), 247-261. <https://doi.org/10.1016/j.jenvp.2006.09.004>

Turiel, E. (1983). *The Development of Social Knowledge: Morality and Convention*. Cambridge University Press.

Turner, J. C. (1987). A self-categorization theory. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher & M. S. Wetherell (Eds.), *Rediscovering the Social Group: A Self-Categorization Theory* (pp. 42–67). Basil Blackwell.

Tverskoi, D., Guido, A., Andrighetto, G., Sánchez, A., & Gavrilets, S. (2023). Disentangling material, social, and cognitive determinants of human behavior and beliefs. *Humanities and Social Sciences Communications*, 10(1), 1-13.
<https://doi.org/10.1057/s41599-023-01745-4>

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.
<https://doi.org/10.1007/BF00122574>

van Kleef, G. A., Gelfand, M. J., & Jetten, J. (2019). The dynamic nature of social norms: New perspectives on norm development, impact, violation, and enforcement. *Journal of Experimental Social Psychology*, 84, 103814. <https://doi.org/10.1016/j.jesp.2019.05.002>

Villatoro, D., Andrighetto, G., Conte, R., & Sabater-Mir, J. (2015). Self-policing through norm internalization: A cognitive solution to the tragedy of the digital commons in social networks. *Journal of Artificial Societies and Social Simulation*, 18(2), 2.
<https://doi.org/10.18564/jasss.2759>

Acknowledgements

The present work was developed within the *ZumWert* project. We acknowledge the funding of the University of Kassel in their profile building initiative. We thank Mirjam Ebersbach and Anna Helfers for stimulating discussions on the experimental design. We also thank Paula Rosendahl for her support in programming of the survey and data processing.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix to

“Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation”

Appendix A: Operationalization of experimental groups

Appendix B: Sample characteristics

Appendix C: Materials

Appendix D: Internal consistencies of norm scales at different measurement time points

Appendix E: Seasonal change in social norms and linear change in personal norms

Appendix F: Regressions of social and personal norms

Appendix G: Group differences in self-oriented and other-oriented personal norms at different measurement time points

Appendix A – Operationalization of experimental groups

		Phase 1					Phase 2				Phase 3				Distraction phase			
Round		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
C-EG	Co-player 1	C	C	C	C	C	C	D	D	D	C	C	C	C	C	C	C	C
	Co-player 3	C	C	C	C	C	D	D	D	D	D	C	C	C	C	D	D	D
D-EG	Co-player 1	D	D	D	D	D	C	C	C	C	C	D	D	D	D	C	C	C
	Co-player 3	D	D	D	D	D	D	C	C	C	D	D	D	D	D	D	D	D
		T1		T2						T3				T4				T5

Note. Participants played the social dilemma game with two artificial co-players (1 and 3), themselves being player 2. The game differed between the cooperative (C-EG) and defective experimental group (D-EG) in the order of their phases. Each experimental group consisted of three phases, characterized by either a cooperation (blue color) or defection (red color) of the co-players, and a distraction phase, characterized by a mixed setting in which one co-player cooperated and the other defected (white color). In between phases, single rounds of a mixed setting were added to make the game more realistically. Social and personal norms were assessed before the game (T1) and roughly after each phase (after rounds 3, 9, 13 and 17) at T2 – T5.

Appendix B – Sample characteristics

- N = 365
- 47% female, 52% male, 1% diverse
- Age: $M = 46$ ($SD = 16$): 21% 18 – 29 years, 17% 30 – 39 years, 18% 40 – 49 years, 20% 50 – 59 years, 20% 60 – 69 years, 5% 70 – 74 years
- Education: 27% no A-levels, 39% apprenticeship, 34% A-levels or higher
- Occupation: 12% students, 3% in apprenticeship, 52% employed, 4% self-employed, 5% unemployed, 20% retired, 4% incapacitated for work
- Income: 19% less than 1300€, 38% 1300 – 3200€, 25% 3200€ – 5000€, 9% more than 5000€ [8% missing values]
- Proficiency in German language: 4% fluent, 96% first language

Appendix C – Materials

Materials were shown to participants in German language and are below presented translated in English (originals can be found here: <https://osf.io/xgucf>). If not indicated otherwise, items were presented with a response slider ranging from 1 “not agree at all” to 101 “absolutely agree”. In addition to the below presented materials, the following measures were assessed at T5, but are not presented here, as they were not included in any analysis or hypothesis: Trustworthiness and predictability of the co-players, personal values, reactance, and strategies during the game. All materials are shown under: <https://osf.io/xgucf>.

Trait cooperativeness (at T1)

The slider measure of social value orientation (Murphy et al., 2011) was applied for the trait cooperativeness measure. Participants were asked to allocate hypothetical money to themselves (“You receive”) and to another unknown person (“The other person receives”). The money is allocated using a single slider for both allocations. The amounts of money received by the person herself and the other person are displayed above and underneath a slider, changing dynamically when the slider is moved. Participants were presented with an example and then asked to make six money allocations. The trait cooperativeness scale was created according to the instructions given by Murphy and colleagues (2011). Higher values indicate a stronger motivation to cooperate.

1. You receive:	85	85
The other person receives:	85	15
2. You receive:	85	100
The other person receives:	15	50
3. You receive:	50	85
The other person receives:	100	85
4. You receive:	50	85
The other person receives:	100	15
5. You receive:	100	50
The other person receives:	50	100
6. You receive:	100	85
The other person receives:	50	85

Comprehension of the game instructions (at T1)

1. The game is about...
 - preparing for a piano concert. (true)
 - preparing for a guitar concert. (false)
2. When I practice out loudly...
 - others are not disturbed. (false)
 - I have a harder time concentrating. (false)
 - I have a better practice experience. (true)
3. What of the following is true?
 - You play the game with four other people, with all practice rooms sharing thin walls. (false)
 - The concert is crucial for your future as a pianist. (true)
 - If it's just you playing the piano via headphones, you'll have the best practicing experience. (false)
4. What of the following is true?
 - Who plays in the practice rooms next door stays the same from day to day. (true)
 - When everyone is playing with headphones, no one can concentrate. (false)
 - If others are practicing loudly, it doesn't bother you. (false)
5. How many practice points you get...
 - is not dependent on the decisions of the other players. (false)
 - is determined each day based on the established awarding rules. (true)

Social norms (at T1 – T5)

Other-oriented social descriptive norms (at T1)

- The others will mostly play the piano loudly.
- The others will mostly play the piano via headphones.

Other-oriented social descriptive norms (at T2 - T5)

- The others mostly play the piano loudly.
- The others mostly play the piano via headphones.

Other-oriented social injunctive norms

- The others believe that they should play the piano loudly.
- The others believe that they should play the piano via headphones.

Self-oriented social descriptive norms (at T1)

- The others believe that I will mostly play the piano loudly.
- The others believe that I will mostly play the piano via headphones.

Self-oriented social descriptive norms (at T2 - T5)

- The others believe that I mostly play the piano loudly.
- The others believe that I mostly play the piano via headphones.

Self-oriented social injunctive norms

- The others believe that I should play the piano loudly.
- The others believe that I should play the piano via headphones.

Personal norms (at T1 – T5)**Other-oriented personal norms**

- I am deeply convinced that the others should play the piano loudly.
- I am deeply convinced that the others should play the piano via headphones.

Self-oriented personal norms

- I am deeply convinced that I should play the piano loudly.
- I am deeply convinced that I should play the piano via headphones.

Manipulation check (at T2 – T5)

- In the past two days, the others have mostly played the piano loudly.
- In the past two days, the others have mostly played the piano via headphones.

Follow-up behavior (at T5)

- Please decide now on how you will practice the next five days.
 - Choice from 0 – 5 days

Perceived realness of the game scenario (at T5)

- I was able to empathize with the situation very well.
- I felt like actually being in the practice room scenario.
- During the game, the scenario felt very real to me.

Supposed goal of the study (at T5)

- What do you think was investigated in this study? [open question]

Credibility of the cover story (at T5)

- Did anything seem strange to you during the course of the study? [open question]

Appendix D – Internal consistencies of norm scales at different measurement time points

Scale	Measurement time point				
	T1	T2	T3	T4	T5
Other-oriented social descriptive norm	0.61	0.94	0.83	0.90	0.56
Self-oriented social descriptive norm	0.48	0.85	0.81	0.80	0.81
Other-oriented social injunctive norm	0.63	0.87	0.75	0.79	0.54
Self-oriented social injunctive norm	0.58	0.82	0.71	0.80	0.73
Other-oriented personal norm	0.68	0.80	0.78	0.81	0.79
Self-oriented personal norm	0.75	0.81	0.80	0.80	0.79
Manipulation check	-	0.95	0.89	0.91	0.51

Note. $N = 365$. Internal consistencies are calculated by Cronbach's alpha. The manipulation check was not assessed at T1.

Appendix E – Seasonal change in social norms and linear change in personal norms

Table E1

Mixed multilevel model of social descriptive and social injunctive norms on the experimental group, a seasonal contrast for the factor time, and their interaction

Multilevel model of social descriptive norms					
	<i>df</i>	<i>B</i>	<i>t</i>	<i>p</i>	<i>r</i>
Experimental group	363	10.46	21.19	< .001***	.74
Seasonal contrast	1458	-0.40	-2.17	.031*	-.06
Interaction	1458	-6.63	-36.02	< .001***	-.69
Multilevel model of social injunctive norms					
	<i>df</i>	<i>B</i>	<i>t</i>	<i>p</i>	<i>r</i>
Experimental group	363	6.82	10.30	< .001***	.48
Seasonal contrast	1458	-0.03	-0.14	.890	-.00
Interaction	1458	-4.24	-22.87	< .001***	-.51

Note. $N = 365$. The seasonal contrast across the five measurement time points was defined as [-1 3 -4 3 -1].
* $p < .05$. *** $p < .001$.

Table E2

Mixed multilevel model of self-oriented and other-oriented personal norms on the experimental group, a linear contrast for the factor time, and their interaction

Multilevel model of self-oriented personal norms					
	<i>df</i>	<i>B</i>	<i>t</i>	<i>p</i>	<i>r</i>
Experimental group	363	2.04	1.68	.094	.09
Linear contrast	1458	-1.90	-7.17	< .001***	-.18
Interaction	1458	0.10	0.37	.713	.01
Multilevel model of other-oriented personal norms					
	<i>df</i>	<i>B</i>	<i>t</i>	<i>p</i>	<i>r</i>
Experimental group	363	1.59	1.45	.147	.08
Linear contrast	1458	-1.26	-4.57	< .001***	-.12
Interaction	1458	0.22	0.79	.428	.02

Note. $N = 365$. The linear contrast across the five measurement time points was defined as [-2 -1 0 1 2].

*** $p < .001$.

Appendix F - Regressions of social and personal norms

Table F1

Regressions of social descriptive and social injunctive norms on experimental group, trait cooperativeness, and their interaction

Regression of social descriptive norms						
	R^2_{adj}	B	β	t	F	p
Model	.55				151.20	***
Experimental group		-17.59	-.62	-8.68		***
Trait cooperativeness		0.12	.21	1.85		.066
Interaction		-0.08	-.24	-1.88		.060
Regression of social injunctive norms						
	R^2_{adj}	B	β	t	F	p
Model	.22				35.71	***
Experimental group		-10.72	-.37	-3.93		***
Trait cooperativeness		0.10	.17	1.15		.250
Interaction		-0.07	-.21	-1.23		.221

Note. $N = 365$. Alpha error corrected by the number of tests, e.g., $\alpha = .05 / 2 = .025$.

*** $p < .0005$.

Table F2

Regressions of self-oriented and other-oriented personal norms on experimental group, trait cooperativeness, and their interaction

Regression of self-oriented personal norms						
	R^2_{adj}	B	β	t	F	p
Model	.09				12.72	***
Experimental group		-5.97	-.13	-1.24		.214
Trait cooperativeness		0.21	.22	1.36		.176
Interaction		0.05	.09	0.50		.618
Regression of other-oriented personal norms						
	R^2_{adj}	B	β	t	F	p
Model	.00				1.18	.319
Experimental group		-7.45	-.18	-1.66		.098
Trait cooperativeness		-0.17	-.20	-1.19		.237
Interaction		0.10	.21	1.08		.279

Note. $N = 365$. Alpha error corrected by the number of tests, e.g., $\alpha = .05 / 2 = .025$.

*** $p < .0005$.

Table F3

Regressions of self-oriented and other-oriented personal norms on social norms, experimental group, trait cooperativeness as well as trustworthiness and predictability of the other players

Regression of self-oriented personal norms						
	R^2_{adj}	B	β	t	F	p
Model	.56				92.59	***
Experimental group		5.77	.12	3.10		.002**
Trait cooperativeness		0.13	.17	3.77		***
Other-oriented social injunctive norm		0.18	.11	2.65		.008*
Self-oriented social descriptive norm		0.78	.63	15.87		***
Self-oriented social injunctive norm		0.20	.16	4.38		***
Regression of other-oriented personal norms						
	R^2_{adj}	B	β	t	F	p
Model	.37				42.99	***
Experimental group		2.53	.06	1.27		.206
Trait cooperativeness		-0.07	-.07	-1.52		.131
Other-oriented social injunctive norm		0.10	.07	1.30		.194
Self-oriented social descriptive norm		0.09	.08	1.70		.091
Self-oriented social injunctive norm		0.66	.58	13.53		***

Note. $N = 365$. Due to multicollinearity (i.e., correlations between predictors above $r = .70$), the factor other-oriented social descriptive norm was excluded from both regressions. Alpha error corrected by the number of tests, e.g., $\alpha = .05 / 2 = .025$.

* $p < .025$. ** $p < .005$. *** $p < .0005$.

Appendix G – Group differences in self-oriented and other-oriented personal norms at different measurement time points

Self-oriented personal norms					
	<i>M (SD)</i>		<i>t</i>	<i>p</i>	δ
	C-EG	D-EG			
T1	51.56 (24.68)	53.02 (25.75)	-0.55	.709	0.06
T2	54.49 (31.12)	44.81 (26.78)	3.18	< .001**	0.33
T3	49.35 (27.93)	43.54 (27.46)	2.00	.023	0.21
T4	47.51 (27.87)	43.48 (27.87)	1.38	.084	0.14
T5	46.05 (26.25)	43.70 (26.68)	0.85	.199	0.09
Other-oriented personal norms					
	<i>M (SD)</i>		<i>t</i>	<i>p</i>	δ
	C-EG	D-EG			
T1	65.04 (22.07)	64.17 (23.17)	0.37	.357	0.04
T2	68.46 (25.76)	64.61 (27.76)	1.37	.085	0.14
T3	65.74 (25.03)	61.86 (26.09)	1.45	.074	0.15
T4	63.82 (25.89)	59.24 (28.28)	1.61	.054	0.17
T5	62.16 (24.23)	59.47 (26.82)	1.01	.157	0.11

Note. $N = 365$. One-sided Welch's t -tests. Degrees of freedom were adapted according to Welch correction. C-EG = cooperative group, D-EG = defective group. δ = effect size Cohen's delta. Alpha error corrected by the number of tests, e.g., $\alpha = .05 / 10 = .005$.

** $p < .001$.

Appendix E

Working Paper: An Experimental Attempt at Validating an Agent-Based Model on Decision-Making, Social Norm Change, and Norm Internalization

The following manuscript has not been published. It was accepted to be presented at the *Social Simulation Conference 2023*.

Batzke, M. C. L., & Ernst, A. (2023a). *An Experimental Attempt at Validating an Agent-Based Model on Decision-Making, Social Norm Change, and Norm Internalization* [Unpublished manuscript]. Center for Environmental Systems Research, University of Kassel.

An Experimental Attempt at Validating an Agent-Based Model on Decision-Making, Social Norm Change, and Norm Internalization

Marlene C. L. Batzke¹ [0000-0001-5882-9813] and Andreas Ernst¹ [0000-0001-5773-4441]

¹ Center for Environmental Systems Research, University of Kassel, Germany

Abstract

Understanding norm internalization remains one of the key questions that are still open in norm research. It refers to the process of individuals' developing personal norms. After formalizing a theory on norm internalization, implementing it into an agent-based model, and conducting an experimental study on norm change processes, the present work attempts at comparing experimental and simulation data. The agent-based DINO model simulates agents' decision-making and actions, social norms, and norm internalization in a 3-person Prisoners' Dilemma Game. The experiment was designed to match the data from the agent-based model. $N = 365$ participants were invited to play the structurally same game, while their social and personal norms were assessed repeatedly.

A first comparison of participants' and agents' behavior, social norms, and norm internalization processes is presented with regard to different social settings (cooperative vs. defective) and participants' willingness to cooperate (cooperator vs. conditional cooperator vs. defector). Results generally show similarities between agents' and study participants' conditional cooperators, suggesting some plausibility of the assumed processes in the agent-based DINO model. The comparison led to assume that one mechanism in norm internalization that was so far missing in the agent-based model and was therefore added to the DINO norm internalization process: asymmetry in internalizing cooperativeness versus defectivity. Further possible mechanisms in norm internalization and limitations of the comparison are discussed.

Keywords: Social Norms, Norm Internalization, Decision-Making, Learning, Social Dilemma, Cooperation.

1. Introduction

The question of how norms are internalized and internalized norms change is one of the key questions still open in norm research. Norm internalization describes the process of how individuals adopt and change their personal norms, being an individual's beliefs about the (in)appropriateness of a behavior in a specific situation (Batzke & Ernst, 2023b). Personal norms can be differentiated from social norms, being beliefs about what other consider appropriate or normal behavior (Bicchieri et al., 2018; Cialdini et al., 1990). The power of social norms has repeatedly been shown (Asch, 1956; Deutsch & Gerard, 1955; Sherif, 1936), with social norms influencing small group cooperation (Ostrom, 2000) and creating tipping points for large-scale transformations (Nyborg et al., 2016).

Yet, many authors have ascribed particular significance to norm internalization, being important for norm maintenance and long-term behavior change (Axelrod, 1986; Gintis, 2004). Studies have shown that the behavioral influence of social norms is largely mediated by personal norms (Hopper & Nielsen, 1991; Thøgersen, 1999). This may lead to the assumption that personal norms are influenced by social norms, yet particularly important for behavioral decisions (Tverskoi et al., 2023) and decisions in the absence of social norm enforcement (Thøgersen, 2006). However, so far there is a lack of specific assumptions about the process of norm internalization (Neumann, 2010). There are few simulation models that conceptualized it (Neumann, 2014) and even fewer empirical studies that investigated it (Bamberg & Möser, 2007).

Norm internalization can be regarded as the product of the complex interplay of individuals' goals, habits, behaviors, et cetera, interacting with the social and physical environment over time, making internalization a suitable candidate to be studied via agent-based simulation (Batzke & Ernst, 2023b). While simulation models on norm internalization may uniquely contribute to the understanding of the underlying mechanisms and dynamics (Andrighetto et al., 2010; Villatoro et al., 2015), it needs a combination of simulation and empirical data to further advance the study and understanding of norm internalization.

The present work provides an attempt at comparing data on norm internalization from an agent-based model with experimental data. The conducted experiment was designed to produce data about variables matching those from agent-based modeling. This allows partly testing, potentially validating, and improving an agent-based model (Gilbert, 2008). A psychologically grounded theory of decision-making, including social norms, goals, and habits was implemented in an agent-based model in the context of a social dilemma game. Hence, the

present approach also addresses a comparison of behavioral data and social norm change processes.

In the following, the agent-based model and the conducted experimental study are presented. Then, agent-based modeling and experimental data are compared regarding participants' and agents' behaviors, social norms, and personal norms. Finally, results are discussed.

2. An Agent-Based Model

The agent-based model DINO model (*D*ynamics of *I*nternalization and *D*issemination of *N*orms) is presented and tested regarding the conditions and effects of norm internalization in Batzke and Ernst (2023b). The model simulates the behavior of three agents in a 3-person *Prisoner's Dilemma Game* (see Dawes, 1980), describing the core of a conflict common to many situations. DINO agents' decision-making is determined by a weighted multi-attribute utility matrix, which represents goals, social norms, and habits as motivational factors in decision-making. There are three types of goals represented, according to Deutsch (1958): the individualistic, cooperative, and competitive goal. Moreover, there are two types of social norms implemented, along the differentiation by Cialdini et al. (1990): social descriptive norms (i.e., what others *do*) and social injunctive norms (i.e., what others (dis)approve).¹ All motivational factors are defined by a situational expectation and a personal value factor, along the *Theory of Planned Behavior* (Ajzen, 1991) and other expectation-value theories (Atkinson, 1957; Fishbein & Ajzen, 1975). Situational factors are adapted over time, based on agents' experiences. Personal value factors represent the individual importance of a motivational factor and are represented statically (see Chapter 4). Agents' behavior is determined by their intention, defined as the weighted sum (i.e., expectation-value products) of all motivational factors.

Over and above the adaptation of situational expectations, which are assumed to be made rather quickly whenever the social situation changes, the norm internalization process was implemented as a slow adaptation process, representing a more aggregated form of learning, depending on DINO agents' personal values and their experiences. Change in personal norms depends on agents' normative judgement regarding their last chosen action. Based on this evaluation, agents adapt their personal norms in a stepwise process. Personal norms influence

¹ In the model, social injunctive norms are represented as constants. Therefore, only agents' and participants' social descriptive norms are compared in Chapter 4. For reasons of simplicity, they are referred to as social norms.

decision-making through emphasizing or inhibiting the importance of the other motivational factors, representing a higher-level factor in decision-making similar to personal values (for details see Batzke & Ernst, 2023b).






3. An Experiment with Participants

The experiment is presented, and temporal differences of personal and social norm change are analyzed in Batzke & Ernst (2023a). Like agents in the agent-based model, participants played a repeated 3-person Prisoner's Dilemma Game, themselves being one of the three players. Their artificial co-players were predefined behavioral sequences. The behavior of the co-players differed in the two experimental groups (see Figure 1). The cooperative experimental group (C-EG) was characterized by predominantly cooperative social setting and the defective group (D-EG) by a defective setting. The experimental variation was expected to influence the norm internalization process. Moreover, the social setting was varied repeatedly within each group (see Figure 1). This was expected to show in changing social norms. In total, the game consisted of 17 rounds.

$N = 365$ participants were sampled via a survey institute and invited to play the online "Concert Game". Participants were asked to imagine themselves being a pianist, preparing for the first grand concert. To practice for the concert, they have rented a practice room with an identical piano for three hours daily. However, the room is in a triangle with two other practice rooms, and their thin walls make it difficult to practice loudly without disturbing each other. While the pianos have headphone options to avoid disturbing others, headphones limit the learning achievements. Hence, participants must choose every day whether to practice loudly or with headphones, knowing that they will share the space with the same two people in the coming days.

Before the game, participants social value orientation (hereafter called: willingness to cooperate) was assessed via the slider measure (Murphy et al., 2011). The game was explained, the payoff matrix introduced, an example round played, and participants' understanding of the instructions tested. Before, during, and after the game, at in total 5 measurement time points (see green triangles in Figure 1) participants were asked to rate their personal norms (e.g., "I am deeply convinced that I should play the piano via headphones.") and social norms (e.g., "The others mostly play the piano via headphones.") from 1 "not agree at all" to 101 "absolutely agree" on each two items.

Figure 1*Operationalization of experimental groups*

Round	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
C-EG	Co-player 1	C	C	C	C	C	C	D	D	D	C	C	C	C	C	C	C
	Co-player 3	C	C	C	C	C	D	D	D	D	D	C	C	C	C	D	D
D-EG	Co-player 1	D	D	D	D	D	C	C	C	C	C	D	D	D	D	C	C
	Co-player 3	D	D	D	D	D	D	C	C	C	D	D	D	D	D	D	D
																	

Note. Participants played the 3-person Prisoner's Dilemma Game with two artificial co-players (1 and 3), themselves being player 2. The game differed between the cooperative (C-EG) and defective experimental group (D-EG). Each experimental group consisted of three phases, characterized by either a cooperation (blue color) or defection (red color) of the co-players, and a final phase, characterized by a mixed setting in which one co-player cooperated and the other defected (white color). In between phases, single rounds of a mixed setting were added to make the game more realistic. Social and personal norms were assessed before the game (T1) and roughly after each phase (after rounds 3, 9, 13 and 17) at T2 – T5.

4. Comparison of Simulation and Experimental Data

To compare simulation and experimental data regarding participants' and agents' behavior, social norms, and personal norms, the agent-based model was modified so that one agent can play the social dilemma game for 17 rounds with two predefined behavioral sequences – like study participants did. The experimental design was similarly applied to the model, with agents playing in the same two conditions: the cooperative and defective experimental group. Hence, agents and participants were put in the exact same situations.

To investigate interindividual differences, results were looked at with respect to participants' and agents' willingness to cooperate. DINO agents were categorized into three groups: cooperators, conditional cooperators, and defectors. Categorization was based on certain ranges of their personal value factors according to the agent type descriptions in Batzke and Ernst (2023b). Within these ranges, values were randomly drawn for 100 agents per category and condition. Hence, in total 600 model runs (2 conditions x 3 categories x 100 agent draws) were conducted. As the agent-based model is deterministic, single model runs were not repeated.

Study participants were grouped along their willingness to cooperate (along the categorization by Murphy et al., 2011) into altruists ($n = 148$), prosocials ($n = 142$), individualists ($n = 63$), and competitiveness ($n = 13$). Due to the small number of competitiveness, the

category was merged with individualists. The resulting three categories are hereafter, like the agent categories, referred to as cooperators, conditional cooperators, and defectors.

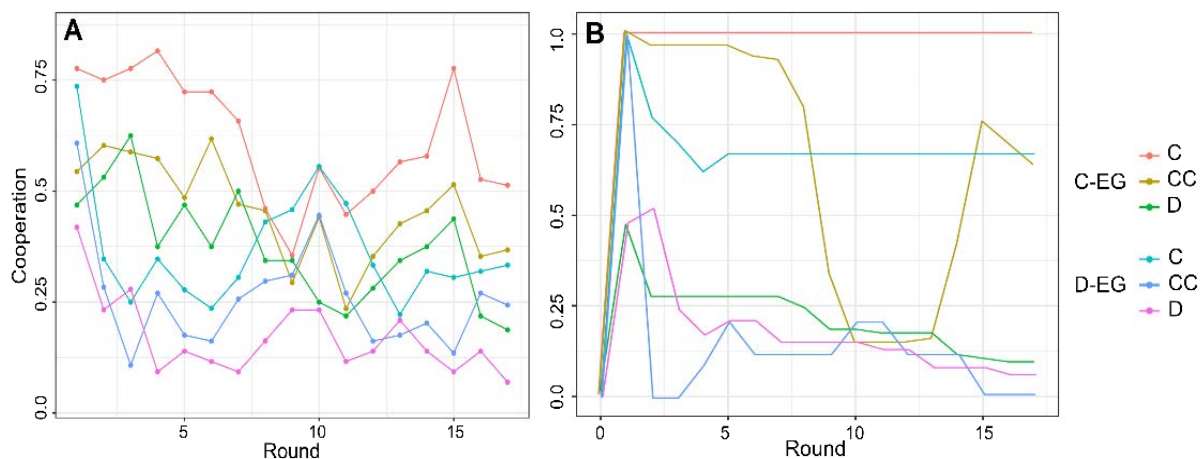
4.1 Behavior

Figure 1 shows study participants' (Figure 2A) and DINO agents' (Figure 2B) cooperative behavior across the 17 rounds of the social dilemma game, depending on the experimental group (cooperative vs. defective) and their willingness to cooperate (cooperators vs. conditional cooperators vs. defectors).

Agents' behavior generally shows to be less volatile than participants' behavior. However, particularly participants and agents categorized as conditional cooperators show similar behavioral developments across time. They are responsive to the first social norm change around round ten as well as to the second change around round 15. Study participants categorized as cooperators show a similar behavioral pattern, while defectors are less but still somewhat influenced by the repeated social norm changes. The respective agent types do not show that pattern. Moreover, agent cooperators generally are generally more cooperative.

Figure 2

Behavior of participants and agents



Note. Study participants' (Figure A) and DINO agents' (Figure B) cooperative behavior across 17 rounds of the social dilemma game, depending on the experimental group (C-EG = cooperative experimental group vs. D-EG = defective experimental group) and their willingness to cooperate (C = cooperators vs. CC = conditional cooperators vs. D = defectors). Cooperation ranges between 0 and 1.

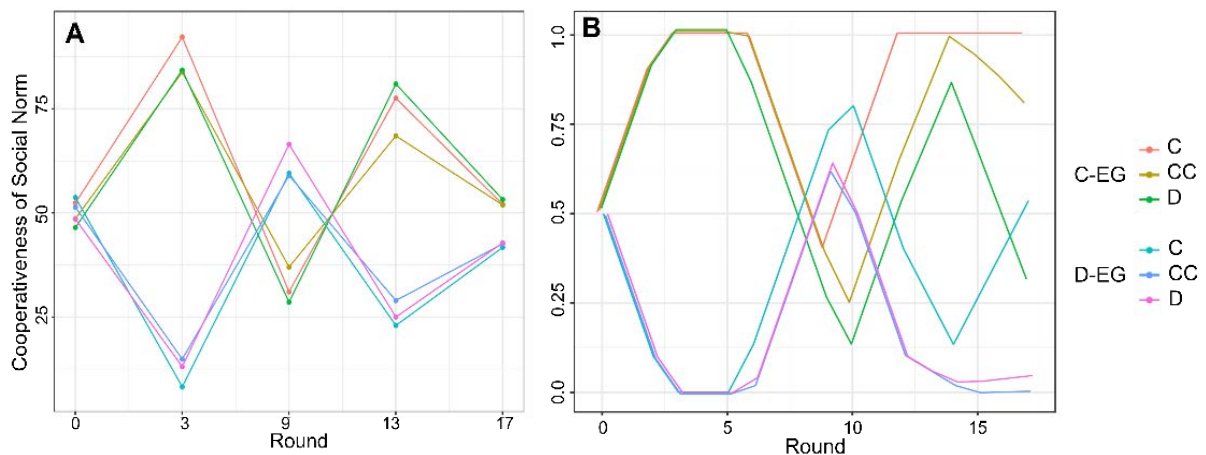
4.2 Social Norm Change

Figure 3 depicts participants' (Figure 3A) and agents' (Figure 3B) cooperativeness of the social norm across time (i.e., 17 rounds), depending on the experimental group (cooperative vs. defective) and their willingness to cooperate (cooperators vs. conditional cooperators vs. defectors).

Participants and agents' developments in social norms show similar patterns across experimental groups and willingness to cooperate categories. Yet, agents' social norm adaptation is faster than participants', plateauing after few rounds of the game. Moreover, in participants' the amplitude of adapting social norms to the social setting decays across time, a characteristic that agents' social norm change do not show.

Figure 3

Social norm changes in participants and agents



Note. Study participants' (Figure A) and DINO agents' (Figure B) change in the cooperativeness of the social norm across 17 rounds of the social dilemma game, depending on the experimental group (C-EG = cooperative experimental group vs. D-EG = defective experimental group) and their willingness to cooperate (C = cooperators vs. CC = conditional cooperators vs. D = defectors). In the study (Figure A), the cooperativeness of social norms ranges between 0 and 100, in the model (Figure B) between 0 and 1.

4.3 Personal Norm Change – Norm Internalization

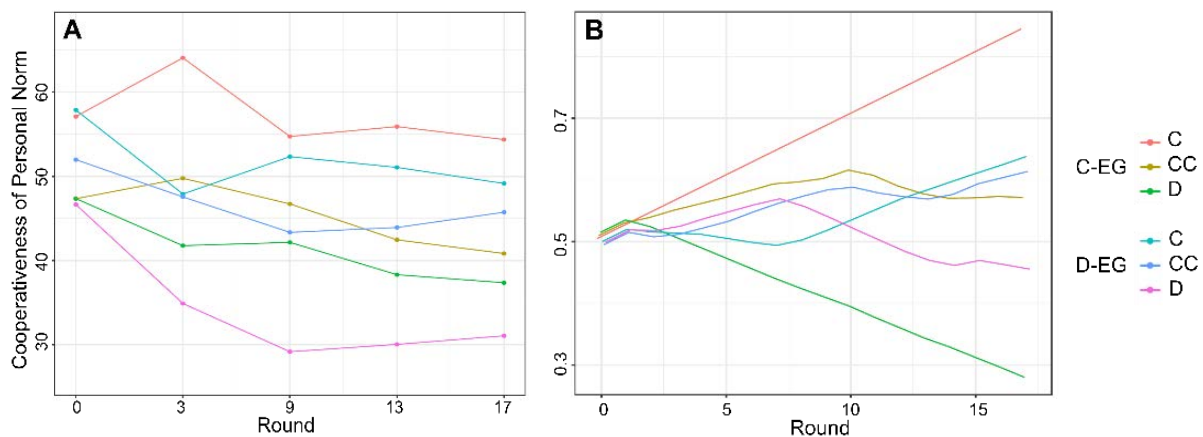
Figure 4 shows participants' (Figure 4A) and agents' (Figure 4B) cooperativeness of the personal norm across time (i.e., the norm internalization process), depending on the experimental group (cooperative vs. defective) and their willingness to cooperate (cooperators vs. conditional cooperators vs. defectors).

When comparing agents' and participants' norm internalization, especially one point strikes the eye: Agents' norm internalization is generally more towards cooperativeness. In

participants' personal norm change, there is no learning of a cooperative norm in any group or category, but rather of a defective norm. Nevertheless, the internalization processes of agents and participants that are categorized as conditional cooperators show similarities in both experimental groups.

Figure 4

Personal norm changes in participants and agents



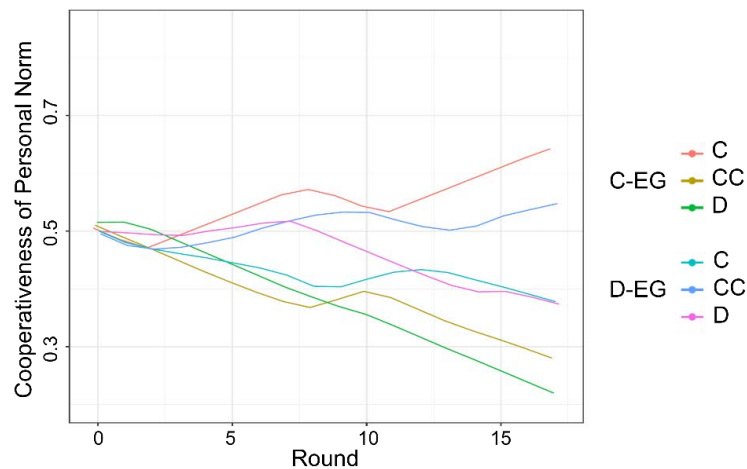
Note. Study participants' (Figure A) and DINO agents' (Figure B) change in the cooperativeness of the personal norm across 17 rounds of the social dilemma game, depending on the experimental group (C-EG = cooperative experimental group vs. D-EG = defective experimental group) and their willingness to cooperate (C = cooperators vs. CC = conditional cooperators vs. D = defectors). In the study (Figure A), the cooperativeness of personal norms ranges between 0 and 100, in the model (Figure B) between 0 and 1.

The DINO norm internalization process was adjusted to account for that point by introducing asymmetry in the ease of internalizing cooperativeness versus defectivity. The threshold to internalize cooperativeness was raised and thus internalizing cooperativeness made more improbable. Results are shown in Figure 5.

That adjustment significantly improved the overall similarity between participants' (Figure 4A) and agents' norm internalization patterns (Figure 5). Particularly the patterns of cooperators have improved by introducing asymmetry. Regarding defectors, there is still a substantial difference between model and experimental data. In the model, defector agents in the cooperative group (green line) internalize a defective personal norm more quickly than those in the defective group (pink line). In participants (see the same lines in Figure 4A), it is the other way round.

Figure 5

Personal norm changes in agents with asymmetry in norm internalization



Note. DINO agents' change in the cooperativeness of the personal norm after implementing asymmetry in the ease to internalize cooperativeness versus defectivity (further explanations in the text). Results are shown across 17 rounds of the social dilemma game, depending on the experimental group (C-EG = cooperative experimental group vs. D-EG = defective experimental group) and their willingness to cooperate (C = cooperators vs. CC = conditional cooperators vs. D = defectors). The cooperativeness of personal norms ranges between 0 and 1.

5. Discussion

The present work represents an attempt at comparing time-series simulation and experimental data on norm internalization – the process of personal norm change. It aimed at testing and improving an agent-based model on decision-making and norm internalization as well as better understanding social norm change and norm internalization. Throughout the comparisons of behavior, social norm change, and norm internalization, DINO agents and study participants categorized as conditional cooperators showed similarities, making the DINO model a good candidate for further testing and exploration of these processes.

The comparison led to assume that a mechanism was missing in the DINO norm internalization process: asymmetry in internalizing cooperativeness versus defectivity. This means that a cooperative norm is more difficult to internalize than a defective norm. The argument of asymmetry relates to Tversky and Kahneman's (1992) *Prospect Theory*. Therein, they describe an asymmetry between losses and gains, stating that negative experiences have a stronger impact than positive. Possibly, this also affects norm internalization. Implementing that aspect improved the overall similarity of agents' and participants' norm internalization. The comparison further led to the assumption that social norm adaptation processes might be slightly slower than assumed in the agent-based model and might decay in their amplitude

across time. One may suggest that there is a decreasing social norm adaptation speed across time. Moreover, the comparison showed that DINO cooperator and defector agents might be unrealistically extreme and require adjustments. Participants generally exhibited stronger similarities with conditional cooperator agents, which especially showed in the behavioral comparison.

There are numerous limitations to this first attempt of comparing participants' and agents' behaviors, social norm changes, and norm internalization processes. First, the comparison was merely based on visual inspection and awaits to be statistically validated in future work. Second, the game lasted only 17 rounds, raising the question whether norm internalization can be observed within such a short time frame. More long-term studies on norm internalization are needed. Third, the results are specific for the 3-person Prisoner's Dilemma Game. Hence, the suggested asymmetry in norm internalization might be limited to the characteristics of the situation. Fourth, the DINO internalization process in defectors facilitates (rather than impedes) learning a defective personal norm in a cooperative setting. In the model, defector agents may exploit others, which leads to goal fulfillment and thus supporting their actions via internalizing the according norm. While this principle seems to explain some dynamics in the norm internalization dynamics, potentially another factor is missing that accounts for participants categorized as defectors learning a defective norm particularly in the defective condition. For instance, (mis)trust in the others could explain these differences. Mistrust might grow with defection of others as well as repeated behavioral changes.

The comparison of experimental and simulated data allows for a deeper understanding of psychological processes. It allows generating hypotheses about underlying mechanisms and testing their consequences across time. The present approach provided one step towards better understanding norm internalization via combining experiment and agent-based modeling. New insights were attained that may inspire future research.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211.
- Andrighetto, G., Villatoro, D., & Conte, R. (2010b). Norm internalization in artificial societies. *AI Communications*, *23*(4), 325–339. <https://doi.org/10.3233/AIC-2010-0477>
- Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and Applied*, *70*(9), 1. <https://doi.org/10.1037/h0093718>
- Atkinson, J. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*, 359–372.
- Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review*, *80*(4), 1095–1111. <https://doi.org/10.2307/1960858>
- Batzke, M. C. L., & Ernst, A. (2023a). *Changing fast, changing slow: Investigating temporal differences between social and personal norm change underlying cooperation*. [Manuscript submitted for publication]. Center for Environmental Systems Research, University of Kassel, Germany.
- Batzke, M. C. L., & Ernst, A. (2023b). Conditions and effects of norm internalization. *Journal of Artificial Societies and Social Simulation*, *26*(1), 1–31. <https://doi.org/10.18564/jasss.5003>
- Bamberg, S., & Möser, G. (2007). Twenty years after Hines, Hungerford, and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behaviour. *Journal of Environmental Psychology*, *27*(1), 14–25. <https://doi.org/10.1016/j.jenvp.2006.12.002>
- Bicchieri, C., Muldoon, R., & Sontuoso, A. (2018). Social Norms. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2018/entries/social-norms/>
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026. <https://doi.org/10.1037/0022-3514.58.6.1015>
- Dawes, R.M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*(1), 169–193. <https://doi.org/10.1146/annurev.ps.31.020180.001125>
- Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution*, *2*(3), 265–279.
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629. <https://doi.org/10.1037/h0046408>
- Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Boston, MA: Addison-Wesley.

- Gilbert, N. (2008). *Agent-based models*. Sage Publications.
<https://doi.org/10.4135/9781412983259>
- Gintis, H. (2004). The genetic side of gene-culture coevolution: Internalization of norms and prosocial emotions. *Journal of Economic Behavior and Organization*, 53, 57–67.
- Hopper, J. R., & Nielsen, J. M. (1991). Recycling as altruistic behavior: Normative and behavioral strategies to expand participation in a community recycling program. *Environment and Behavior*, 23(2), 195-220. <https://doi.org/10.1177/0013916591232004>
- Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771-781.
<https://doi.org/10.1017/S1930297500004204>
- Neumann, M. (2010). Norm internalisation in human and artificial intelligence. *Journal of Artificial Societies and Social Simulation*, 13(1), 12. <https://doi.org/10.18564/jasss.1582>
- Neumann, M. (2014). How are norms brought about? A state of the art of current research. In R. Conte, G. Andrighetto, & M. Campenni (Eds.), *Minding norms: Mechanisms and dynamics of social order in agent societies* (pp. 50–67). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199812677.003.0004>
- Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W.N., Arrow, K. J., Barrett, S., Carpenter, S., Chapin III, F. S., Crépin, A.-S., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S. A., Madsen, O. J., ... De Zeeuw, A. (2016). Social norms as solutions. *Science*, 354(6308), 42-43.
<https://doi.org/10.1126/science.aaf8317>
- Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3), 137-158. <http://www.jstor.org/stable/2646923>
- Sherif, M. (1936). *The psychology of social norms*. Harper.
- Thøgersen, J. (1999). The ethical consumer: Moral norms and packaging choice. *Journal of Consumer Policy*, 22(4), 439-460. <https://doi.org/10.1023/A:1006225711603>
- Thøgersen, J. (2006). Norms for environmentally responsible behaviour: An extended taxonomy. *Journal of Environmental Psychology*, 26(4), 247-261.
- Tverskoi, D., Guido, A., Andrighetto, G., Sánchez, A., & Gavrilets, S. (2023). Disentangling material, social, and cognitive determinants of human behavior and beliefs. *Humanities and Social Sciences Communications*, 10(1), 1-13.
<https://doi.org/10.1057/s41599-023-01745-4>
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.
<https://doi.org/10.1007/BF00122574>

Villatoro, D., Andrighetto, G., Conte, R., & Sabater-Mir, J. (2015). Self-policing through norm internalization: A cognitive solution to the tragedy of the digital commons in social networks. *Journal of Artificial Societies and Social Simulation*, 18(2), 2. <https://doi.org/10.18564/jasss.2759>

Erklärung zum Eigenanteil

Universität Kassel, Fachbereich Humanwissenschaften

Erklärung zu kumulativen Dissertationen im Promotionsfach Psychologie

Erklärung über den Eigenanteil an den veröffentlichten oder zur Veröffentlichung vorgesehenen wissenschaftlichen Schriften innerhalb meiner Dissertationsschrift, Ergänzung zu § 7 Abs. 4 der Allgemeinen Bestimmungen für Promotionen an der Universität Kassel vom 14. August 2021

I. Allgemeine Angaben

Name: Batzke, Marlene

Institut: Center for Environmental Systems Research, Universität Kassel,
Fachgebiet Umweltpsychologie/Umweltsystemanalyse

Thema der Dissertation: Dynamics of Norms in Decision-Making – A Psychological
Analysis Combining Theory, Experiment, and Social Simulation

II. Nummerierte Aufstellung der eingereichten Schriften

1. Batzke, M. C. L., & Ernst, A. (2023b). *Changing Fast, Changing Slow: Investigating Temporal Differences Between Social and Personal Norm Change Underlying Cooperation* [Manuscript submitted for publication]. Center for Environmental Systems Research, University of Kassel.
2. Batzke, M. C. L., & Ernst, A. (2023c). Conditions and Effects of Norm Internalization. *Journal of Artificial Societies and Social Simulation*, 26(1), 1–31. <https://doi.org/10.18564/jasss.5003>
3. Dannenberg, A., Gutsche, G., Batzke, M. C. L., Christens, S., Engler, D., Mankat, F., Möller, S., Weingärtner, E., Ernst, A., Lumkowsky, M., von Wangenheim, G., Hornung, G., & Ziegler, A. (in press). The effects of norms on environmental behavior. *Review of Environmental Economics and Policy*.

III. Darlegung des eigenen Anteils an diesen Schriften:

Zu Nr. 1

Entwicklung der Konzeption: überwiegend

Literaturrecherche: vollständig

Methodenentwicklung: überwiegend

Entwicklung des Versuchsdesigns: vollständig

Datenerhebung: vollständig

Datenauswertung: vollständig

Ergebnisdiskussion: vollständig

Erstellen des Manuskripts: überwiegend

Bewältigung des Review-Prozesses: vollständig

Zu Nr. 2

Entwicklung der Konzeption: überwiegend

Literaturrecherche: vollständig

Entwicklung des Modells: überwiegend

Datengenerierung: vollständig

Datenauswertung: überwiegend

Ergebnisdiskussion: überwiegend

Erstellen des Manuskripts: überwiegend

Programmierung: vollständig

Bewältigung des Review-Prozesses: vollständig

Zu Nr. 3

Entwicklung der Taxonomie: mehrheitlich

Entwicklung des konzeptuellen Modells inklusive Literaturrecherche: überwiegend

Entwicklung des Studien-Reviews: in Teilen

Erstellen des Manuskripts: mehrheitlich

Erstellen des Anhangs: überwiegend

Beweisführung: überwiegend

Bewältigung des Review-Prozesses: mehrheitlich

Eidesstattliche Versicherung und Erklärung

Erklärung gemäß § 8 der Allgemeinen Bestimmungen für Promotionen der Universität Kassel vom 14.07.2021.

1. Bei der eingereichten Dissertation zu dem Thema
„Dynamics of Norms in Decision-Making – A Psychological Analysis Combining Theory, Experiment, and Social Simulation“
handelt es sich um meine eigenständig erbrachte Leistung.
2. Anderer als der von mir angegebenen Quellen und Hilfsmittel habe ich mich nicht bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen veröffentlichten oder unveröffentlichten Werken übernommene Inhalte als solche kenntlich gemacht.
3. Die Dissertation oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.
4. Die abgegebenen digitalen Versionen stimmen mit den abgegebenen schriftlichen Versionen überein.
5. Ich habe mich keiner unzulässigen Hilfe Dritter bedient und insbesondere die Hilfe einer kommerziellen Promotionsberatung nicht in Anspruch genommen.
6. Im Fall einer kumulativen Dissertation: Die Mitwirkung von Koautoren habe ich durch eine von diesen unterschriebene Erklärung dokumentiert. Eine Übersicht, in der die einzelnen Beiträge nach Ko-Autoren und deren Anteil aufgeführt sind, füge ich anbei.
7. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

Unterschrift der Antragstellerin:

Ort und Datum

Marlene Batzke, M.Sc.