Motivational Aspects of Moral Decision Making

Dissertation

zur Erlangung des akademischen Grades eines
Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt von Johannes Theodor Doerflinger

an der

Universität Konstanz

Mathematisch-Naturwissenschaftliche Sektion
Fachbereich Psychologie

Konstanz, den 08.01.2019
Tag der mündlichen Prüfung: 20.12.2018
1. Referent: Prof. Dr. Peter M. Gollwitzer
2. Referent: Prof. Dr. Urs Fischbacher
3. Referent: Privatdozent Dr. Frank Wieber
“… for there is nothing either good or bad, but thinking makes it so.”

Shakespeare, Hamlet (Act II, Scene 2)

Acknowledgments

I am very grateful for the support I received while writing this thesis. Many people helped me in my scientific endeavors. I want to thank especially the co-authors on the research papers presented here and elsewhere, but also many others without whom this thesis would not have been possible.

First, I want to thank Peter M. Gollwitzer. He has not only been a supportive supervisor, but also a mentor since I worked as an intern at his chair. I am thankful for your guidance and your enthusiasm for psychological science. Thank you for always being open to discuss my ideas, no matter how crazy.

I also want to thank Urs Fischbacher for inspiring me to work across disciplines. Your feedback and support as head of the Graduate School of Decision Sciences was invaluable. I also want to thank you and Frank Wieber for your willingness to review my thesis.

Thank you to Martin Bruder and Nora Heinzelmann for showing me the first steps to a psychological exploration of human morality.

To the chair of Social Psychology and Motivation at the University of Konstanz – every former and current member: Torsten Martiny-Hünger, Lukas Thürmer, Klaus Harnack, Lucas Keller, Maik Bieleke, Lucia Görke, Anja Weiergräber, Inge Schweiger Gallo, Baasandulam Strube, Andreas Danielowski, Sigmar Papendick – Thank you for all the discussions, your help with small and big challenges, feedback, and ideas.

Thank you to Gabriele Oettingen. Your work and enthusiasm has been an inspiration. I would also like to thank Christina Crosby, Ana Gantman, Bryan Sim, Sandra Wittleder, and other former and current team members of the motivation lab at NYU. You welcomed me into a friendly and intellectually stimulating group.
The research assistants who have supported my studies over the years also deserve my gratitude: Katrina Wahl, Dina Hansen and Hamilkar Oueslati. Similarly, I am especially grateful to Erkan Poyraz, for his great assistance in the design and data collection for Research Paper II.

I am very grateful for the support and funding by the Graduate School of Decision Sciences (GSDS). The GSDS gave me the opportunity to work and learn in an exceptional interdisciplinary research environment. Particular thanks to David Grammling, who shared my enthusiasm for morality as a subject of research. To Jutta Obenland and Justine Overall – Thank you for your help at all times.

To Katrin Platzer and Markus Thiemel, thank you for opening my eyes for the curiosities of philosophy.

To my parents, without your unconditional support and love, this would not have been possible. Thank you!

Lea, Rosenpuck – thank you for listening to my ideas. Thank you for your patience and your support. Thank you for enduring me at times when I was absorbed by my research. Thank you for your love.
Abstract

In the present thesis I investigate moral identity from a goal perspective (Research Paper I and II), and the moderating role of cognitive orientations on the influence of incidental emotional information in moral dilemma judgments (Research Paper III). The overarching theme of these topics is an integration of moral psychology and motivation science.

The first research paper is a theory paper in which the goal perspective on moral identity is developed. It builds on symbolic self-completion theory and contrasts the conceptualization of moral identity as a goal with two prominent theories in moral psychology: the Self Model and the Social Cognitive Theory. Central properties of the goal construct are reviewed and their implications for moral identity as a goal are discussed. Emerging research questions and a comprehensive review of empirical findings related to identity goals and the moral self point toward the usefulness of goals to conceptualize moral identity. In particular dynamics of goal activation and goal completion, the interrelationship of multiple goals as well as cultural variations of goals and morality are highlighted. Taken together, Research Paper I suggests that moral identity research can benefit from taking motivational processes into account.

While research Paper I lays the theoretical groundwork for an investigation of moral identity from the vantage point of motivation science, the second research paper goes one step further by testing specific hypotheses derived from the goal perspective on moral identity. In particular, I predict that individuals typically have more than one identity goal and if multiple identity goals are active at the same time than goal striving efforts depend on the configuration of active identity goals. In Research Paper II goal dynamics of multiple simultaneously active identity goals are explored in two experiments. In Experiment 1, participants with an active moral identity goal exhibited more effort toward that goal than participants with an inactive moral identity goal, but only when no other competing goal was active. Importantly, the moral identity goal and the secondary goal could not be pursued with the same means. In Experiment 2, two identity goals were simultaneously activated and then a means serving both identity goals was presented. Under these conditions, both identity goals had an additive effect on goal striving. The experiments provide evidence in support of the goal perspective developed in Research Paper I and...
indicate that the available means for identity goals determine the effect of multiple active identity goals on behavior.

Finally, in the third research paper I shift from the global level of moral identity to a more in-depth analysis of moral judgments. In particular, the joint effect of incidental emotional information and mindsets that influence information processing is investigated. Furthermore, the studies provide insight into the underlying cognitive processes of the deliberative and implemental mindsets. In three experiments moral judgments of trolley-type dilemmas are investigated. The experiments demonstrate that an emotional emphasis on consequences can sway moral judgments toward consequentialism or deontology depending on the target of the emotional emphasis. In Experiment 1 and 2, a moderation effect of mindsets on the influence of emphasis was observed with stronger emphasis effects in the deliberative mindset and weaker emphasis effects in the implemental mindset. In Experiment 2, eye-tracking data of participants observing a visual presentation of dilemma stories revealed a narrower focus of attention for participants in the implemental mindset. This is a likely mechanism underlying the moderating role of mindsets on emphasis biases in moral judgments. In the third experiment, planning to engage in thorough reflective thinking was tested as an alternative method to control the emphasis bias. Emotional emphasis affected moral judgments irrespective of whether participants were in an impulsive or a reflective mode of thought.

Taken together, the three research papers show how motivational processes influence moral identity and moral judgments. Each of the three research papers makes a substantial contribution to the integration of moral psychology and motivational science. Research Paper I provides a theoretical framework for conceptualizing moral identity goals. Research Paper II shows that identity goals do not affect behavior in isolation; instead, the entire configuration of active goals determines how identity goals are translated into behavior. Research Paper III demonstrates that the cognitive tuning arising from different goal states affects moral judgments. In sum, the thesis offers a novel perspective on the role of goals in moral psychology.
Zusammenfassung

In der vorliegenden Arbeit untersuche ich moralische Identität aus der Sicht der Motivationspsychologie (Forschungsartikel I & II) und den Moderationseffekt von kognitiven Orientierungen auf den Einfluss von nebensächlicher emotionaler Information auf moralische Urteile (Forschungsartikel III). Das übergeordnete Thema der Arbeit ist damit die Integration von Moralspsychologie und Motivationsforschung.


Aufbauend auf dem theoretischen Fundament aus Forschungsartikel I werden in Forschungsartikel II spezifische Hypothesen getestet, die sich aus der Zielperspektive auf moralische Identität ergeben. Es wird angenommen, dass Menschen typischerweise mehr als ein Identitätsziel haben. Wenn mehrere Identitätsziele gleichzeitig aktiv sind, dann ist zu erwarten, dass Anstrengungen im Zielstreben von der Konfiguration der aktiven Identitätsziele abhängen. In Experiment 1 wandten Teilnehmer mit einem aktiven moralischen Identitätsziel mehr Anstrengung an, um nach diesem Ziel zu streben, als Teilnehmer mit einem inaktiven moralischen Identitätsziel, aber nur, wenn kein weiteres Ziel aktiv war, das um Ressourcen mit dem moralischen Identitätsziel im Konflikt stand. Entscheidend ist, dass in Experiment 1 das moralische Identitätsziel und das zweite Ziel nicht mit denselben Mitteln verfolgt werden konnten. In Experiment 2 wurden zwei Identitätsziele gleichzeitig aktiviert. Anschließend wurde ein Mittel präsentiert, mit
dem beide Ziele verfolgt werden konnten. Unter diesen Umständen hatten beide Identitätsziele einen additiven Effekt auf das Zielstreben. Die Ergebnisse der Experimente stützen die Zielperspektive, die in Forschungsartikel I entwickelt wurde, und deuten darauf hin, dass die vorhandenen Mittel für Identitätsziele bestimmen, wie mehrere aktive Identitätsziele gemeinsam Verhalten beeinflussen.


Table of Contents

Acknowledgments V
Abstract VII
Zusammenfassung IX
Table of Contents XI
List of Figures XIV
List of Tables XV
Synopsis 1

A Brief Review of Moral Psychology 2
  Moral identity 2
  Moral judgment 3
The Goal Concept 4
  Identity goals – Symbolic Self-Completion Theory 5
  Mindset Theory of Action Phases 6
Research Paper I: A Goal Perspective on Moral Identity 7
Research Paper II: The Pursuit of Multiple Simultaneous Identity Goals Depends on the Multifinality of Available Means 13
Research Paper III: Controlling Emphasis Effects on Moral Judgment 15
General Discussion 17
  Implications 18
  Future Directions 19
Summary and Conclusion 21
Research Paper I: A Goal Perspective on Moral Identity 23
  Abstract 24
  Introduction 25
  Moral Identity 26
    The Self Model of Moral Identity 27
    The Social Cognitive Theory of Moral Identity 32
    The Moral Self as an Identity Goal 39
Emerging Research Questions 48
  Cultural Influences 49
  The Moral Self and Other Identities 59
  Motivational Dynamics 66
Summary and Conclusion: The Goal Perspective 69
Experiment 3: The Moderating Role of Mindsets 133
and Visual Attention
Method 133
Results 135
Discussion 137
General Discussion 138
Conclusion 94
References 141
Record of Achievement 173
List of Figures

Figure 1. An example of a goal system. ................................................................. 78
Figure 2. Goal systems investigated in Research Paper II. ............................. 81
Figure 3. Experiment 1: Donations in € by experimental conditions. .............. 86
Figure 4. Experiment 1: Donations in € by experimental conditions and commitment to being a psychotherapist. ................................................................. 87
Figure 5. Experiment 2: Change in the sense of belonging to the Turkish culture after the experimental manipulation. ......................................................... 95
Figure 6. Experiment 2: Change in the sense of belonging to the German culture after the experimental manipulation. ......................................................... 95
Figure 7. Experiment 2: Difference between the desire to belong the Turkish culture and current sense of belonging. ......................................................... 96
Figure 8. Experiment 2: Difference between the desire to belong the German culture and current sense of belonging. ......................................................... 97
Figure 9. Experiment 2: Symbolization effort, indicated by the number of voluntary tasks. ...................................................................................................................... 98
Figure 10. Mean endorsement of the consequentialist options as a function of emphasis and mindset condition in Experiment 1 ............................................. 122
Figure 11. Example of a picture used for the moral dilemma task in Experiment 2. 127
Figure 12. Sequence of screens in a trial in Experiment 2. ................................. 128
Figure 13. Proportion of endorsement of the consequentialist options by experimental conditions in Experiment 2. ................................................................. 130
Figure 14. Distribution and density of fixations for each dilemma separated by mindset conditions in Experiment 2. ................................................................. 132
Figure 15. Mean endorsement of the consequentialist options by experimental conditions in Experiment 3. ................................................................. 137
List of Tables

Table 1. Mixed linear model estimating moral judgments in Experiment 1 .......... 121

Table 2. Mixed linear logit models estimating moral judgments and focus on means in Experiment 2 .......................................................................................................... 129

Table 3. Mixed linear models estimating moral judgments in Experiment 2 .......... 136
Introduction

Synopsis

There is no consensus in regards to what constitutes a moral or immoral act (Heinzelmann, Ugazio, & Tobler, 2012). Yet morality is almost ubiquitous. Moral rules can be found across widely different cultures (Shweder, Much, Mahaputra, & Park, 1997) and thinking of oneself in moral terms is essential for self-construal (Strohminger & Nichols, 2014). At the most basic level, human behavior, and moral behavior in particular, is dependent on goals (Kruglanski & Köpetz, 2009). In the present thesis goal processes influencing how people think about themselves in moral terms and about moral choices are investigated.

Early theories of moral psychology defined human morality as a product of moral judgments derived from conscious reasoning (e.g., Kohlberg, 1981). It has been shown, however, that there is often a discrepancy between moral judgment and moral behavior, referred to as the moral judgment-action-gap. The degree to which morality is integrated into an individual’s identity may explain this gap: according to moral identity theories, moral judgments motivate behavior only when they are self-relevant (Blasi, 1983, 1984; Walker, 2004). The research presented in the present thesis builds on this framework and combines it with theories of goal pursuit. The first two research papers explore moral identity as an identity goal and the third research paper investigates how cognitive orientations associated with goal striving influence moral judgment.

In the first research paper, a theoretical framework capturing the moral self as an identity goal is developed and compared to two moral identity theories: the Self Model (Blasi, 1984, 2005) and the Social Cognitive Theory of moral identity (Aquino & Reed, 2002; Lapsley, 2016). In the second research paper, the joint effect of two simultaneously active identity goals on behavior is tested. In the third research paper, I examine how the effects on incidental emotional information in moral dilemmas can be controlled by cognitive orientations elicited by the phases of goal pursuit (i.e., mindsets; Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016). Further, the cognitive processes underlying these mindsets are explored.

The aim of this synopsis is to integrate the three papers into a common framework and to discuss how each paper contributes to the idea that moral judgment and behavior are goal-driven processes. To this end I will proceed as follows: First, theoretical stances in moral psychology relevant for the thesis will be presented. This section will focus on moral identity and moral judgment. Next, I will introduce the
goal concept in motivational science. In this section, I will provide an account of Symbolic Self-Completion Theory (Gollwitzer, Marquardt, Scherer, & Fujita, 2013; Wicklund & Gollwitzer, 1982) which deals with identity goals, the central topic of the first two research papers. The process of goal pursuit will be discussed in the framework of Mindset Theory of Action Phases (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016) which will be taken up again in Research Paper III. In the subsequent section, I will present a summary of each of the three research papers of the present thesis. Finally, in a general discussion, implications for moral psychology and motivational science are pointed out.

**A Brief Review of Moral Psychology**

Morality is fundamentally evaluative. One can distinguish two major classes of normative moral theories: those concerned with the evaluation of acts (e.g., classic utilitarianism, Kantian ethics) and those concerned with the evaluation of character (i.e., virtue ethics). This distinction has its roots in moral philosophy (Driver, 2013), but it also reflects moral psychological schools and it is a useful categorization for the topics of the present thesis. Act evaluation and virtue ethics are not mutually exclusive. Rather, the approaches are complementary in describing human morality at different levels. In psychology, personality theories and moral identity theories map onto virtue ethics while models of moral judgment, decision making, and behavior correspond to act evaluation as descriptive accounts (Lapsley, 2016). In the following section, I will provide an overview of current moral identity theories, and then I will summarize the important literature on moral judgment.

**Moral Identity.** Recent theories have attempted to elucidate the gap between moral judgment and behavior with the concept of moral identity (Jennings, Mitchell, & Hannah, 2015). Identity in this context is defined by two properties termed by Jennings et al. (2015) as the having side and the doing side of identity. The having side refers to individual characteristics of a person and the doing side refers to the individuals’ agency (i.e., intentionally regulating one’s behavior). Moral identity specifically reflects how important moral concerns are for an individuals’ sense of self. According to Damon (1984), moral judgments only have behavioral implications for an individual if the moral beliefs represented by them are part of the individual’s identity. Along this line of reasoning, one could argue that moral judgments need to be self-relevant to motivate behavior.
Currently, the psychological discourse about moral identity builds on two major theories: the Self Model (Blasi, 2005) and the Social Cognitive Theory (Lapsley, 2016). According to the Self Model, human morality is rooted in reasoning. Through reasoning individuals gain moral knowledge, and based on that knowledge they can integrate morality into their identity. The degree of integration is an individual difference. According to the model, whether individuals act in accordance with their moral values depends on three factors: the degree to which morality is integrated into their identity, an assessment of personal responsibility in the situation, and the individuals’ self-regulation abilities.

In the Social Cognitive Theory, moral identity is conceptualized as a self-schema, a memory network that connects the various self-relevant components related to morality. These components can be factual knowledge, emotional association, attitudes, or even behavioral scripts. The schema concept allows for nonconscious processes and automatic responses. In contrast to the Self-Model, the Social Cognitive Theory emphasizes situational factors. Moral behavior is explained as an interaction of person and environment.

In the present thesis, I add to these theories by proposing an integrative view on the moral self from the perspective of motivation psychology. In this new theoretical stance, I conceptualize moral identity as an identity goal. I review evidence in support of goal striving processes related to moral identity (Research Paper I) and present two experiments testing the interaction of simultaneously active identity goals (Research Paper II).

**Moral judgment.** Early psychological theories of moral judgment dealt primarily with the controlled processes involved in moral reasoning (Kohlberg, 1971; Piaget, 1932). More recent research has shifted the focus to include intuitive processes in models of moral judgment (Cushman, Young, & Greene, 2010; Greene, 2017). One of the most widely used paradigms to investigate the role of reasoning and intuition in moral judgments is based on a thought experiment called the trolley dilemma (Foot, 1983). In the standard trolley dilemma, the following situation is described: Five people are on a railway track and a trolley is heading toward them. If you do nothing, the trolley will hit these people and kill them. The only way to avoid the death of these five people is to sacrifice another person’s life. In such cases two moral principles are at odds: a deontological ethic based on fundamental principles (e.g., it is wrong to kill another person) and a consequentialist ethic based on the outcome (e.g.,
Synopsis

one person dying is better than five people dying). Trolley-type dilemmas can be defined as any sacrificial moral dilemma that pits a consequentialist option against a deontological option.

Answers to trolley-type dilemmas have been shown to correspond to distinct neural patterns (e.g., Greene, 2018; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Hutcherson, Montaser-Kouhsari, Woodward, & Rangel, 2015; Schirmann, 2013; Shenhav & Greene, 2014). Furthermore, a dual dissociation of deontological versus consequentialist judgments on the one hand and intuitive versus reflective processes on the other hand has been reported (Cushman et al., 2010). While other methods may be needed to test more basic properties of moral judgments, trolley-type dilemmas have become a standard method of moral psychology (Greene, 2018).

In the present thesis, I use trolley-type dilemmas to investigate whether an emotional emphasis of consequences can affect moral judgments in a directed way. In addition to providing evidence for emphasis effects, I also test how mindsets (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016) control emphasis biases in moral dilemma judgments (Research Paper III). The mindsets investigated in the present thesis are part of the goal striving process.

The Goal Concept

According to Moskowitz and Grant (2009), most human behavior is at the most basic level controlled by goals. A goal is a mental representation of a positively or negatively valenced future state (Custers & Aarts, 2005) that the individual pursuing the goal is committed to attain or avoid (Oettingen & Stephens, 2009). Being committed to a goal means that the pursuer of the goal is determined to act toward that goal (Oettingen, 2012). In memory, goals are represented as part of organized knowledge structures in which different goals are arranged in a goal hierarchy. The goal hierarchy ranges from relatively abstract, high-level goals (e.g., identity goals) on the one end to relatively concrete, low-level means (e.g., behavioral intentions) on the other end. Subgoals function as means for higher level goals. The various goals are connected via associative links. Furthermore, other goal-relevant memory content can be linked within a goal system (Kruglanski & Köpetz, 2009). The strength of the links varies between goals and depends on the constellation of goals and subgoals. The constellation of multiple means associated with the same goal is referred to as
equifinality. Goal systems with multiple goals served by the same means are referred to as multifinal. A higher degree of both equifinality and multifinality is negatively related to the strength of the link between means and goals (Kruglanski, Shah, Fishbach, & Friedman, 2018; Zhang, Fishbach, & Kruglanski, 2007). Generally, such goal systems follow the same rules as other memory structures (i.e., if an element of the system is activated, activation can spread to excitatorily linked elements especially when links are strong).

An important difference between goals and other knowledge structures is that goals have motivational properties. Activation in associative networks depends primarily on the strength of the links and the recency of encountering elements of the network. In addition, activation in a goal system depends on the state of goal pursuit. Spreading activation effects can be expected while a goal is actively pursued. Once the goal reaches its endpoint, such effects should cease (Kruglanski & Köpetz, 2009).

**Identity goals – Symbolic self-completion theory.** Identity goals are defined as goals aimed at attaining a lasting personal quality (i.e., a self-definition; Wicklund & Gollwitzer, 1982). An individual pursuing an identity goal does so by seeking out indicators of having attained the desired identity (Ledgerwood, Liviatan, & Carnevale, 2007; Marquardt, Gantman, Gollwitzer, & Oettingen, 2016; Moskowitz, Li, Ignarri, & Stone, 2011). These indicators are referred to as *symbols* and can be physical objects as well as social cues and behaviors signaling possession of the desired identity.

Individuals who are committed to an identity goal and do not possess the relevant symbols of that goal are in a state of identity goal incompleteness. Incompleteness is experienced as an unpleasant tension that motivates individuals to self-symbolize. Symbols are interchangeable within the domain of an identity goal. Thus, individuals can compensate for the loss of a goal-relevant symbol with other means signifying possession of the identity aspired to. Once enough symbols of the identity are acquired, the individual enters a state of identity goal completeness. The experienced tension subsides and the motivation to act toward the goal is reduced until the individual experiences the next setback related to their identity goal (Gollwitzer et al., 2013).

In the first research paper, we review the literature on moral identity and goals and propose an integrative model that explains the moral self as an identity goal. In the model, we propose that moral identity is part of a goal system in which multiple
goals can interact. In the second research paper, we investigate the joint effect of simultaneously active identity goals. This research builds partly on Symbolic Self-Completion Theory.

**Mindset Theory of Action Phases.** According to Mindset Theory of Action Phases (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016), a goal-directed action consists of four sequential phases. In each phase, the agent must deal with a specific challenge. This is aided by phase-specific cognitive orientations (i.e., mindsets) that facilitate overcoming the challenges. Phases are: 1) the *pre-decisional* phase, concerned with goal setting, 2) the *pre-actional* phase, concerned with finding opportunities for goal-directed action and planning their implementation, 3) the *actional* phase, concerned with the execution of goal-directed action, and 4) the *post-actional* phase, concerned with evaluating the outcome of an action. Of particular interest for the present thesis are the first two phases and their corresponding mindsets, because the tuning of attention and processing is drastically different in these phases.

In the pre-decisional phase, the agent must decide which goal to pursue. The corresponding mindset is called the *deliberative mindset*. In this phase, the agent has not yet committed to a goal and must consider the feasibility and desirability of the potential goals. Therefore, impartial processing of all available information is adaptive. In line with this, it has been shown that processing in a deliberative mindset is open-minded regarding new information (Fujita, Gollwitzer, & Oettingen, 2007), desirability and feasibility are judged realistically (Bayer & Gollwitzer, 2005), and attention is undirected and broad (Büttner et al., 2014; Fujita et al., 2007).

Once agents have committed to a goal, they enter the pre-actional phase. The challenge in this phase is no longer selecting the optimal goal of a range of goal candidates but rather finding opportunities to act, getting started, and staying on track. Consequently, not all available information is relevant in the pre-actional phase. Rather, the task is supported by an implemental mindset which is characterized by a narrow breadth of attention and selective information processing (Büttner et al., 2014; Fujita et al., 2007).

Mindsets can carry over from the task that elicited them to an unrelated subsequent task. We use this property in Research Paper III to investigate how the influence of incidental emotional information in moral dilemmas can be controlled with a deliberative versus an implemental mindset. We predict that the open-minded
attention and processing associated with the deliberative mindset enhances the impact of incidental information, whereas the focus and narrower attention associated with the implemental mindset can be expected to shield decision makers from goal-irrelevant influences.

**Research Paper I: A Goal Perspective on Moral Identity**

Being moral is a core quality of personal identity (Strohminger & Nichols, 2014). Moral identity is defined as a system of various psychological processes and attributes that determines self-construal in moral terms and regulates behavior (Jennings et al., 2015). The major psychological theories describing this construct are the Self Model (Blasi, 2004, 2005) and the Social Cognitive Theory of moral identity (Aquino & Reed, 2002; Lapsley, 2016).

In Research Paper I, we present a review of the current literature on moral identity and propose a goal perspective as a synthesis of moral identity theories and identity goal research. We first outline the Self Model and the Social Cognitive Theory, then we describe how moral identity would be conceptualized as an identity goal. We discuss emerging research question and propose how each of the three models would tackle these question.

**The Self Model.** The Self Model stands in the tradition of cognitivist accounts of morality (e.g., Kohlberg, 1981) where morality is rooted in moral reasoning. According to the Self Model, for an action to be morally good or bad it must be the consequence of a reasoned moral judgment (Blasi, 1983). Likewise, individuals construct their moral identity based on moral reasoning. Yet, at the same time moral identity has deeply motivational properties and allows the individual to experience agency (Blasi, 1999). According to the Self Model, moral beliefs derived from reasoned judgment have motivational force if they are integrated into a person’s sense of identity (Blasi, 1995). The degree of integration is an individual difference labeled moral identity centrality. In addition, according to Blasi (1983) moral judgments are composed of two parts: a judgment about universal right and wrong and a judgment about the individual’s personal responsibility. Higher moral identity centrality leads to judgments of more personal responsibility. A need for self-consistency drives individuals to act in accordance with the moral values incorporated in their identity. If individuals fail to do so, this self-betrayal is experienced as guilt (Blasi, 1993).
Blasi (2005) defined three essential virtues that determine the makeup of a person’s moral identity and its impact: First, people’s desire for moral ends varies between individuals. Moral desire as a basic moral motivation is a prerequisite for the formation of moral identity. Second, the need for self-consistency is labeled integrity. Third, the term willpower is used as a collective name for the self-regulatory mechanisms that aid goal-directed behavior.

Thus, according to the Self Model the process of moral behavior begins with moral understanding on the basis of reasoning, upon which moral identity is constructed. If individuals feel responsible in a situation and have a high need for self-consistency, they are motivated to act on their moral beliefs. Willpower helps to resist temptations and overcome external obstacles (Bergman, 2004).

Importantly, moral identity has been shown to predict a wide range of behaviors typically deemed moral, including charitable donations (Aquino & Reed, 2002, Study 6), civic engagement (Younis & Yates, 1999), prosocial volunteering (Black & Reynolds, 2016), honesty (Reynolds & Ceranic, 2007; Rua, Lawter, & Andreassi, 2017), sportsmanship (Kavussanu & Ring, 2017; Kavussanu, Stanger, & Ring, 2015; Sage & Kavussanu, 2010), and even exceptional heroism (Colby & Damon, 1992; Frimer, Walker, Dunlop, Lee, & Riches, 2011; Monroe, 2003; Oliner & Oliner, 1992).

Two important limitations of the Self Models lie in the scope of the model. First, in the Self Model morality is defined as the result of higher order cognition (i.e., reasoning). This is not problematic if one merely aims to describe the relation of reasoned judgment and behavior. However, much of human everyday moral behavior is governed by non-conscious, automatic processes and moral intuitions (Sinnott-Armstrong, Young, & Cushman, 2010). Such behaviors are not addressed by the Self Model. Second, situational influences affect to what degree moral identity is translated into behavior (Leavitt, Zhu, & Aquino, 2016), but such interactions of identity and context are not explained in the Self Model.

**The Social Cognitive Theory of moral identity.** The Social Cognitive Theory of moral identity aims to explain automatic behavior as well as situational influences (Lapsley, 2016). According to the Social Cognitive Theory, actions are only meaningful within their situational context. Therefore, the stability of identity is understood as an interaction of the situation and the person (Bandura, 1999).
According to the Social Cognitive Theory, moral identity is represented as a self-schema, a memory structure which facilitates the effective management of information in long-term memory (Aquino, Freeman, Reed, Lim, & Felps, 2009). Schemata are organized around a core concept (e.g., the moral self) and link relevant memory content to that concept. A schema can contain explicit, abstract knowledge, autobiographical memories, affective associations, behavioral responses, and other relevant information. The links in a schema are associative and guide cognition and behavior, especially in complex situations (Fiske & Linville, 1980). If one element of a schema is activated, this activation can spread to the other elements. The selection of a schema can, therefore, be stimulus driven. Once a schema is active in working memory, further processing and response selection is determined by the active schema (Markus & Sentis, 1982). Schemata are a highly efficient means to navigate situations where either very little or a large amount of information is available. In information-scarce situations, schemata fill in the blanks with information from the experience of the agent. In situations where information is abundant, schemata provide a filter and an interpretation (Markus, 1977). Self-schemata refer to schemata about a self-relevant domain (e.g., I as a moral person). Not all self-schemata can be active at the same time. The subset of active schemata is referred to as the working self-concept. Schemata are to a degree malleable and apt to change. However, those schemata that contain core values (e.g., moral identity) are typically relatively stable (Janoff-Bulman, 1989). Some self-schemata may be chronically accessible. The degree to which this is the case for moral identity represents moral identity centrality (Aquino & Reed, 2002), which can be seen as parallel to moral identity centrality in the Self Model.

In support of the schema account of moral identity, Narvaez (1998) has shown that the accessibility of moral schemata predicts how well children understand moral narratives. Further, the spreading activation effect of moral content predicted by the Social Cognitive Theory has been demonstrated (Lapsley & Lasky, 2001) and studies assessing implicit moral self-schemata have found that implicit moral identity is related to honesty (Perugini & Leone, 2009) and an aversion against third-party immoral behavior (Johnston, Sherman, & Grusec, 2013). Activation of moral self-schemata also depends on situational factors. Priming morality via situational cues can increase the accessibility of a moral self-schema, whereas competing schemata or goals can decrease it (Aquino et al., 2009).
The goal perspective of moral identity. We argue that both the Self Model and Social Cognitive Theory miss a crucial point regarding the motivational dynamics of moral identity. It has been shown that individuals tend to compensate for acting inconsistently with their moral self-concept. After moral actions individuals often relax their moral standards and after immoral actions they increase their moral effort (review by Mullen & Monin, 2016). This pattern is not predicted by the Self Model or Social Cognitive theory. However, the findings on compensation are in line with the identity goal concept. We propose a framework based on Symbolic Self-Completion Theory (Gollwitzer et al., 2013; Wicklund & Gollwitzer, 1982) and Goal Systems Theory (Kruglanski et al., 2018; Shah, Kruglanski, & Friedman, 2003) to conceptualize moral identity as an identity goal.

According to Goal Systems Theory, goals are represented in memory, similarly to schemata, as organized knowledge structures. In contrast to schemata, activation of content in a goal system depends not only on associations between elements but also on the state of goal progress (Kruglanski & Köpetz, 2009). When goals are complete, the goal system is inactive and its accessibility is inhibited, whereas when a goal is actively pursued, the accessibility of the corresponding goal system is increased.

An individuals’ commitment to a goal reflects the determination to act toward goal completion (Oettingen, 2012). Commitment to moral identity goals corresponds to moral identity centrality in the Self Model and Social Cognitive Theory. We predict that individuals who are highly committed to being moral are more likely to act on this goal. If individuals are committed to moral identity goals, the process of symbolic self-completion can be expected to drive behavior (Wicklund & Gollwitzer, 1982). The motivational force of moral identity goals is expected to be greater for individuals in a state of identity goal incompleteness than individuals in a state of identity goal completeness. In addition, symbolic Self-Completion Theory predicts that one moral symbol can be exchanged for another. Thus, one would be able to compensate for one moral transgression with an unrelated moral action. These predictions match with findings of moral compensation (Mullen & Monin, 2016).

Emerging research questions. After outlining the Self Model, the Social Cognitive Theory, and the goal perspective, we discuss three major emerging research questions related to moral identity and propose answers from the perspective of each theoretical stance. First, how does the cultural background influence the moral self?
Second, how is the moral self related to other identities and how does it interact with them? Third, when does a moral identity motivate agents to act in accordance with its values and when does it not?

**Morality and culture.** Although similarities exist, moral values vary widely between cultures (Graham et al., 2018). The core structure of the Self Model is compatible with moral pluralism (i.e., the notion that moral values differ between cultures). Yet, the concrete conceptualization of morality in the current version of the Self Model is based on Western societies and may not adequately reflect Eastern cultures (Jia & Krettenauer, 2017). In addition, the Self Model defines morality as rooted in reasoning. Reasoning styles have been shown to differ between cultures and to correlate with culture-specific values and concepts of personal identity (Henrich, Heine, & Norenzayan, 2010). Thus, to be compatible with moral pluralism the Self Model would have to be revised to consider these influences. According to Social Cognitive Theory, self-schemata are culturally learned (Lapsley, 2016). Cultural influences can thus be explained through social learning processes on the level of the individual. From the goal perspective, the effectiveness of identity goal symbolization depends on the social recognition of the symbols (Gollwitzer, 1981). Other people need to register displayed symbols as adequate indicators of possessing the desired identity. This way, one’s social environment (and consequently one’s cultural surroundings) directly affects which identity symbols are chosen. Thus, if in different cultures different symbols are seen as indicators of being moral, individuals from these cultures can be expected to symbolize morality differently.

**Identity interactions.** According to the Self Model, the content of morality does not easily change between situations. The subjective meaning of morality according to the model is relatively stable (Blasi, 2005). Other identities could interfere with the moral self in form of temptations. Fluctuations of subjective moral norms would be treated by the model as nonmoral interference. On the basis of Social Cognitive Theory, three types of identity interactions can be predicted. First, other self-schemata can be in competition with the moral self. Second, moral identity and other schemata could be excitatorily linked. Third, moral identity and other self-schemata could partially overlap without clear boundaries. In the goal perspective on moral identity, we treat moral identity as part of a goal hierarchy in which different goals are organized. Moral identity is conceptualized as a high-level goal that is interconnected with most other identity goals. We propose that the content and
activation of moral identity goals depends on other active goals. This proposition builds on the idea that the self is not unified but rather composed of multiple subidentities (i.e., self-complexity; Linville, 1987). In this framework, stability and malleability of moral identity content depend on the overlap of the moral values that are integrated into an individuals’ various identity goals. This also implies that the meaning of morality is not stable within a person but changes as a function of active goals. Numerous studies support this proposition, showing how an individuals’ moral behavior and judgment depend on other actively pursued identities (Babcock, Loewenstein, Issacharoff, & Camerer, 1995; Kavussanu, Boardley, Sagar, & Ring, 2013; Marquardt et al., 2016; Zimbardo, 2004).

Motivational dynamics. The Self Model cannot explain motivation dynamics such as compensating for past moral transgressions or indulging in immoral behavior after virtuous acts. According to the Self Model, moral Identity drives moral consistency (Blasi, 2005). In the Social Cognitive Theory, moral compensation has been explained as the result of different types of goals aimed at self-improvement or self-presentation (Lapsley, 2016). The goal perspective provides the most coherent answer to the question: ‘When does moral identity motivate behavior?’. Moral identity goals lead to moral behavior while they are active (i.e., the agent is in a state of identity goal incompleteness); once enough evidence for the possession of a moral identity has been acquired (i.e., the agent is in a state of identity goal completeness) no more effort is needed and moral standards are relaxed.

Conclusion. In Research Paper I, a goal perspective is developed as a theoretical framework for moral identity. This framework is a synthesis of other models describing the moral self (i.e., the Self Model; Blasi, 2005, and Social Cognitive Theory, Lapsley, 2016), Goal Systems Theory (Kruglanski et al., 2018; Shah et al., 2003), and Symbolic Self-Completion Theory (Wicklund & Gollwitzer, 1982). The central propositions of the goal perspective are: 1) Moral identity is represented as an identity goal. The commitment to this goal determines the relative importance of morality for an agent. Committed individuals strive toward moral identity completeness. Incomplete individuals are more likely to seek out moral symbols (e.g., act morally) than incomplete individuals. 2) As part of a goal system, moral identity interacts with other goals. The content of the moral identity goal depends at least partly on other active goals that overlap with moral identity. 3) Goal systems vary between individuals. Therefore, the moral content derived from
overlapping goals constitutes an individual difference. Overall, the goal perspective emphasizes the self-regulation aspect of moral identity. Its focus is on moral action and dynamic motivational processes.

**Research Paper II: The Pursuit of Multiple Simultaneous Identity Goals Depends on the Multifinality of Available Means**

A prediction derived from the goal perspective presented in Research Paper I is that the configuration of the active goal system shapes an individuals’ effort toward symbolizing moral identity. On the one hand, if other active goals are in conflict with the moral identity goal, one can expect interference of these goals and a decreased symbolization effort toward the goal of being moral. On the other hand, if the active goals are compatible, then goal synergy and increased effort can be expected. We hypothesize that the interference versus synergy of identity goals depends on whether the goals share available means for symbolization. We tested this hypothesis in Research Paper II.

**Identity Goals in Interaction.** According to Goal Systems Theory (Kruglanski et al., 2018; Shah et al., 2003), goals are organized in hierarchical systems. Goals on a lower hierarchical level are linked as means to hierarchically higher level goals. This way, a means can be associated with multiple goals, a configuration termed *multifinal*. Studies have shown that individuals prefer multifinal means over unifinal means (serving only one goal) as long as multiple goals are active (Chun, Kruglanski, Sleeth-Keppler, & Friedman, 2011; Köpetz, Faber, Fishbach, & Kruglanski, 2011). Active background goals can interfere with a currently pursued focal goal because they take up attentional, cognitive and motivational resources (Emmons, King, & Sheldon, 1993) unless a potential goal conflict can be resolved by the use of multifinal means (Köpetz et al., 2011). Accordingly, we predict that striving toward a moral identity goal is hampered if at the same time a competing goal is active and no multifinal means is available that also serves the moral identity goal.

This prediction was tested in Experiment 1. For the experiment, students enrolled in psychology were recruited. Identity goal incompleteness was induced for the goal of becoming a psychotherapist. This goal is a common professional identity goal among psychology students. Importantly, it has a moral and a competence subgoal. Activation of these subgoals was manipulated between participants. The participants had either the opportunity to symbolize (i.e., deactivate the corresponding
identity goal) morality, symbolize competence, or no opportunity for goal relevant symbolization. This procedure led to three experimental conditions: 1) an inactive moral identity goal and an active competence identity goal, 2) an active moral identity goal and an inactive competence identity goal, and 3) active moral and competence identity goals. After these conditions had been established, the participants received 2 € as compensation for taking part in the experiment. They could keep this money either for themselves or donate it to a charitable organization. Participants with an active moral identity goal donated more money than participants with an inactive moral identity goal, but only when no competence goal was simultaneously active. This effect was moderated by the participants’ commitment to the higher order goal of becoming a psychotherapist. The results of the first experiment provide suggestive evidence in support of the goal perspective on moral identity. The impact of the moral identity goal on behavior depended on another identity goal (i.e., becoming a psychotherapist) that was linked to moral identity in the active goal system. An additional competence identity goal that shared no available means with the moral identity goal reduced moral symbolization effort. A limitation of this experiment is that it lacks an experimental condition where both identity goals are inactive. We expect the same pattern of behavioral results as in the condition with an inactive moral identity goal and an active competence identity goal in such a condition.

Goal conflict can be resolved if multifinal means are found (Köpetz et al., 2011). Based on this, we propose that concurrently active identity goals do not interfere with each other if multifinal means are available. We tested this hypothesis in a second experiment by simultaneously activating two cultural identity goals of bicultural individuals. Identity goal incompleteness versus completeness of cultural identity goals was manipulated for Turkish-German biculturals. The feeling of belongingness and the desire to belong to the respective cultures was measured afterward. Participants reported that they felt as if they belonged less to their culture and a stronger desire to belong to it if incompleteness in that culture was induced. This effect occurred independently for the Turkish and German culture. At the end of the experiment, the participants could answer additional questions based on cultural dimensions (Hofstede, 1983, 2011). This task was designed as a multifinal means and participants could symbolize both the German and the Turkish identity goals by answering questions. Participants with an incomplete identity worked on more questions. Importantly, no interference occurred when both identity goals were active.
Instead, participants with active Turkish and German identity goals answered the most questions and participants for whom both identity goals were inactive answered the fewest questions.

**Conclusion.** Taken together, the results of the two experiments indicate that identity goal striving in a goal system is the product of active goals on a similar hierarchical level and the available means. Typically, identity goals compete for resources. This leads to goal interference if the goals cannot be pursued with shared means. If means to symbolize both goals overlap, such conflict can be avoided. To our knowledge, this study is among the first to systematically test the interaction of multiple identity goals in the framework of Symbolic Self Completion Theory. As such, it is only a first step in the exploration of the interplay of multiple identity goals and points to the need for future replication and expansion. A limitation of the present experiments is that a separate goal system was explored in each experiment. To draw strong causal conclusions, future research should test the motivational and behavioral effects of the different goals system configurations within one experiment.

**Research Paper III: Controlling Emphasis Effects on Moral Judgment**

In the third research paper, we investigated how the cognitive consequences of goal striving influence moral judgments. A model for the goal striving process is provided by Mindset Theory of Action Phases (Gollwitzer & Keller, 2016). It has been shown that the breadth of attention and focusing differs systematically between the deliberative and the implemental mindset. The deliberative mindset is characterized by undirected, broad attention and open-minded processing of information, whereas the implemental mindset is characterized by focused, narrow attention and selective processing (Büttner et al., 2014; Fujita et al., 2007). We hypothesize that, due to these properties, individuals in a deliberative mindset are more prone to be influenced by incidental emotional information in moral dilemmas. In contrast, implemental mindsets are expected to shield decision makers from such influences.

Incidental emotional information is operationalized as an emphasis on the consequences of either the deontological or the consequentialist option in trolley-type dilemmas (Lotto, Manfrinati, & Sarlo, 2014). This allows us to test the association of affective content and judgments of moral dilemmas. We show across three experiments that emotionally emphasizing content in trolley-type dilemmas biases
moral judgments in a directed way. Emphasizing the harmful consequences of a consequentialist option decreases consequentialist preferences, whereas emphasizing the harmful consequences of a deontological option increases consequentialist preferences. We refer to this effect as emphasis bias.

To what extent the emphasis bias occurred depended on the current mindset of the participants. We experimentally activated a deliberative versus implemental mindset in Experiments 1 and 2. The emphasis bias was enhanced by a deliberative mindset and diminished by an implemental mindset. We propose that the increased focus in the implemental mindset contributes to this effect.

Fujita et al. (2007) have shown that individuals in a deliberative mindset recall incidental, task-irrelevant information better than individuals in an implemental mindset. This finding is congruent with the prediction that attention and processing is more selective in the implemental mindset. In line with this reasoning, Büttner et al. (2014) have shown that the visual attention of individuals in an implemental mindset is more focused on central objects, whereas the gaze of individuals in a deliberative mindset is more evenly spread.

Building on this work, we assess and analyze gaze data of individuals looking at schematic pictures of trolley-type dilemmas in Experiment 2. Conceptually replicating the findings of Büttner et al. (2014), we show that the gaze fixations of participants in the implemental mindset were indeed more focused than the fixations of participants in the deliberative mindset. Using pictures of trolley-type dilemmas allowed us to assess viewing patterns of meaningful content. Trolley-type dilemmas are especially suited for the exploration of mindset effects because all trolley-type dilemmas share the same core structure. They can be constructed to contain a visually presentable means (e.g., a lever). Our Experiment goes beyond previous work by showing that individuals in the implemental mindset focused their attention more on goal-relevant means compared to individuals in the deliberative mindset.

In a third experiment, we test whether planning to think thoroughly about the moral judgment works as an alternative method to control the emphasis bias. Previous work has shown that planning thorough reflective thinking can improve decision quality (Doerflinger, Martiny-Huenger, & Gollwitzer, 2017). However, the emphasis bias was present irrespective of whether participants planned thorough reflection or spontaneous decision making.
Conclusion. The present research shows that incidental emotional content affects moral judgments. We have shown across three Experiments that a negative emotional emphasis can bias moral judgments toward an increased or decreased preference for the consequentialist option depending on whether the emphasis targets the consequences of the deontological or the consequentialist option. This finding is an important addition to research on moral dilemma judgments because it highlights that contrary to predictions of prominent models (e.g., Cushman et al., 2010) more emotional situations can lead to a preference of a consequentialist option.

Importantly for the overall theme of this thesis, the emphasis bias was moderated by the currently active mindset. Stronger emphasis effects were found for participants in a deliberative mindset, while the emphasis bias was weaker for participants in the implemental mindset. Mindsets arise in the course of goal pursuit (Gollwitzer & Keller, 2016). Understanding how such action phase related cognitive procedures impact moral judgments can lead to better predictions about moral reasoning. Finally, the study provides new insights into the cognitive-attentional consequences of mindsets. We replicate earlier findings on attention in deliberative versus implemental mindsets (Büttner et al., 2014). We show in addition to an increased focus in the implemental mindset that this focus is directed at goal relevant means.

General Discussion

Normative moral theories can be classified as virtue ethics and action evaluation. A psychological construct related to virtue ethics is moral identity, and descriptive theories of moral judgments correspond to action evaluation. The three research papers reported in the present thesis demonstrate that goal striving processes are highly relevant for both of these psychological counterparts of normative theories. Moral identity has properties of identity goals and the mindsets evoked during goal pursuit control our susceptibility to incidental information during moral judgment. In the following section, I will discuss implications of these findings for research on moral identity, identity goal theories, moral judgment, and Mindset Theory of Action Phases.
Implications

The evidence reviewed in Research Paper I points toward the conclusion that goal processes are necessary to explain the motivational dynamics involved in action control by moral identity. We propose that conceptualizing moral identity itself as an identity goal is a suitable model to explain how the moral self interacts with behavior.

In line with established models and theories describing moral identity (Blasi, 2005; Lapsley, 2016), the question for the moral self is: ‘Who am I as a moral person?’. Following the new theoretical perspective presented in this thesis, one might think of moral identity as much more dynamic. If moral identity is understood as a goal, then progress and setbacks in the pursuit of this goal should influence how people think of themselves in moral terms. In this view, a more apt question for the moral self would be: ‘What kind of moral person do I want to be?’.

Obviously, the goal perspective emphasizes situational variability that arises as a consequence of goal pursuit. Cross-situational stability in the model is explained as goal commitment. One could argue that the construct of moral identity goal commitment overlaps with more stable accounts of moral identity. I agree with this argument. However, goal commitment is only part of the equation. I propose that taking the entire goal construct into account, including structure and process, will lead to a fuller understanding of moral identity.

A second proposition is that moral identity goals are part of goal systems. Goal systems represent the interconnected constellation of the various goals a person holds. Within a goal system, active goals can be expected to influence each other’s impact on behavior. For example, the pursuit of a moral identity goal might be hindered by competing goals. Experiment 1 in the second research paper provides an example of such a case: An active competence identity goal reduced symbolization of moral identity. Other interactions are also likely to occur. Some goals may be hierarchically linked as means and ends. We found evidence for this in the first experiment of Research Paper II: The degree to which individuals symbolized their moral identity depended on the commitment to a higher-level vocational identity goal.

The possible means in a situation pose further constraints on identity goal pursuit in situations where multiple identity goals are active. Consistent with Goal Systems Theory (Kruglanski et al., 2013; Kruglanski et al., 2018; Shah et al., 2003), we found that simultaneously active identity goals interfere with each other in
situations where no multifinal means are at hand that would be able to satisfy all available goals (Research Paper I, Experiment 1). In contrast, when a shared means was available goal synergy was observed (Research Paper I, Experiment 2).

Research Paper III addressed how goal striving influences moral judgment. Specifically, we investigated how the cognitive processes associated with action phases moderate the influence of emotional information in moral dilemmas. In regards to moral judgment, a prominent assumption is that consequentialist judgments are strongly aligned with reflective reasoning and deontological judgments are primarily driven by emotion and intuition (Greene, 2009). The results presented in Research Paper III suggest that emotional content in moral dilemmas can favor both consequentialist and deontological judgments. This underscores that emotions are targeted affective states that refer to an object (Colman, 2008). Therefore, negative affect can be attached to a consequentialist option as well as a deontological option. An interesting finding is that the impact of emotional presentation depended on action phase-related mindsets. A deliberative mindset led to moral judgments that were more in line with the emphasis and an implemental mindset led to a decreased effect of emphasis. The second experiment of Research Paper III corroborates core properties of the action phase-related mindsets. Visual attention in the deliberative mindset was undirected, whereas attention in the implemental mindset was focused on goal-relevant means. Narrowing of attention is likely one mechanism by which the implemental mindset protects selected goals from detrimental influences and distractions.

Future Directions

While the research papers in the present thesis rely on the goal concept as a common framework to understand human morality, a broad spectrum of phenomena both from moral psychology and from goal psychology are investigated. The thesis is composed of two parts. The first two research papers are concerned with moral identity and describe moral identity goals based on Symbolic Self-Completion Theory. The third research paper is about moral judgment and explores the influence of mindsets. The research papers open up several possible avenues for future research. I focus here on those future directions I find most interesting.

First, the perspective described in Research Paper I is a foundation for the understanding of moral identity as a goal. At this point, the theoretical ideas presented
in the paper are a framework for future exploration. A central proposition of the paper is that moral identity goals are part of a goal system. As such, the activation and pursuit of moral identity should be dependent on the other goals moral identity is linked to. Evidence for interference and synergy of identity goals is provided in Research Paper II. Furthermore, the second research paper shows that in some instances moral identity functions as a subgoal of another identity goal. The hierarchical relationship of the moral self and other identity goals is likely a fruitful topic for future research. While moral identity is generally a high-level goal, in some instances it may be a means for a higher order goal. Such cases are certainly not unique within a person. For an individual, being moral can be a subgoal of a professional identity goal. At the same time, for this person being a parent can have a moral component. An open question is: Is moral identity stable, or would its content change with the activation of higher order goals? If the latter is the case, then the moral values individuals hold would be much more context-dependent than current moral identity theories would predict. Future studies could systematically test this hypothesis by activating different higher order identity goals and afterward assessing moral judgments and behavior. A related question pertains to the degree that the moral values of a specific moral identity goal (e.g., sportsmanship) can be influenced by subtle situational cues. Moral identity can be activated outside of the agent’s awareness via priming (Aquino, McFerran, & Laven, 2011). I predict, based on the goal perspective, that priming different aspects of morality will lead to qualitatively different moral judgments and different behavior.

Second, Hannah, Avolio, and May (2011) have proposed a self-complexity approach to moral identity. They suggested that moral identity may be composed of several sub-identities based on social roles. Moral content can differ between sub-identities. As an individual moves between social roles, different aspects of their moral self become activated. The goal perspective offers a framework for the structure and mechanisms underlying such phenomena. If moral identity is multidimensional and malleable by situational factors, a challenge for research will be to measure and mathematically model the complexity of moral identity goal systems.

Third, in the third research paper, the interaction of emotional presentation and mindsets on judgments of trolley-type dilemmas have been tested. There is ongoing debate as to whether trolley-type judgments are actually informative of more general moral thinking and real-world decision making (Conway, Goldstein-
Summary and Conclusion

The present research investigated whether people construct their moral identity as an identity goal and how mindsets related to the action phases in goal pursuit influence moral judgment. In Research Paper I, we review current theories of moral identity and add to this our perspective based on the psychology of goals. In Research Paper II, we observed that effort toward a moral identity goal is reduced when a competence goal is simultaneously active and no shared means is available. In this research paper, we also show that two cultural identity goals worked in synergy toward a shared means. In Research Paper III, we found that the influence of emotional emphasis is increased by a deliberative mindset and decreased by an implemental mindset. Taken together, this thesis is a first step toward bringing the psychology of goals into the realm of moral identity and moral judgment. This is likely to be beneficial for both fields. The moral domain is a rich field to test and evaluate theories from motivational science. At the same time, goal theories will likely provide a fuller understanding of moral identity, judgment, and behavior.
Research Paper I

A Goal Perspective on Moral Identity

Johannes T. Doerflinger¹ & Peter M. Gollwitzer¹²

¹University of Konstanz, Germany ²New York University, USA
Abstract
The present article reviews the Self Model (Blasi, 1984, 2005) and the Social Cognitive Theory of moral identity (Aquino & Reed, 2002; Lapsley, 2016), and then introduces a goal perspective on moral identity. The aim is to ready the research on moral identity to taking on a motivation science perspective. From this perspective it is suggested to conceptualize the moral self as a potential identity goal. The question of moral identity thus shifts from “Who am I?” to “Who do I want to be?”, and the motivational mechanisms behind striving to be a moral person are analyzed. We propose that moral identity goals should not be seen in isolation but as part of a larger network encompassing various identity goals. We also discuss three research questions related to moral identity goals. First, how does culture affect moral identity goals? Second, how are moral identity goals related to other identity goals? Third, under what circumstances do moral identity goals motivate moral behavior? Answers to these three questions are given on the basis of the three theoretical positions: the Self Model, the Social Cognitive Theory, and the goal perspective. We see the goal perspective as complementary to other models of moral identity by highlighting the importance of motivational processes.
Introduction

A Goal Perspective on Moral Identity

Morality has been described as a defining feature of people’s self-concept, such that moral concerns are considered to be among the most central parts of personal identity. Strohminger and Nichols (2014) have shown in a series of studies that, according to laypeople’s intuitions, morality is an important core component of the self, more so than any other mental faculty, including autobiographical memory, personality, preferences, and basic cognitive functions. Yet, there is no commonly accepted definition of morality, and moral norms do come in many different forms. The content of what is considered to be moral or immoral does substantially diverge between individuals, groups, and cultures. For instance, Haidt and Graham (2007) have demonstrated that politically liberal and politically conservative people hold different sets of moral intuitions. These authors have argued that, when analyzing morality, a broader spectrum of norms and values should be considered than is typically done in moral psychology (Graham et al., 2013; Graham, Haidt, & Nosek, 2009; Haidt, 2013a). According to Haidt (2007), “morality is about more than harm and fairness.” Surprisingly, most theories of moral identity remain silent about the content of moral identity or rely on implicit assumptions about it.

In this article, we adhere to a goal perspective on moral identity, a theoretical framework of moral identity that is compatible with the moral pluralism perspective (e.g., Graham et al. 2013). Taking a goal perspective, we outline the structural relation of moral identity to other identity components. This is both of theoretical and empirical importance, because understanding how moral identity is related to other identities may be the key to understanding moral pluralism and situational influences on moral judgment and behavior. Furthermore, moral identity bridges moral judgments and moral behavior (Lapsley & Narvaez, 2004), and many studies suggest that personal identity has a deeply motivational function (Gollwitzer, Bayer, Scherer, & Seifert, 1999; Gollwitzer & Kirchhof, 1998; Markus & Nurius, 1987). Thus, in the present article, an account of moral identity from the perspective of motivation psychology is presented, building on the concept of identity goals (Gollwitzer & Wicklund, 1985a).

Our central propositions are: 1) The moral self can be construed as an identity goal. People strive towards realizing their identity goals. Identity goals represent commitments to realize potential future selves. 2) A person can hold multiple identity goals. Identity goals can be interrelated, overlapping, or distinct. 3)
Moral identity goals overlap with most other identity goals. Typically, the moral identity goal is not an independent entity. It is best understood as imbued with content by other identity goals the individual feels committed to.

The present paper is structured in three parts. To make the case for a goal perspective, we first provide a theoretical review of moral identity in the psychological literature and then outline the goal perspective. We discuss Augusto Blasi’s Self Model (Blasi, 1984, 1995, 2005) and more recent social cognitive approaches (Lapsley, 2016; Lapsley & Narvaez, 2004; Shao, Aquino, & Freeman, 2008). In the second part, we will raise various research questions related to moral identity and discuss how each of the three theoretical approaches (i.e., the Self Model, the Social Cognitive Theory, and the goal perspective) might deal with these questions: How does the cultural background of a person influence moral identity? How does the moral self interact with other identities? And, when does activation of moral identity increase or decrease moral behavior? Finally, in the third part of the paper, we will summarize the moral identity goal framework and discuss implications for future research.

Moral Identity

One of the major challenges to cognitive developmental theories of morality (Kohlberg, 1981; Kohlberg & Candee, 1984; Piaget, 1932) is the often occurring discrepancy between moral judgment and moral conduct (Walker, 2004). Much of the early psychological theorizing about morality focused almost exclusively on moral understanding and explicitly reasoned moral judgments. This line of thinking is strongly rooted in continental European moral philosophy exemplified by Kantian ethics (Kant, 1786/1986, 1788/1995). According to Kant only actions that follow willful, conscious, and deliberate considerations can have moral value (Kant, 1788/1995). Building on the Kantian notion that morality is by definition ultimately rooted in rational thought, Kohlberg (1981) developed a stage model of moral judgment. In his model, an idealized trajectory of moral cognition, understood as effortful conscious reasoning about issues of justice, is presented according to which the highest form of moral judgment is reminiscent of a Kantian categorical imperative. Kohlberg, in the philosophical tradition of rationalists like Socrates, Plato, and Kant, was convinced that moral reasoning is the primary motivator of moral action.
However, explicit moral judgment and moral reasoning is at best a modest predictor of actual behavior (Blasi, 1980; Walker, 2004). Furthermore, moral exemplars (i.e., individuals who consistently demonstrate exceptional moral behavior over a long time) don’t differ much from the general population, when it comes to moral reasoning (Colby & Damon, 1992; Walker & Frimer, 2007). The Socratic idea that good reasoning motivates moral behavior thus seems to be empirically unfounded. This problem of moral motivation has been termed the judgment-action gap and moral identity has been proposed as the concept bridging the judgment-action-gap (Blasi, 1983, 1984; Walker, 2004). For example, Damon (1984) argued that for moral beliefs to be translated into moral behavior, morality must be integrated into a person’s understanding of their identity. One can argue that for morality to have motivational force on an individual, morality must be relevant to that person’s sense of self. Following this line of reasoning, any adequate notion of the moral self must be motivational and agentic in nature.

Moral identity, as discussed in the present paper, is to be understood not as a mere trait, but as a motivated and motivating mental faculty. Currently, the most influential theories of moral identity are Blasi’s Self Model (Blasi, 1984, 2004b, 2005) and the Social Cognitive Theory (Bandura, 1991). Both theories are in agreement that static personality traits are insufficient to describe the moral self. Most psychological theories of the self are built on William James’ (1890/2013) account of the self as both an object (Me) and a subject (I). This idea is reflected in contemporary notions of the moral self. For example, Jennings et al. (2015) describe the moral self as composed of a having side comprised of individual dispositions and characteristics and a doing side which agentically regulates behavior. Both Blasi (2004a) and Bandura (1982) emphasize agency tied to the subjective side of the self as one of its core functions, where agency means being conscious of one’s experiences and consciously and intentionally regulating one’s own actions (Bandura, 1999b). Experiencing oneself as a subject means to experience agency as well as self-appropriation (i.e., the feeling that an aspect belongs to oneself and is integrated into one’s identity; Blasi, 2004b).

The Self Model of Moral Identity

The Self Model proposed by Blasi defines morality as fundamentally cognitive (where the term cognitive takes the meaning of involving reasoning).
According to Blasi (1983), only actions following some form of moral reasoning have moral significance (good or bad). It is clear from Blasi’s writing that he sees reasoning and judgment as the essential foundations of morality; any noncognitivist theory of morality would lack the essential quality of its subject (Blasi, 1984). Thus, the major challenge for Blasian moral psychology is to bridge the moral judgment–action gap while retaining the primacy of cognition in moral conduct. Consequently, the first component laid out in the Self Model is a description of cognitive motivation. Cognitive motivation, according to Blasi (1983) can be understood in terms of Piaget’s concept of equilibration (Mischel, 1971): The desire to understand reality motivates cognitive activity. Then, if an agent perceives rational beliefs to be true, they can become reasons for action. In the Self Model, as in most psychological theories of morality, moral norms (e.g., Turiel, 1983) are perceived as universally prescriptive (Blasi, 1985). This means, from the point of view of the agent, that morality is not just about subjective preferences (e.g., “Who do I want to be?”) but rather about objective standards (e.g., “Who should I be?”). Moral motives are thus characterized by a duality between an “ought” perceived by the agent as universally objective and the relevance for the self which is uniquely subjective (Bergman, 2004). This duality has important implications for the concept of cognitive motivation.

**Agency.** First, a moral individual needs to have at least a minimal degree of agency over their actions. The crucial process by which moral agency is enacted is what philosopher Harry Frankfurt called higher-order volition (Blasi, 1999; Frankfurt, 1988a, 1988b). In this framework, desires, needs, and drives concerning direct consequences of an action are first-order volition (e.g., the desire to eat a piece of cake). Higher order volitions are desires about other desires. For example, one might have the desire to eat cake (first-order volition), but at the same time want to lose weight (another first-order volition). This can lead the individual to wish, they wouldn’t desire the cake (second order volition). Higher order volitions, thus, serve to evaluate other volitions, order them hierarchically, approve some, but reject others. By doing so, “we agentically structure our motives and desires (we structure our will in Frankfurt’s language) and begin to establish our identity” (Blasi, 1999, p. 11).

**Responsibility and self-consistency.** Second, moral beliefs (as higher-order volitions) have a motivational impact on an agent, if they are personally important to the agent. This is the case if the individual constructs their identity around moral notions. Cognitive moral reasoning becomes motivational when it is
integrated into a person’s structure of self, in other words into their moral identity (Bergman, 2004; Blasi, 1995). Further, moral judgments are composed of two components: first, an agent makes a deontic judgment about what is morally right or wrong, and then the deontic judgment is followed by a responsibility judgment determining whether the agent is personally responsible for taking the moral course of action. Modelling the moral judgment process as a combination of both a judgment of universal moral right or wrong and one’s individual personal responsibility, has been adopted in a wide range of psychological theories (Kohlberg & Candee, 1984; Rest, Thoma, & Bebeau, 1999; for a critical review see Minnameier, 2013). Importantly, responsibility judgments create a link between moral judgments integrated into one’s identity and one’s behavior. How responsibility judgments are made is a case of individual differences and depends on an agent’s self-definition. The more strongly one’s identity is constructed around moral notions, the more central morality is for the self. The mechanism by which moral identity is translated into judgments of responsibility and ultimately into action rests on the need for self-consistency. Although often described as a drive to reduce tension, for Blasi (1983) the need for self-consistency is a cognitive motive. According to Blasi, self-inconsistency can only be resolved by realigning one’s actions with the self, not the other way around. Behavior that is inconsistent with an individual’s identity results in guilt as an emotional reaction to that inconsistency. According to the Self Model, guilt represents more than a consequence of betraying an abstract rule. For a person with an integrated moral identity, guilt is a consequence of betraying one’s own identity. Experiencing self-betrayal thus requires a sense of agency because only an agentic self can experience responsibility (Blasi, 1993).

**Essential virtues.** A person is never only motivated by moral motives but also by other needs (e.g., self-interest, biological needs, temptations) that can potentially be in conflict with moral motives. Moral behavior, and thus self-consistency, is more likely if a person has effective self-regulatory mechanisms to deal with such conflicting motives (Blasi, 1983). Three such mechanisms have been proposed as essential virtues for individuals with a moral identity: moral desire, willpower, and integrity (Blasi, 2005). Moral desire refers to moral ends like fairness, honesty, courage, or kindness. Willpower in this framework refers to the ability and degree to which a person can exert self-control, avoid temptations, and consistently enact goal-directed behavior. As such it is an amalgamation of various self-regulatory
mechanisms. Integrity refers to the importance of a unified sense of self which drives self-consistency motives. Both willpower and integrity are morally neutral. A person could use their willpower to achieve either moral or immoral goals, and whether integrity leads to moral or immoral behavior depends on the self-concept. However, both virtues (i.e., willpower and integrity) are deemed necessary for moral motivation. They are at the core of moral agency and enable acting on second-order volitions (e.g., the desire to be an honest person).

The Self Model can thus be summarized as follows: Moral understanding (norms, values, and moral beliefs) is derived from moral reasoning. Individuals construct their moral identity based on their moral understanding, integrating morality into their sense of personal identity. Being true to that identity motivates action if the individual judges themselves to be responsible in the given situation. If moral identity is central to an individual’s identity and the person has the necessary self-regulatory capacity, they should be motivated to act in a moral way (Bergman, 2004; Blasi, 2005; Lapsley, 2016).

**Moral identity and moral action.** As others have noted (e.g., Lapsley, 2016; Shao et al., 2008), the Self Model advanced moral psychological theorizing significantly by putting personal identity at center stage. In the model, mechanisms are proposed by which moral cognition motivates moral action – albeit indirectly via the self and the need for self-consistency. Evidence for the influence of moral identity on moral behavior is provided most clearly by studies on moral exemplars. For instance, Colby and Damon (1992) interviewed individuals identified as living a morally exemplary life. Their participants were selected based on criteria provided by twenty-two experts in the fields of moral philosophy, theology, ethics, history, and social sciences. The participants in their study were highly diverse in regards to gender, ethnicity, religion, geographic location, and political orientation. What was typical of the individuals interviewed by Colby and Damon was that their personal and moral goals were closely integrated. What was morally right and personally necessary for these individuals was in fact one and the same, so much so that many exemplars stated that they had no choice but to act the way they did. For these individuals, their self was defined by morality. Similar response patterns were found in interviews with rescuers of Jews during the Holocaust, who often stated that helping people in dire need was an act of necessity rather than heroism (Monroe, 2003; Oliner & Oliner, 1992). Nonetheless, this felt necessity does not seem to
diminish experienced agency. Rather, as shown by Frimer et al. (2011), moral exemplars experience more agency than comparison peers. Other studies with individuals who demonstrate exceptional care for others or their community revealed that for the sense of self of such care exemplars morality is more important than for comparable peers (Hart & Fegley, 1995).

Moral identity also seems to drive everyday moral behavior such as youth and civic activism (Younis & Yates, 1999), voluntarism (Black & Reynolds, 2016), and fair play in sports (Kavussanu et al., 2015; Sage, Kavussanu, & Duda, 2006). Furthermore, self-report measures of moral identity predict frequency of moral behavior (Aquino & Reed, 2002; Hardy, 2006; Stets & Carter, 2012), reduced cheating behavior in experimental settings (Reynolds & Ceramic, 2007; Rua et al., 2017), and academic cheating (Reynolds & Ceramic, 2007; Wowra, 2007). They also correlate with experienced moral elevation after being exposed to moral behavior (Aquino et al., 2011), donation behavior (Aquino & Reed, 2002), reductions in outgroup hostility (Reed & Aquino, 2003), and doping in sports (Kavussanu & Ring, 2017). Psychopathy, a personality attribute commonly associated with immoral behavior, is negatively related to moral identity. Interestingly, this is independent of moral judgments in ethical dilemmas, indicating that not moral reasoning but the self-importance of morality might drive immoral behavior in people with psychopathy (Glenn, Koleva, Iyer, Graham, & Ditto, 2010). Experimental manipulations of the accessibility of moral identity also demonstrate behavioral effects. Priming moral identity leads to more moral behavior in response to other people’s acts of uncommon goodness (e.g., after reading a story about forgiveness; Aquino et al., 2011) and reduced moral disengagement effects (Aquino, Reed, Thau, & Freeman, 2007). Also, individuals cheat less if a situation is framed in a way that implicates the self (Bryan, Adams, & Monin, 2013). A recent meta-analysis of 112 studies demonstrated that moral identity is a significant positive/negative predictor of prosocial and anti-social behavior, respectively, across various behavioral domains and assessment methods (Hertz & Krettenauer, 2016).

**Limitations.** Despite the evidence underscoring the importance of moral identity for moral conduct, the Self Model is limited in several respects. It only covers behavior that is produced by moral reasoning and explicit judgments (Shao et al., 2008). For Blasi, reasoning is at the very heart of morality. However, even from a normative perspective, the dominance of rationalism is far from settled and
throughout history, alternatives have been discussed (e.g., sentimentalism; Hume, 1738/2003). From an empirical perspective, reducing the moral realm to reasoned choices and consequent actions brings with it the disadvantage of missing much of everyday morality. First, nonconscious, automatic processes do account for a large proportion of human behavior and this should be true for moral behavior as well (Bargh & Chartrand, 1999). Second, the intuitive knowledge of right and wrong, heuristics, and automated behavioral patterns seem to guide most of our everyday moral decisions (Sinnott-Armstrong et al., 2010; for an extensive discussion of the role of intuition, heuristics, and automaticity in moral conduct, see Sinnott-Armstrong, 2008). Some researchers even argue that moral judgments are mostly based on gut feelings and moral emotions, while explicit reasoning serves mainly to create post-hoc rationalizations (Haidt, 2001). Importantly, in moral judgment as in moral conduct both reasoning and intuition play an important role (Cushman et al., 2010). Thus, it follows that a comprehensive psychological account of morality must widen its scope to explain these phenomena. A theory of the moral self that restricts itself only to reasoned judgments will always be incomplete.

Another limitation of the Self Model is that it does not address how the situational context affects the impact of the moral self on behavior (Nucci, 2002). Moral identity is certainly not the only component of the self (Shao et al., 2008). In fact, it has been shown that the situational social context shapes individuals’ self-concepts (Leavitt et al., 2016; Markus & Kunda, 1986). If the self is malleable to situational and social influences, the questions for theories of the moral self are thus: Under what circumstances do agents construct their identity in moral terms, in what sense is the moral self stable across different situations, how is the moral self related to other identity components, and how do the individual characteristics of a person interact with the environment?

**The Social Cognitive Theory of Moral Identity**

The Social Cognitive Theory (Bandura, 1985, 1989) applied to moral identity builds on the key elements of the Self Model (summary by Lapsley, 2016). Moral identity is seen as central to the self, but the degree of this is a matter of individual differences. Moral identity is seen as cognitively constructed, being based in the persons processing of relevant information (Shao et al., 2008). The construction and development of moral identity is understood, according to Social Cognitive
The Social Cognitive Theory of Moral Identity

Theory, in terms of a combination of associative and observational learning (Bandura, 1999b). Two major differences between the Self Model and the Social Cognitive Theory must be noted, however. First, in contrast to the Self Model, automatic behavior and implicit cognition are considered to be of high relevance in Social Cognitive Theory (Lapsley & Narvaez, 2004; Narvaez & Lapsley, 2005). Second, social cognitive theorists argue that in explaining the self, special attention must be paid to situational influences (Lapsley, 2016). The context in which behaviors occur is crucial because actions acquire meaning only in a given situation; when abstracted from that context behavior is often ambiguous (Bandura, 1999b; Wiley & Alexander, 1987). In this framework, consistent behavior and stable dispositions can best be understood as the interaction of person and situation variables (Shoda, 1999). Moral identity is thus captured in Social Cognitive Theory as a self-schema in which moral values, traits, goals, behavioral scripts, and morally relevant situational cues are encoded (Aquino et al., 2009).

Self-schemata. Schemata are memory structures in which information is efficiently selected, stored, and organized. Each schema can be understood as a network of information related to a core concept. The schema determines which new information is congruent or incongruent with the existing pattern. Schemata also determine how new information is processed, and stored information is retrieved from long-term memory. Generally speaking, schemata facilitate the effective use of a memory system, especially when the individual deals with complex situations (Fiske & Linville, 1980). Information stored in schemata is typically a mixture of abstracted content (e.g., the idea of fairness) and specific instances (e.g., the time I shared candy with my sibling). The notion of schemata represents a model for structuring information that can be applied to various domains and different levels of abstraction and complexity. At the most basic level, schemata represent learned categories that can be used in perceptual or cognitive tasks (Sweegers, Coleman, van Poppel, Cox, & Talamini, 2015). The schema concept can be used to describe organized structures consisting of environmental cues, mental processes, and actions, or as models of actions themselves. In this context, schemata are sometimes referred to as scripts (Schank & Abelson, 1977).

Schemata facilitate human behaviors in two ways. They reduce information to generalized meaningful structures if the agent is faced with too much information and a complex environment, and they provide a guiding framework to
operate in if too little information is at hand. In this sense, both bottom-up (stimulus-driven) processes are involved to select the appropriate schema and top-down (schema-driven) processes in turn guide subsequent processing and behavior (Markus & Sentis, 1982). The construction of schemata can be understood in terms of the acquisition of expertise. As experts become knowledgeable and skillful in a certain domain, they represent domain-specific information chunked into meaningful information clusters. For example, expert chess players have been shown to recall constellations of chess pieces better than novices do, because they construe them as larger meaningful units (i.e., schemata), whereas novices might only see individual pieces (Chase & Simon, 1973; Gobet & Simon, 1998). Self-schemata represent expertise in a self-relevant domain (Markus & Sentis, 1982).

According to Markus (1977), self-schemata are the frameworks individuals use to make sense of their past behavior. These frameworks are specific to life domains and constitute a connected set of autobiographic memories, affective reactions, motivational inclinations, behavioral patterns, and thoughts. Within the domain of a given self-schema, information processing about the self in relation to that domain is facilitated. This includes directing attention towards schema-relevant cues, filtering of information, knowledge integration, retrieving related memory content, and directing future behavior. The more experience an individual has in a specific domain, the more interconnected are the elements of the schema (Markus, 1977). Such well-developed schemata are resistant to change and their content is chronically accessible and salient (Lapsley, 2016). Because new experiences are typically interpreted in terms of existing schemata and individuals are likely to seek out situations that are congruent with their self-schemata, schemata are likely to verify themselves (Cantor, 1990). This effect is amplified by the tendency to resist counterschematic information (Markus & Sentis, 1982; Sweeney & Moreland, 1980) and to seek out validation of self-schemata (Gollwitzer & Kirchhof, 1998; Stryker, 1987). However, people are also able to restructure their self-schemata, rejecting unwanted selves and working towards desired, possible selves (Markus & Kunda, 1986; Markus & Nurius, 1986). As human cognitive capacities are limited, not all self-schemata can be active at the same time. The subset that is active at any given moment can be understood as the working self-concept (Markus & Nurius, 1987). Also, not all self-schemata are equally important to any given individual. Especially core values and fundamental beliefs about the self and the world are typically the
The most stable self-schemata (Janoff-Bulman, 1989). In the social cognitive framework, moral identity represents such a central self-schema (Aquino et al., 2009; Walker & Frimer, 2009).

**Research on moral self-schemata.** If the moral self-schema is readily accessible to an individual, then it is central to that person’s identity and can be expected to strongly influence behavior and social information processing (Shao et al., 2008). A highly developed moral identity is, thus, from the social cognitive perspective a chronically accessible self-schema with moral content (Lapsley, 2016). James Rest developed the Defining Issues Test to assess the development of moral reasoning schemata (Rest, 1989). The original test was later revised for practical improvements and higher validity (Rest, Narvaez, Thoma, & Bebeau, 1999). In this newer version, participants are presented with short moral dilemmas followed by a set of issue statements. Respondents are asked to rate the statements in terms of importance for making moral judgments. Each statement is only a short fragment, rather than a complete moral justification. If a fragment activates an existing cognitive schema and the respondent agrees with the statement, they are expected to rate it as important. The schemata underlying the issue statement in the Defining Issues Test correspond to a stage structure. Thus, the test can be seen as a neo-Kohlbergian approach that incorporates social cognitive schemata (Rest, Narvaez, et al., 1999; Rest, Thoma, et al., 1999). Importantly, it has been shown that moral development assessed with the Defining Issues Test is related to the understanding of moral narratives (Narvaez, 1998, 2001), suggesting that the available moral schemata guide information processing.

Further support for the theory that the moral self is represented in terms of cognitive schemata is provided by Walker and Pitts (1998), who showed that people have a prototypical knowledge of moral maturity and excellence. Extending these findings, Lapsley and Lasky (2001) tested the spreading activation effect of knowledge accessibility in moral schemata. They first activated the moral person schemata by presenting subjects with statements about a fictitious person. Then the participants were asked in a recognition test to indicate whether statements were previously presented to them or not. The items in the recognition test included the statements shown before, new schema-congruent statements and new schema-incongruent statements. The false recall rate was higher for the congruent than the incongruent statements.
Moral identity schemata can also include implicit associations and automatic response patterns. If a schema is active, the content in this information network can be expected to be highly accessible. This should also be reflected in measures of cognitive association strength like the implicit association test (IAT). Studies using adaptations of the IAT show that a higher implicit moral identity is related to more honest behavior (Perugini & Leone, 2009), and stronger physiological reactions to moral violations of others (Johnston et al., 2013). Explicit moral identity, on the other hand, was predictive of responses to hypothetical moral dilemmas (Johnston et al., 2013; Perugini & Leone, 2009).

The moral self-schema is a network of all information relevant to morality. Knowledge of both moral as well as immoral conduct are included in such a network. Schema theory (Markus, 1977) suggests that encountering information mapped into a schema increases the current accessibility of the schema. Thus, both encountering moral as well as immoral content should increase the accessibility of moral identity. This seems to work even with subconscious priming procedures. Welsh and Ordóñez (2014) primed participants subconsciously using a sentence unscrambling task that included words related to ethical conduct, words related to unethical conduct, or neutral words. Subsequently, the participants primed with moral and immoral words indicated higher moral identity scores than participants in a neutral control condition; they also categorized ambiguous dilemmas more in moral terms (Study 1), reported less endorsement of unethical behavior, and were more honest when given the opportunity to overreport performance in an incentivized task. Taken together, these studies present converging evidence that the moral self and the meaning of being a moral person are stored schematically.

The situation-identity interaction. Aquino and Reed (2002) built on the Social Cognitive Theory, defining moral identity as a self-schema into which moral trait associations related to justice and care are integrated. Moral identity is not the only self-schema. Accordingly, it is in competition with other self-schemata and its motivational power is influenced by situational factors. The person-situation interaction offers an explanation for when moral identity motivates behavior. The more central morality is to an agent, the higher is the self-importance of moral identity. This is a point of individual differences and represents the accessibility of the moral self-schema. On the other hand, morality can be more or less central and salient
in a situation. Thereby, the context can facilitate or inhibit activation of the moral self-schema (Aquino et al., 2009; Aquino & Reed, 2002; Leavitt et al., 2016).

Aquino et al. (2009) systematically investigated how situational factors influence the accessibility of moral identity. They hypothesized that priming morality (Study 1) influences moral intentions. However, this effect should be moderated by moral identity centrality. Moral identity centrality is defined as chronic accessibility of moral self-schemata. Increasing the accessibility via a prime should affect individuals less for whom moral identity is already highly accessible (i.e., those high in moral identity centrality). Furthermore, the interactive effect of moral identity centrality and moral primes can be expected to affect intentions via accessibility of the moral self-schema. In line with this reasoning, the authors found a moderated mediation effect. The prime increased moral intentions of participants with low moral identity centrality more than participants with high moral identity centrality. This effect was mediated by the accessibility of the moral self assessed with a ranking task for self-descriptions. Conversely, situational factors should also be able to decrease the accessibility of moral identity or increase the accessibility of competing self-schemata (Studies 2 & 3). Providing financial incentives for being dishonest indeed led to more intentions to lie, especially for individuals with high moral identity centrality. Again, this effect was mediated by the current accessibility of moral identity. Priming moral identity counteracted the effect of financial incentives (Study 4). Other studies have found that priming moral identity can alter the current association strength between morality and other concepts and increase one’s concern for others. For example, Leavitt et al. (2016) have demonstrated that priming moral identity weakened the implicit association of the concepts “business” and “ethics”, an association that has been shown to be related to unethical decision making (Reynolds, Leavitt, & DeCelles, 2010).

Beyond subtle situational factors outside of the agent’s awareness, moral identity also interacts with overt and consciously experienced features of the environment. For example, H. Zhang et al. (2017) measured explicit moral identity. Afterward, their participants were asked to recall and describe either a situation where another person acted more morally than them (upwards moral comparison), a situation where another person demonstrated more competence (upwards competence comparison), or a situation where they were impressed by something (control condition). The participants in the upwards moral comparison group indicated more
prosocial intentions than the participants in the other two groups. This was mediated by experienced guilt. Importantly, moral identity moderated the effect: those participants with higher moral identity experienced more guilt after recalling an upwards moral comparison and subsequently had a stronger desire to act prosocially.

In sum, the Social Cognitive Theory brings three major advancements to moral identity theories. *First*, by introducing the self-schema concept, the theory offers a powerful yet elegant framework for the organization of self-knowledge that can encapsulate affective, cognitive, and motivational components. *Second*, schemata can include nonconscious content and automatic response patterns. Thus, Social Cognitive Theory opens the explanatory range of moral identity for the tacit behaviors and psychological processes that make up most of our everyday life. *Third*, the importance of situational influences and the interaction of person and context variables is emphasized. While in Blasian self-theory (Blasi, 1984, 2005) moral identity is a relatively static construct, the accessibility of moral schemata depends largely on the situation. At the same time, individual differences can be explained in terms of different degrees of moral identity centrality.

So far, we have reviewed prominent theories of moral identity, the Self Model (Blasi, 2005) and the Social Cognitive Theory (Lapsley, 2016). Both theories have answered important questions pertaining to the moral self and advanced our understanding of the relation of moral judgment and moral conduct. The Self Model is well suited to describe moral excellence and exemplary behavior (Shao et al., 2008), and it has laid the groundwork for the scientific analysis of the moral judgment-action gap (Blasi, 1980). One of the major contributions of this model is the proposition that moral judgment only then has a behavioral impact when the issue at hand and morality have self-relevance for the individual (Blasi, 1995, 2005). Morality in this model is more than just reflective reasoning – for the Self Model, an actively constructed personal identity is at the heart of human morality. Importantly, Social Cognitive Theory shifts the focus of the analysis of the moral self towards everyday morality and situational influences. In treating moral identity as a self-schema, it offers a powerful tool to model relevant cognition, behavior, and memory processes (Lapsley, 2016; Shao et al., 2008).

In the next section, we want to complement research presented so far with an account of moral identity from the perspective of motivational psychology. The core assumption of this perspective is that moral identity can be treated as an identity
goal. Thus, while the central questions of the Self Model and Social Cognitive Theory are questions like: “Who am I?”, and “How does identity affect judgment and behavior?”, the goal perspective focusses on questions like: “Who do I want to become?”, and “What are the motivational processes involved in the pursuit of being a moral person?”. We will first outline the properties of moral identity goals, based on goal theories and relate the resulting model to empirical findings about moral identity and moral conduct. Then, we will expand the framework to the level of goal systems and discuss the potential interactions of moral identity goals with others goals. This is crucial because being moral is not an isolated goal. Treating it as such would be an unnecessary reductionism that does not match the reality of human self-construction. Finally, we will present research questions that do emerge when one applies the goal perspective to the moral self. In the subsequent section, we will then analyze these questions from the point of view of all of the three models mentioned: the Self Model, Social Cognitive Theory, and the goal perspective.

**The Moral Self as an Identity Goal**

Identity formation and implementation are ongoing processes throughout the lifespan. Individuals move constantly through the space of actual and possible selves. Markus and Nurius (1986) defined possible selves as “the cognitive manifestations of enduring goals, aspirations, motives, fears, and threats.” Possible selves are distinct from mere self-concepts insofar as they almost always exert a motivational pull (in case of a desired possible self) or a motivational push (in case of a feared possible self) on the agent. Mere self-concepts can be thought of as a cognitive representation of who we are (cf., the Self Model and Social Cognitive Theory). In contrast, possible selves represent who we want to be, who we don’t want to be, and who we could be. Wicklund and Gollwitzer (1982) went a step further, they distinguish identity goals from possible selves. Possible selves are wishes carrying positive or negative valence (Heckhausen & Kuhl, 1985). However, only after an individual selects out of the possible selves (wishes) those they actually want to realize and makes a binding commitment towards attaining them, they do become goals.

Being moral goes beyond mere self-concepts because the moral self is inherently evaluative. Having a moral identity is for most people a desired self that they are committed to attain. It is, therefore, reasonable to conceptualize moral
identity as an identity goal. For this analysis, it is necessary to first provide a
description of general properties of the goal concept and identity goals. Subsequently,
we develop the goal perspective on moral identity.

**Goal processes.** Goals are elementary components of human self-
regulation. They guide behavior and allow us to exert control over our actions and
interact with the environment in a meaningful way (Moskowitz & Grant, 2009).
Although goals have been conceptualized in various different ways, a set of core
features of the goal concept congruent with major goal theories can be defined (Elliot
& Niesta, 2009). Goals are mental representations that specify a future state and link
that state to the person holding the goal. The end state carries positive or negative
valence (Custers & Aarts, 2005) and the link connecting it to the goal pursuer
contains an element of commitment towards approaching or avoiding that state
(Oettingen & Stephens, 2009).

**Committing to a goal.** Goals can be distinguished from other mental
representations of future states such as wishes (i.e., positive fantasies about the future;
Oettingen, 1996) because goals are characterized by the feature that the agent is
committed (determined) to acting towards the goal (Oettingen, 2012; Oettingen &
Stephens, 2009). Goal pursuit begins with an individual selecting a goal out of many
goal candidates (wishes). In this first phase of goal pursuit, the agent has to weigh the
desirability and feasibility of these options. According to the mindset theory of action
phases (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016), this is called the pre-
decisional phase, in which agents are in a deliberative mindset, a cognitive orientation
that is tuned towards processing desirability and feasibility concerns. The higher the
desirability and feasibility of an option, the more likely is it that this option is chosen
and the agent commits to it thus transforming it from a goal candidate (wish) into a
binding goal (Oettingen, Pak, & Schnetter, 2001). The strength of commitment can
vary as a function of the importance (desirability) of a goal and the agent’s self-
efficacy beliefs (i.e., the belief that one is able to perform instrumental goal-directed
actions; Locke & Latham, 2002). Commitment energizes goal pursuing individuals to
act on the goal and leads to increased effort in the face of difficulties, especially when
the goal is specific and challenging (Locke & Latham, 2002, 2013).

**Goal pursuit.** Once agents have committed to a goal, they enter a new
phase in which the primary task is to successfully initiate and execute goal-directed
action. Mindset theory of action phases (Gollwitzer & Keller, 2016) distinguishes a
The Moral Self as an Identity Goal

pre-actional and an actional phase. In the pre-actional phase, the implementation of goals is planned. Cognitive tuning in this phase is towards potential means and opportunities for action. Thus, if individuals are already committed to a goal, their focus is less on the desirability and feasibility of the goal but more on the steps and conditions necessary to achieve that goal. In the actional phase, the individual is oriented towards successfully move towards the goal (e.g., shiedling goal striving from distractions).

A defining feature of goals is that they are pursued until the goal is reached or the agent disengages from them. Consequently, goal effects on attention, memory, and cognition are persistent as long as the goal is active (Martin & Tesser, 2009). Active goals lead to a motivational tension driving the agent to reduce the discrepancy between the status quo and the desired end state of the goal (Lewin, 1951). This requires progress monitoring, which can be modeled as a feedback loop (G. A. Miller, Galanter, & Pribram, 1960). At the minimum, such a feedback loop would have to monitor the discrepancy between the desired goal state and the current state. During goal pursuit, attention is selectively directed at information relevant for the attainment of the active goal and away from alternative goals and distractions (Shah, Friedman, & Kruglanski, 2002; Shah, Hall, & Leander, 2009).

**Goal framing.** Past research has paid considerable attention to the way goals are framed (Gollwitzer & Oettingen, 2012). One can construe the same goal content in various ways with different consequences for goal pursuit. Elliot and Niesta (2009) argued that a central feature of goals is that they involve approach or avoidance of an object. The object of a goal can be anything the agent interacts with (from concrete physical objects to abstract mental representations). Approach and avoidance can occur as physical movements or as mental processes. For example, one can construe the goal of being a kind person as either acting in a prosocial way (approach) or refraining from antisocial behavior (avoidance). Another commonly discussed variation of goal framing is between performance goals aimed at succeeding at a given task and confirming one’s competence, and learning goals aimed at acquiring new skills and increasing competence (Dweck, 1996). According to Molden and Dweck (2006), performance goals are associated with the agent’s belief that abilities cannot easily be altered, whereas learning goals correlate with the belief that abilities can be improved. In the moral realm, implicit theories about the stability and malleability of moral character influence how individuals make moral
judgments. People who hold the belief that a person’s moral character is unchangeable tend to attribute transgressions more to the character of transgressors and favor punishment as a response, whereas people who believe that moral character can change and develop take more situational factors into account when making a moral judgment thus favoring rehabilitation and education as a response to transgressions (Dweck, Chiu, & Hong, 1995).

The way an agent construes their identity goals has also implications for the framing of goals. Higgins (1987) has demonstrated that individuals thinking of themselves in terms of aspired selves (ideal selves) tend to use proactive promotion strategies in their goal pursuit, whereas individuals who construe their identity goals in terms of external demands (ought selves) tend to use prevention-based strategies. A focus on promotion strategies has been shown to increase risk seeking and, as a consequence, dishonest behavior. However, individuals with a promotion focus, more than individuals with a prevention focus, compensated for dishonesty with subsequent ethical behavior (Gino & Margolis, 2011).

Goal structure and mental representation. Kruglanski and Köpetz (2009) argued that on a basic cognitive level, goals function as organized knowledge structures. In this view, similar to the account of social cognitive schemata detailed above, goals are represented as an interconnected memory constellation. Information relevant to the goal is stored and linked. This information can include many different forms of memory, such as desired end states, semantic associates of a goal (Chartrand & Bargh, 1996), other people (Aarts, Gollwitzer, & Hassin, 2004), or means and opportunities to act (Shah & Kruglanski, 2003). Activation of knowledge within a goal constellation can be described in similar terms as other knowledge structures (Bargh, 1990; Bargh & Ferguson, 2000). The accessibility of goals and content of the goal constellation is sensitive to changes in the environment and internal motivational and volitional states. Facing elements of the goal constellation affects the accessibility of other goal-related elements if they are connected via associative links. This way, encountering a stimulus that is cognitively linked to a goal can activate the entire goal structure via spreading activation (Kruglanski & Köpetz, 2009). The spreading activation effect is bidirectional. Having an active goal may increase the likelihood of attending to and processing the elements of a goal system (Shah & Kruglanski, 2003). On the other hand, inhibitory associations prevent conflicting goals and temptations from being activated while a goal is pursued (Shah et al., 2002). Furthermore, these
effects are a function of the strength of the association between elements in the goal constellation.

In contrast to other knowledge structures, goals are motivational in nature. The associative links within a goal system are responsive to the state of goal pursuit and serve self-regulatory functions. As long as a goal is actively pursued, it is important to direct cognitive and motivational resources towards goal pursuit, but upon goal completion investment into the goal is no longer necessary. Thus the relevance of goal-related concepts for an agent is a function of goal completion (Kruglanski & Köpetz, 2009). For example, Moskowitz (2002) has demonstrated that individuals in a state of goal incompleteness direct their attention more than individuals in a state of goal completeness towards goal-relevant items in a non-related task. Similarly, Masicampo and Baumeister (2011) showed that active, unfinished goals increased intrusive thoughts about the goal (Study 1), heightened the mental accessibility of goal-related words (Studies 2 & 3), and decreased performance in an unrelated, cognitively demanding task (Study 4). Working towards the completion of the goal by formulating a specific plan decreased these cognitive effects of unfulfilled goals (Masicampo & Baumeister, 2011). Bargh, Green, and Fitzsimons (2008) showed that goals (e.g., evaluate a job applicant) spill over to goal-related content that was not originally the target of the goal (e.g., impression formation of incidental bystanders), but only if the goal is still incomplete. Spillover effects cease once the goal has been satisfied. In sum, the main difference between goals and other mental representations is that the accessibility of goal content is contingent on the state of goal (in)completion and goal activation (i.e., the mental representation of the goal has been activated by relevant situational cues pertaining to opportunities, hindrances, or instrumental means. Higher accessibility of a goal can only be expected as long as the goal is not attained yet and it is important to note that goal activation differs from goal strength (Sheeran, Webb, & Gollwitzer, 2005). Goal strength is a relatively stable property of a goal as it refers to a person’s degree of goal commitment.

**Symbolic self-completion theory.** Following the theoretical tradition of Kurt Lewin (1927), Wicklund and Gollwitzer (1982) describe in symbolic self-completion theory the process of committing to a goal as entering a tension state. The person pursuing the goal experiences tension until the goal is reached or abandoned. The tension state motivates individuals to work towards the completion of the goal
and once it is attained the tension subsides. For example, a boy who forms the goal of helping an old lady cross a street would experience this tension until he and the lady have safely reached the other side of the street. If his goal was not self-defining, goal attainment can be thought of as complete and the goal is no longer a behavioral motivator. The case is different if the boy acted out of an identity goal (Gollwitzer & Wicklund, 1985a). For example, he could work towards the self-definition of being a nice person. The external goal of helping the lady has a definite endpoint. A single criterion (having safely reached the other side of the road) indicates whether the goal pursuit was successful. In contrast, the internal identity goal of being a nice person is open-ended. It is part of a constructive process to create a persisting personal quality. This quality cannot be attained in the same definite way as an external goal can be completed. A person pursuing an identity goal will be faced repeatedly with evidence for and against possessing the desired identity. Therefore, identity goals can only be relatively complete to the degree that the present evidence indicates the possession of the identity. An individual pursuing an identity goal can move back and forth from a state of goal completeness and incompleteness.

People who are committed to an identity goal seek out indicators of possessing the identity. Such indicators serve as symbols representing the identity goal, and identity goals typically have multiple possible symbols. According to self-completion theory, the symbols of an identity are interchangeable. Individuals can symbolize their identity by demonstrating identity-relevant skills, showing behavior typical of the identity, displaying physical objects, or merely stating intentions to implement the identity goal. Most socially defined identities have numerous potential symbols and it is not possible to acquire all of them. Thus, further symbolizing is always possible (Gollwitzer, 1986). For example, an individual wanting to be a nice person could symbolize being nice by alternative means such as offering someone his seat in the bus, or by merely stating, that he is a nice person. Each of these symbols can reduce the tension that comes along with an active identity goal. Symbols are especially effective when they are noticed by others (i.e., they have become a social reality). This is not merely a matter of strategic self-representation towards others. Because identity goals are typically socially defined, indicators of goal completion (identity symbols) depend on being noticed by others (Gollwitzer, 1986; Gollwitzer & Wicklund, 1985c; Wicklund & Gollwitzer, 1982).
As long as an individual has accumulated symbols of an identity she or he can point to, this person would experience completeness concerning the pursuit of the respective identity. At this point, there is little need for further symbolization and efforts for goal striving are reduced. Individuals remain motivationally complete until they encounter a disruption that calls their attention towards the identity goal and they enter an evaluative state. Losing or failing to acquire an identity-relevant symbol leads to a sense of incompleteness eliciting the psychological tension (Gollwitzer, Wicklund, & Hilton, 1982) that motivates individuals to compensate by symbolizing the identity with alternative means.

Theoretical implication of the goal concept for the moral self. The goal perspective on identity has three important implications for our present discussion of moral identity. First, individuals are committed to a differing degree to identity goals. Commitment determines the motivational power of an identity goal. Moral identity can thus be expected to be more influential over individuals who are committed to being a moral person. Identity goal commitment is reflected in the Self Model and Social Cognitive Theory by moral identity centrality, although those theories don’t explicitly treat moral identity as a self-defining goal. Still, we argue that the most commonly used measures of moral identity centrality, Aquino and Reed’s internalization and symbolization scales (Aquino & Reed, 2002) tap at least partly into moral identity goal commitment (e.g., “I strongly desire to have these characteristics.”).

Second, symbolic self-completion theory (Wicklund & Gollwitzer, 1982) predicts that identity goals regulate behavior in a predictable pattern. Their motivational influence is greater if the agent is in a state of moral identity incompleteness (as compared to identity completeness). Individuals committed to a moral identity are motivated to work towards this identity via the acquisition of symbols. As reported in the moral compensation literature (Monin & Jordan, 2009) – individuals who feel morally complete relax their standards and individuals who feel morally incomplete experience guilt and increase moral efforts. Importantly, commitment is crucial for compensation. Only individuals committed to an identity goal should experience the goal specific tension that drives symbolization. Further support comes from a recent study by Ding et al. (2016), who induced moral identity incompleteness (vs. a neutral control condition) by having participants recall past immoral actions (vs. everyday occurrences). They measured moral identity centrality,
guilt, and willingness to volunteer time. The study found moral compensation effects, indicated by a higher tendency to volunteer by incomplete subjects, which was mediated by guilt. Further, moral identity centrality, which can be thought of as commitment to the moral identity goal in the context of the present discussion, moderated compensation effects. Higher moral identity centrality correlated with more compensation. Similar results are reported by Mulder and Aquino (2013), who found that people high in moral identity centrality are more honest and more willing to donate to charities after lying in economic games. A further prediction of symbolic self-completion theory is that symbols are relatively interchangeable within the same domain. Any symbol representing morality could induce a state completeness that would license future immoral behavior. Across several studies, reliable small to medium sized effects of licensing could be demonstrated after participant reflected on past moral deeds, demonstrated moral behavior, thought about immoral behavior they have refrained from, or merely planned future moral acts (Effron & Conway, 2015).

Third, identity goals can be thought of as organized knowledge structures (Kruglanski & Köpetz, 2009) but in contrast to mere self-schemata, identity goals have a directed self-regulatory function that depends on goal progress and completion. Elements of identity goals are associatively linked. Thus, encountering environmental cues encoded in the goal constellation can make the goal more accessible (cf. goal priming; Bargh & Chartrand, 1999; Moskowitz & Yuichu, 2009). In turn, holding a goal active increases the accessibility of its content. However, the cognitive principles of frequency and recency of activation that one expects to increase accessibility of associated content in other cognitive structures are overridden by motivational principles of goal completion: once a goal is completed, accessibility of goal content decreases (Kruglanski & Köpetz, 2009).

**Goal hierarchies.** In light of competing goals and motives, it is critical to understand how the moral self is related to other goals – self-defining and otherwise. Goals have been conceptualized as hierarchically organized (Carver & Scheier, 1982). This notion is comparable to the concept of higher-order volition (Frankfurt, 1988a, 1988b). At the highest level of such an action hierarchy, people have abstract goals (e.g., be a responsible person). At a lower level, more concrete subgoals translate higher-level goals into principles that guide behavior (e.g., follow through on your promises). These subgoals may still be completed via multiple routes of behavior. One could map out such a goal hierarchy with a multitude of levels from an aspired-to
identity at one end to direct behavioral intentions at the other end. Thus, being a moral person would represent a relatively high-level goal within the action hierarchy. Hierarchical models for goal structures are useful to map out abstract and concrete, self-defining and instrumental goals. However, individuals have typically more than one goal at the same level and a lower level goal may be attached to more than one higher level goal (Austin & Vancouver, 1996).

**Moral self-complexity – moral identity and other goals.** We propose that the moral self can be modeled as a high-level self-defining goal within an interconnected, hierarchical goal network (Shah & Kruglanski, 2000). If this goal approach is applicable to the moral realm, a crucial question arises: What is the scope of the moral identity goal? Thus, we return to the question of whether the moral self is best understood as global and unified or as a multifaceted construct. Hannah et al. (2011) argued for a self-complexity approach to moral identity.

According to self-complexity theory (Linville, 1987; Woolfolk, Gara, Allen, & Beaver, 2004), individuals construe themselves as consisting of various subidentities. These subidentities can be interconnected or independent, resulting in different degrees of complexity. The subidentities determine the values relevant for moral identity. Instead of a singular identity, this approach assumes that the moral self is distributed across other identities. Moral values that are shared among most self-definitions of an agent are this agent’s core values. An individual has a unified moral self if the set of values is similar in most self-definitions.

Importantly, a moral agent can differ on morally crucial dimensions between situations. For example, different social roles can evoke different moral values. Classic social psychological experiments have demonstrated the power of social roles (Zimbardo, 2004). Beyond that, it has been found that everyday morality is partially suspended in sports and replaced by ego-centrism, a phenomenon known as bracketed morality (Bredemeier & Shields, 1986; Kavussanu et al., 2013). This can be explained by the different identity goals pursued in the sports context, which would entail a different set of moral values (e.g., loyalty towards one’s team and promoting one’s own team at the expense of others).

An identity shaping the content of the moral self might even entail values typically considered to be immoral. For example, Marquardt et al. (2016) have demonstrated that individuals who were threatened in their professional identity as a business student (Study 1) or as a law student (Study 2) compensated for this
incompleteness by endorsing immoral behavior that symbolized their identity goals. Furthermore, after being presented with a professional exemplar (Study 3) who was described as having immoral traits, incomplete students claimed to possess traits associated with immorality. Findings on bracketed morality and on self-symbolization via immoral means both support the notion that the content of moral identity depends on other active identities. Further, Marquardt et al. (2016) demonstrated that goal striving dynamics as predicted by symbolic self-completion theory operate on moral identity. Individuals who were incomplete in their professional identity goal adopted values ostensibly consistent with those professions but inconsistent with moral values. Adopting immoral values served as a symbol to complete the professional identity goal. The state of professional identity goal incompleteness affected the participants’ moral value system.

**Emerging Research Questions**

In this section, we present three research questions related to goals and morality and discuss how the Self Model, Social Cognitive Theory, and the goal perspective could inform these topics. These questions emerge as a consequence of conceptualizing the moral self as an identity goal, but they pertain to the moral self in general.

*First*, how does the cultural background influence the moral self? Oettingen, Sevincer, and Gollwitzer (2008) have analyzed the process of goal pursuit in a cross-cultural comparison and found that content, assessments of desirability and feasibility, and the use of self-regulatory strategies differ between cultures. Regarding morality, Graham et al. (2018) argued that there is a substantial variation concerning the moral values different individuals, groups, and cultures include in their moral value system. Thus, to conceptualize the moral self, one has to take cultural influences into account, especially if one understands the moral self as a goal.

*Second*, how is the moral self related to other identities and how does it interact with them? A person usually holds multiple goals at the same time and they can be described as connected within a goal network (Shah & Kruglanski, 2000). Goals can hierarchically depend on each other, they can be in conflict, and multiple goals can be associated with the same means. Similarly, the moral self is not an isolated phenomenon. Aside from a moral identity, people may hold any number of
other identities. We argue that factoring in the interaction of moral identity, other goals, and other identities will enrich moral identity theories.

Third, when does a moral identity motivate agents to act in accordance with its values and when does it not? Within the literature about moral identity, both compensatory as well as consistency effects of moral identity have been reported (Conway & Peetz, 2012; Joosten, Van Dijke, Van Hiel, & De Cremer, 2014; Mullen & Monin, 2016). Having a goal does not necessarily lead to goal-directed action. Only if goal strength is high and the goal is activated, can we expect behavioral consequences. Even that may not be enough for successful goal attainment. Self-regulation is needed to overcome problems during goal striving (Gollwitzer & Oettingen, 2012). In this section, we will examine motivational dynamics of moral identity.

Cultural Influences

It is important to clarify that our understanding of morality and moral identity in psychology is a purely descriptive notion. We hold the view that on a descriptive level human morality is best understood in pluralistic terms. Other pluralistic theories have been criticized on the ground that a broad definition of morality that is compatible with ideological diversity would treat certain preferences (e.g., right-wing ideology) as genuine moral concerns, even though they should be considered immoral or at least amoral (Kugler, Jost, & Noorbaloochi, 2014). This criticism suffers from confusing is and ought. Descriptive moral pluralism is about how people do experience the world and how they act in it. It is not about how they ought to behave or experience the world. A scientific account of morality outlining what people experience as domains of moral concern should not be misunderstood as a normative theory about how people should construct their moral systems. Human morality is diverse and what people consider to be moral need not necessarily be beneficial or good. In fact, extreme moralizing seems to be a major cause of atrocities around the world (Baumeister, 1996; Graham & Haidt, 2012). Here, our aim is not to condemn or condone any moral system but purely to describe. No matter whether we agree or disagree with any particular moral system, as long as they share their underlying psychological mechanisms it is sensible to assume a common principle to understand how moral systems work.
In the following, we will provide a review of research concerning cultural variations of moral values. We interpret these findings as evidence that a pluralistic conception of morality is best suited for descriptive moral theories. We conclude the section with brief suggestions on how to include cultural variations and moral pluralism in each of the three models described in the present article.

**Kohlberg and Turiel: Justice, care, and individual rights.** Most moral psychological research has been conducted within the cultural context of western Europe or the USA (Henrich et al., 2010). Culture and the moral philosophical history has shaped the moral theories brought forward by psychologists. As discussed above, the influence of Kantian ethics is apparent in Kohlberg’s theories (Kohlberg, 1981). Kohlbergian ethics are rationalistic, but beyond the process of moral reasoning which is at its highest stage not unlike a categorical imperative, the content of morality according to early Kohlbergian theories was very much in line with Kantian ideals of justice and fairness. This has been criticized as a too narrow view of morality, most prominently by Gilligan (1982), who proposed that female morality developed differently from male morality. While men and boys relied more on principles of justice, women and girls relied more on principles of caring. Although empirical findings don’t offer strong support for the claim of sex differences in moral reasoning (Jaffee & Hyde, 2000; Walker, 1984; Woods, 1996), the argument made it clear that morality is about more than fairness, a point acknowledged in later formulations of Kohlberg’s theory (Kohlberg, Levine, & Hewer, 1983).

Contemporary feminist ethics underscore that moral theorizing needs to build its values on diverse perspectives to correct biases (Driver, 2013c). Eliot Turiel, a student of Kohlberg’s, defined moral rules in contrast to conventional rules, a distinction that according to his argument is culturally universal. Moral rules are perceived in contrast to conventional rules as independent of authority with an objective prescriptive force, and violations of moral rules are typically seen as more serious than violations of conventional rules. A commonality of moral rules in the Turiel tradition is that violations of moral rules tend to involve a victim who has either been harmed, whose rights have been violated, or who has been a victim of injustice (Turiel, 1983; Turiel, Killen, & Helwig, 1987; Turiel & Nucci, 2018). The central values of morality in Turiel’s project can be ascribed broadly to concerns of fairness (justice and rights) and care (prevent harm). It has been shown that a general distinction of moral and conventional transgressions is already made in early
childhood (Turiel & Nucci, 2018). Furthermore, the distinction of the moral and conventional domain has been demonstrated in multiple cultural contexts (Hollos, Leis, & Turiel, 1986; Nucci, Turiel, & Encarnacion-Gawrych, 1983; Tisak & Turiel, 1988). The underlying values of fairness and harm avoidance are nowadays reflected in the majority of psychological work about morality (Graham et al., 2018; Haidt, 2013a; Haidt & Graham, 2007) including the widely used moral identity internalization and symbolization scales constructed on the basis of Social Cognitive Theory (Aquino & Reed, 2002).

**WEIRD populations.** The case for moral universalism on the basis of fairness and care (Driver, 2013a) becomes a lot murkier, however, when additional sources are considered, especially the anthropological literature (Cassaniti & Hickman, 2014). Cross-cultural comparisons revealed that the psychological components likely to underlie human morality vary significantly between cultures (Henrich et al., 2010). In fact, in many respects, the Western urban population (which makes up the majority of participants in published psychological studies) seems to be an outlier compared to humanity as a whole (Haidt, 2013b; Henrich et al., 2010). Western individuals tend to prefer an analytic thinking style (i.e., using abstract representations, partitioning issues into distinct categories, and focusing on discrete objects and their attributes). Non-western individuals, on the other hand, tend to prefer a holistic thinking style (i.e., representing the world relational and processing issues and objects contextualized). Although this distinction has initially been drawn between Western and East Asian populations, almost all non-western populations seem to demonstrate a similar pattern in this regard (Nisbett, Peng, Choi, & Norenzayan, 2001).

Henrich et al. (2010) draw the conclusion that Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies are not representative of humanity at large. The unusual pattern of thinking styles and values is even more pronounced in liberal, middle class urbanized populations. The cultural differences in basic cognitive styles are related to both self-concepts and morality. WEIRD individuals typically have an independent self-concept, understanding themselves in terms of unique individual features that apply to them, whereas nonWEIRD individuals typically have an interdependent self-concept, emphasizing their social roles and relationships (Heine, 2001). Assessments of Kohlberg’s developmental model demonstrate the full range of stages in WEIRD populations, but post-
conventional thinking, the highest level in Kohlberg’s structure, is rare in small-scale societies (Edwards, 1986; Snarey, 1985). This cannot just be attributed to a lack of formal education since post-conventional reasoning is also absent in some samples of highly educated non-Westerners (Al-Shehab, 2002; Henrich et al., 2010; J. G. Miller, Bersoff, & Harwood, 1990).

There are three likely reasons underlying this cultural bias. The first is the aforementioned difference in thinking styles. Kohlberg’s theories favor analytic over holistic thinking. The second is the related difference in self-concepts. The independent self-concept predominantly found in WEIRD societies values autonomy, which is essential for the post-conventional level. The interdependent self-concept in contrast values community over autonomy. The third reason is that WEIRD individuals construct their moral values mostly around principles of fairness and care, whereas non-WEIRD individuals rely also on other principles (Haidt, Koller, & Dias, 1993; J. G. Miller & Bersoff, 1992). This divide can even be found within one and the same culture: liberal North Americans consider morality to be a matter of harm and fairness, whereas for conservative North Americans morality is also about loyalty, hierarchy, and purity (Haidt & Graham, 2007).

**Schweder: The “big three of morality”.** In comparing orthodox Hindu individuals from Bhubaneswar in Eastern India and a matched sample of Jewish, Christian, and secular individuals from Chicago, Shweder, Mahapatra, and Miller (1987) challenged Kohlberg’s and Turiel’s notions of ethical universalism based on justice, harm, individual rights, and welfare. They found that the Indian sample moralized issues that were not considered to fall in the moral domain for the American sample, such as food, clothing, or sex roles. In fact, the authors argue, the separation of issues into the moral and conventional domains is a peculiarity of Western culture rather than a psychological universal. As an alternative to a morality based solely on the wellbeing of individuals, Shweder et al. (1997) proposed a categorization of morality into three types: First, WEIRD morality maps on to an ethic of autonomy that is concerned with the welfare of individuals, their rights, and justice. Second, an ethic of community encompasses people’s obligations in their social groups, their roles in larger entities, hierarchies, and the order of relationships in society; loyalty and patriotism are examples of an ethic of community. Finally, the third type of morality is an ethic of divinity. It includes concerns about purity,
sanctity, and elevation. Rules about food, sexuality, and hygiene fall into this category.

Haidt et al. (1993) substantiated Shweder’s claim that moral norms are culture-dependent by presenting Brazilian and U.S. subjects from high and low socio-economic status (SES) groups with harmless taboo violations – short stories of victimless actions that are perceived by many as offensive (e.g., a family eating their dog who had been killed in a car accident). They asked questions to assess moral and conventional norms (interference: “Should the [actor] be stopped or punished in any way?”, universality: “Suppose you learn about two different foreign countries. In country A, people [do that act] very often, and in country B, they never [do that act]. Are both of these customs OK, or is one of them bad or wrong?”). In addition, they probed whether their research participants thought anyone was harmed in the scenarios and whether they were bothered by the described acts. The main finding of this study is that SES systematically predicted the moral judgments. Across both countries, individuals with low SES endorsed interference with harmless taboo violations more and they universalized their judgments more than individuals with high SES. The same pattern emerged in an analysis of only those cases for which the participants indicated that the described action was indeed harmless, and that they were bothered by it. Thus, moral judgments of these cases could not be reduced to the condemnation of perceived harm. Other work (Kelly, Stich, Haley, Eng, & Fessler, 2007) has demonstrated that the moral conventional distinction does not reliably emerge when a wide range of harmful transgressions (e.g., slavery in ancient Rome vs. the American South, corporal punishment – legally allowed vs. legally prohibited) is used instead of typical schoolyard behavior. In their study, the authors found that judgments of the permissibility and wrongness of harmful acts were contingent on authority and the historical context – signature conventional responses. If the moral-conventional distinction were universal as Turiel suggests (Turiel, 1983; Turiel et al., 1987; Turiel & Nucci, 2018), then condemnation of harmfull acts should be universally independent of authority and context.

**Moral Foundations Theory.** Haidt and his colleagues (Haidt & Graham, 2007; Haidt & Joseph, 2004, 2008) proposed an alternative framework to Kohlberg’s and Turiel’s individualist morality. Their Moral Foundations Theory draws from evolutionary psychology to provide a common set of systems underlying human morality that is able to account for cultural variations in moral values. The core idea
of Moral Foundations Theory is that humans have a set of innate modules that provide us with an *a priori* preparedness to learn moral norms (Graham et al., 2013). Thus, domain-specific learning is facilitated by the moral modules. Just as humans (and other animals) have an innate preparedness that makes it easier to learn to fear snakes than flowers (DeLoache & LoBue, 2009; Mineka & Cook, 1988), the authors argue that humans also have learning modules specifically for moral learning (Graham et al., 2013). As individuals grow up and gather experience in their social environment, they built their system of morality on these modules (i.e., the moral foundations). The foundations are assumed to be universal, innate modules on which values are constructed via cultural learning. These learning processes are compatible with the schema account in Social Cognitive Theory (Lapsley, 2016) and can lead to constructions of information networks including explicit and tacit knowledge as well as behavioral scripts and automatic associations. The foundations do not completely determine which moral values will be developed, but they constrain the possible moral systems and determine which moral values can easily be learned (Graham et al., 2013).

Graham and his colleagues formulated five defining features of moral foundations. First, the values built on a foundation must be common in third-party normative judgments (i.e., judgments about what someone ought to do if the action does not directly affect the person judging the act). Second, violations of moral foundations evoke automatic affective evaluations. Observing an immoral act makes the observer immediately feel that it was wrong. Each foundation has a signature affective response. In particular, anger and disgust seem to be typical responses to moral transgressions (Russell & Giner-Sorolla, 2011, 2013), and distinct neural systems for different transgressions (harm vs. cheating vs. sexual disgust) typically associated with these emotions are correlated with foundation specific moral judgments (Parkinson et al., 2011). Third, the foundation must be present across a wide array of cultures. Even though moral foundations are malleable to environmental and social influences because the underlying structure is assumed to be universal, traces of the foundations should be found in most cultures. Fourth, moral foundations facilitate moral learning through innate preparedness. Evidence for this can be provided by investigating infants (e.g., Sloane, Baillargeon, & Premack, 2012) or non-human primates (e.g., de Waal, 2009). Fifth, moral foundations must provide a clear evolutionary advantage.
Moral foundations theory embraces moral pluralism and can in some respects be seen as an extension of Shweder’s model. In its current formulation (Graham et al., 2018), the theory provides detailed accounts of five foundations. The foundation of care/harm, which is primarily about the well-being of other people is assumed to have evolutionarily developed from the mammalian need to protect and care for one’s offspring. It is related to compassion, kindness, and anger towards individuals who harm others. A foundation of fairness/cheating developed according to the theory as a mechanism to increase benefits in reciprocal partnerships. It underlies notions of justice and fairness. The first two foundations resemble the values of Kohlberg’s and Turiel morality and fall into Shweder’s category of ethics referred to as autonomy. In moral foundations theory, they are referred to as the individualizing foundations. Two foundations have been suggested that map on to the ethics of community. The loyalty/betrayal foundation serves to increase group cohesion and cooperation within social groups, and the authority/subversion foundation is the basis for social hierarchies. The fifth foundation is labeled purity/degradation and corresponds to ethics of divinity, encapsulating, for example, sexual morality. These latter three foundations are thought of as binding foundations because they bind individuals into larger groups. The authors point out that this list of foundations is not exhaustive and morality might not be limited to these five concerns. For example, liberty/oppression (concern for personal freedom) is a candidate for foundationhood (Iyer, Koleva, Graham, Ditto, & Haidt, 2012). The moral foundations are the ground on which culture-specific moral systems are built; the degree to which each foundation is relevant to a person is a point of individual differences.

For example, individual differences in the endorsement of moral foundations correlate with attitudes towards foundation corresponding policy issues and recreational drug use (Koleva, Graham, Iyer, Ditto, & Haidt, 2012). The loyalty/betrayal score predicts identification with one’s favorite sports team (Winegard & Deaner, 2010), and the care/harm dimension is linked to perception and appeal of violent media content (Tamborini, Eden, Bowman, Grizzard, & Lachlan, 2012). Importantly, it has been shown that concern for binding foundations, especially the purity/degradation foundation, has explanatory power above and beyond concern for the individualizing foundations. This has been demonstrated with various methodologies and in diverse contexts (Clifford & Jerit, 2013; Dehghani et al., 2016; Graham et al., 2018; Koleva et al., 2012). Furthermore, framing liberal issues in terms
of purity increased conservatives support for these issues (Day, Fiske, Downing, & Trail, 2014; Feinberg & Willer, 2013, 2015).

Critics of moral foundations theory have questioned the central claims of innateness (Suhler & Churchland, 2011) and intuitionism (Narvaez, 2008). Most important for the present article is the debate over moral pluralism. While some authors have suggested alternative pluralist theories (e.g., Janoff-Bulman & Carnes, 2013; Rai & Fiske, 2011), others reject pluralism in favor of more parsimonious models. Gray and colleagues developed a theory of dyadic morality (Gray, Schein, & Ward, 2014; Gray, Waytz, & Young, 2012). They propose that moral transgressions are always psychologically construed in a dyad of perpetrator and victim. Even if no objective victim is present (as in Haidt’s harmless taboo violations), the observer would automatically think in terms of a dyad and detect some kind of victim (e.g., a spiritual entity, society, or the perpetrator). Indeed, they have found across multiple studies that ostensibly harmless violations are perceived as harmful (Gray et al., 2014; Schein, Ritter, & Gray, 2016). Thus, the authors conclude that a single value underlies human morality: harm. Graham et al. (2018) agree that care/harm is the most prototypical dimension of morality. However, they disagree with the idea that all of morality can be reduced to perceived harm. According to Graham and his colleagues, harm is either too narrow a principle to account for moral pluralism, or too vague to serve as a useful concept of a parsimonious theory (Graham, 2015).

**The Road to Moral Pluralism.** There is a strong case to be made for adopting a pluralist position in moral psychology. Moral values vary with culture (Henrich, 2015; Saucier et al., 2015) and cultural values, in turn, affect individual moral values (Shilling & Brown, 2015; Vauclair et al., 2015). Even moral principles that seem to be based on the same foundation differ between cultures. For instance, German children have different fairness norms than children from hunter-gatherer cultures (Schäfer, Haun, & Tomasello, 2015). We propose that moral identity theories would benefit from asking the fundamental question: How does culture influence the content, structure, and processes of moral identity?

**Culture and the Self Model.** The central propositions of the Self Model are that human morality is fundamentally linked to an individual’s self-identity. Further, only if a moral judgment is perceived as self-relevant and the agent feels responsible to act on that judgment, can a strong effect on behavior be expected (Blasi, 1984, 2005). Blasi (2005) proposed that individuals with a moral identity have three
essential virtues: moral desire, willpower, and integrity. This framework is theoretically compatible with a pluralist notion of morality. Moral desire determines the content of morality and it can be shaped by cultural influences. Willpower represents the self-regulatory capacity of the agent, and integrity the quasi-need to behave in a self-consistent way.

However, it has recently been pointed out (Jia & Krettenauer, 2017) that the moral identity construct as conceptualized in current empirical research while adequately suited to reflect Western morality is only weakly related to non-Western morality (Hertz & Krettenauer, 2016). This is likely due to the fact that the common moral identity concept in psychological research is strongly influenced by the Western, individualist culture. Some of this cultural bias may be traced back to the Kohlbergian roots of the Self Model, which emphasize western justice and care norms, and an individualist morality. The self and moral values differ between WEIRD and sociocentric, Eastern cultures (Buchtel et al., 2015; Hwang, 2006; Markus & Kitayama, 1991, 2010). According to Jia and Krettenauer (2017), while there is an overlap of moral values between Eastern and Western societies, some values used to describe morality are culturally specific; with Western values emphasizing an individually oriented morality and Eastern values emphasizing a societally oriented morality. The authors suggest that assessments of moral identity must include culture-specific moral attributes, because what it means to be a moral person is not the same across cultures. Moral identity in individualistic societies may be construed as an independent self with personal values. Moral identity in sociocentric societies, in contrast, may be more akin to a social identity, as its values depend to a larger degree on shared group norms.

Culture-specific moral content could be described within a common framework based on the central assumptions of the Self Model, but some components of the model require adjustment or further elaboration to fit with non-WEIRD societies. Blasi (2004a) stated that morality is fundamentally cognitive (based on reflective reasoned judgments). Yet, reasoning differs between cultures. Westerners rely more on an analytic style of reasoning, whereas Easterners rely more on a holistic reasoning style. This difference has been directly linked to a difference in self-construals. Independent self-construals correlate with a propensity for analytic thought and interdependent self-construals correlate with a propensity for holistic thought (Henrich et al., 2010). Thus, if one argues, as Blasi (2004a) did, that morality
is based on reasoning, then a pluralistic perspective would have to consider cultural influences on reasoning styles.

**Culture and Social Cognitive Theory.** According to Social Cognitive Theory, the moral values of a person are stored in self-schemata (Lapsley, 2016). Modern social cognitive accounts of moral identity remain largely silent about moral content. However, the most commonly used measure of moral identity centrality derived from Social Cognitive Theory, the self-importance scale of morality (Aquino & Reed, 2002), relies on moral attributes that reflect WEIRD cultural values; accordingly, it has been suggested that a more inclusive measure would better reflect moral identity of non-Western cultures (Jia & Krettenauer, 2017). Bandura (1991) proposed that moral values are acquired in the process of social learning as a combination of reward and punishment as well as modeling. Norms and values are transmitted by one’s social environment. Learned values are internalized, but this is not just a matter of copying and passively adopting. Rather, individuals learn values from diverse sources (e.g., institutions, parents, siblings, peers, and media), modify them, and selectively construct their own system of values as a composite of the social influences they encounter. Bandura’s (1991) notion of social learning meshes well with moral pluralism. In fact, it goes one step further by describing differences not only on the cultural level but also on the level of the individual.

**Culture and Moral Identity Goals.** Culture affects what kind of goals people pursue (Oishi & Diener, 2009). People in individualist cultures pursue more goals directed at personal wishes, whereas people in collectivist cultures pursue more goals directed at the interests of others (Triandis, 1995). Markus and Kitayama (1991) proposed that in collectivist cultures “the goals of others may become so focal in consciousness that the goals of others may be experienced as personal goals” (p. 229). Cultural background clearly influences which values are shared among one’s group and these values, in turn, affect appraisals of the desirability of goal content (Markus & Kitayama, 1991). In addition to desirability judgments, feasibility judgments and efficacy beliefs also are culture dependent (Oettingen et al., 2008).

The social environment and cultural background determine which moral values are part of one’s moral identity. From a hierarchical goal perspective, achieving one’s moral values can be seen as the desired end state of a goal (e.g., “I want to be a person who is honest”). One can also construe the moral identity goal as a higher level goal and specific values as means to attain that goal (e.g., “I want to be
a moral person, and in order to achieve this I will be honest.”). Identity goals are socially defined. Gollwitzer (1981) has shown, that symbols need to be taken notice of by others to be effective means of reaching a state of completeness. Furthermore, the content of identity goals is determined in social interactions with others (Wicklund & Gollwitzer, 1982). Interpersonal differences in moral identity can stem from two sources. First, an individual can hold different self-defining goals that overlap with morality. For example, one person may have an identity goal to be a religious person and their religion promotes certain moral values. Another person may have a different religion or no religion at all. The corresponding self-definitions are likely to entail different moral values. Second, the social definitions of identities that are similarly labeled may differ between groups and cultures. For example, American Jews and Protestants differ in their self-construal and in their religious moral values (Cohen, 2015).

In sum, the goal perspective needs to include culture as an important explanatory variable for two reasons. First, cultural influences on goal selection and goal pursuit have been reported (Oettingen et al., 2008). Second, the moral values a person adheres to are influenced by culture. Differences in goal striving and moral values are thus likely to be present at the individual members of different cultures.

The Moral Self and Other Identities

Moral identity research in the tradition of the Self Model and Social Cognitive Theory has mostly focused on a global, domain-general moral-self construct (Jennings et al., 2015). These approaches pay little attention to the fact that individuals usually hold multiple domain-specific identities. Stets and Serpe (2013) defined three types of identities. Role identities describe the meanings of an individual’s roles in their social contexts. Social identities indicate membership of a group and a self-definition of belonging to that group. Person identities refer to idiosyncratic characteristics that make an individual unique. Traditional domain-general accounts of moral identity can be classified as person identity. This type of identity matches most closely the Western cultural understanding of the self (Henrich et al., 2010; Markus & Kitayama, 2010). We propose that this view of the moral self, which assumes that moral identity is unified and separate from other identities misses non-WEIRD morality and situational changes in moral identity. It has been
demonstrated that activating role identities (e.g., family member) and social identities (e.g., nationality) influences people’s preferences (LeBoeuf, Shafir, & Bayuk, 2010).

More directly, Leavitt, Reynolds, Barnes, Schilpzand, and Hannah (2012) investigated how occupational identities are related to moral judgments. They assumed that the sphere of moral concern (i.e., which people are considered to be morally relevant) varies between occupations. Accordingly, active occupational identities should be related to predictably broad (i.e., many different people are thought of as morally relevant) or narrow spheres of moral concern (i.e., only a small subset of people is considered to be morally relevant). In two studies they examined soldiers who were trained combat medics (Study 1) or graduating to become medics (Study 2). Importantly in both studies the participants had dual occupational identities. The soldier identity was predicted to be associated with a narrow sphere of moral concern. In contrast, the medic identity was predicted to be associated with a wider sphere of moral concern. This was demonstrated using implicit measures of moral identity centrality (Study 1). In addition, the authors showed that priming the medic identity (Study 2) led to more principled moral judgments, and priming the soldier identity led to more flexible (outcome-oriented) moral judgments. In a third study, they conceptually replicated this finding with people who held a dual engineering and management identity, and found a similar pattern for people in the private sector (but not for those working in the public sector).

Other studies point into a similar direction. Babcock et al. (1995) tested how experimentally assigned roles in a negotiation task biased perceptions of fairness. They presented their participants with materials from a legal case regarding a traffic accident. The participants were assigned the roles of plaintiff or defendant, and later on had to negotiate for compensation of damages. Importantly, both parties received the exact same information about the case. Before the negotiation they were asked to indicate what they thought a fair settlement would be from a neutral observer perspective and they had to estimate how much the judge would award. The materials were drawn from a real case. Both parties were incentivized to be as accurate as possible; they would receive a bonus if they were close to the real settlement. Despite the personal incentive for accuracy, the participants demonstrated a self-serving bias in favor of their clients. These results can be interpreted as biased fairness influenced by role identity.
Other findings also point towards within-person domain specific moral systems. For example, athletes’ moral systems differ between the sport context and other areas of life. Although general moral identity is negatively related to antisocial behavior in sports (Kavussanu et al., 2015), student-athletes’ self-ratings of the likelihood to behave antisocially were higher towards opponents in sports than towards other students. Additionally, ingroup biases in favor of teammates emerged in the sports context (Kavussanu & Ring, 2016). Others have found diverging patterns in moral reasoning about sports and everyday life, suggesting that different moral mechanisms are at work (Bredemeier & Shields, 1986). Taken together, one can conclude from these studies that the moral self should be treated as contextualized (Jennings et al., 2015).

**Moral Disengagement.** The most dramatic and controversial case of role-dependent moral identity in psychological experiments is probably provided by the Stanford Prison Experiment (Haney, Banks, & Zimbardo, 1973). In an attempt to understand the power of the social environment on individual behavior, Zimbardo and his colleagues set up a simulated prison. The participants were randomly assigned the roles of guard or inmate. Zimbardo himself played the role of prison superintendent. The guards were instructed to abstain from physical violence but were allowed to exhibit psychological pressure. Both the prisoners and inmates were deindividualized (with uniforms and designated numbers for the prisoners). The experiment was planned to last for two weeks. However, the behavior of the participants escalated so dramatically that it was aborted after only six days. The prisoners became increasingly apathetic. The guards, on the other hand, showed sadistic tendencies and severe cruelty as they abused the inmates. Zimbardo noticed that just as the guards and prisoners, he himself became absorbed in his role as superintendent. Despite the fact that the study remains highly problematic for methodological as well as ethical reasons, partial replications have been attempted (Lovibond & Adams, 1979). Haslam and Reicher (2005) criticized Zimbardo’s method. The original study lacked proper control conditions and because Zimbardo was heavily involved in the experiment as superintendent, he was no neutral observer. In cooperation with the British Broadcasting Company, Haslam and Reicher conducted their own prison study. The BBC Prison Study (Haslam, 2006) followed the template of the Stanford Prison Study, with two major differences. First, the events of the study were broadcasted in television, and second, the researchers had no formal role during the experiment, thus
remaining in an observer position. The findings line up with earlier findings showing the influence of the assigned roles. Taken together, the prison studies provide evidence for the profound impact of role identity on behavior.

Even if it varies with moral identity centrality, humans have a desire to see themselves as moral persons (Aquino & Reed, 2002). Yet behavior like the cruelty exhibited by the guards in the prison studies above clashes with widely held moral beliefs. According to moral disengagement theories (Bandura, 1991, 1999a, 2016; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Bandura, Caprara, Barbaranelli, Pastorelli, & Regalia, 2001), several mechanisms can protect agents from moral self-sanctions in case of transgression. Moral disengagement mechanisms have the function to preserve a sense of moral identity. An important mechanism is diffusion of responsibility (Darley & Latane, 1968), that is, if decisions are made by groups or people act as a collective instead of individuals, responsibility for actions and decisions is not clearly defined. Because no single individual is responsible for the performed actions (or inactions in case of bystander interventions; Darley & Latane, 1968), ethical concerns have less force in such situations. This is amplified if the agent identifies strongly with a group or as the holder of a role. Thus deindividuation on the side of the actor or decision maker can lead to moral disengagement. Deindividuation and dehumanization on the side of the victim can increase moral disengagement as well. It is far easier to perform moral atrocities towards an anonymous stranger or to a distant group of people than toward an individual (Bandura et al., 1996). Nazi war criminals operating concentration camps stripped Prisoners of their identity not only to degrade them but mainly to make it easier for the guards to operate the camps (Baumeister, 1996). Other factors influencing moral disengagement are related to the construal of the performed action and its consequences. Framing a moral transgression in terms of euphemistic labeling (e.g. “collateral damage” instead of “killing civilians”), using palliative comparisons (e.g. “compared to the killer, the thief is a moral person”), and ignoring or misconstruing the consequences of one’s actions can lead to moral disengagement.

Moral disengagement shows drastically, how moral norms can easily be adjusted to group norms and role affordances. One mechanism of moral disengagement employs moral ideals and ideology. “Exonerative moral reasoning can make the immoral inconsequential or even moral” (Bandura, 1991, p. 56). In the name of one’s country, religion, or moral ideals, inflicting harm can become a moral
imperative. This phenomenon is seen for example in terrorist ideologies (Reich, 1990) which can take a religious (e.g., permission to kill nonbelievers) or secular, often utilitarian (e.g., the ends justify the means) form. This form of moral disengagement can also be understood as a particularistic, local morality, that is embedded in certain groups (Greene, 2014). Local morality influences individuals by way of their identification with the group or their role. Interestingly, findings about moral exemplars line up with such a particularistic morality. Case studies of rescuers, bystanders, and Nazis during the Holocaust demonstrate this impressively. Monroe (2003, 2008) interviewed people who rescued Jews, bystanders, and supporters as well as perpetrators of genocide. The groups differed dramatically in terms of the moral values and their sense of moral efficacy which they incorporated into their self-concepts. Yet all three groups felt that their identity constrained their choices. Being who they were, they could not have acted differently.

In sum, the examples above illustrate that moral identity often interacts with other identities and situational factors. We propose that human value systems have a certain degree of flexibility. Depending on the current configuration of the working self-concept (in the langue of the Self Model and Social Cognitive Theory) or the set of active identity goals (in the language of the goal perspective), the content of morality is malleable by situational influences. Especially role and group identities seem to be powerful in shaping situated moral identity.

Identity interactions and the Self Model. In the Self Model, other identities than the moral self and competing goals are treated as temptations that can be overcome with willpower (Blasi, 2005). The model assumes that moral content is a relatively stable set of values and other identities do not easily alter this set. Following this assumption, neither other identities nor the mechanisms of moral disengagement would affect what it means to be a moral person. This implies that role and group dependent values do not have the status of moral norms, but are treated as non-moral preferences. This position reflects the assumption that morality, while embedded in personal identity has a universally objective quality (Bergman, 2004)

However, identity is typically construed in a situational context (Diehl & Hay, 2007). Krettenauer and Hertz (2015) argued that the endorsement of moral values and the centrality of morality varies across social contexts, especially roles. The degree of differentiation seems to be a developmental process. Differentiation
increases during adolescence (Daniel et al., 2012) and decreases with old age (Krettenauer, Murua, & Jia, 2016).

If the Self Model were expanded to include differentiation of moral identity, two types of interactions with other identities could be predicted by the model. First, non-moral identities may be in conflict with moral identity and self-regulation is needed to resolve that conflict. Second, there is a variability in the centrality of moral values in different life domains, where different roles and social identities are active. This conceptualization of morality is however incompatible with Blasi’s (2005) notion of moral objectivity. While a stable Blasian account of morality may be suitable to describe exemplary individuals who are defined by moral consistency across many contexts, a dynamic account would be a more suitable description of everyday fluctuations in moral value endorsements.

Identity interactions and Social Cognitive Theory. Social Cognitive Theory emphasizes situational factors more than the Self Model. Thus far, however, the interplay of moral identity and other identities has received little attention in the research following this theoretical tradition. It will be important for future work on social cognitive models of moral identity to elaborate how the moral self is related to other selves. One critical question is: What are the boundaries of moral identity? For example, a person may have in addition to a moral identity the professional identity of being a physician. It needs to be clarified if such identities are seen as independent or overlapping with the moral self. Independent schemata would be competing for cognitive resources and activation of one schema in the working self-concept would inhibit the other.

Alternatively, it is equally possible, that the professional and moral identity schemata are connected via associative links. In that case, activation of one identity might be able to increase the accessibility of the other. Finally, there might be no clear boundaries between the moral self and other self-schemata (i.e., the schemata overlap). In the latter case, moral content strongly associated with an overlapping self-schema would be more accessible, as long as the other schema is active. It is likely that all three types of self-schema interactions exist. We propose that these three types of interactions could be used to explain situated moral identity and moral disengagement in terms of Social Cognitive Theory.

Identity interactions and the Goal Perspective. Kruglanski & Köpetz (2009) described goal systems as “cognitive webs of interconnected constructs”
Aside from means, episodic memory content, semantic and procedural knowledge, goal systems also include links to alternative goals. Such associations between goals can be facilitative or inhibitory. It has been shown for example that activation of a goal inhibits processing of distracting, temptation-related information (Fishbach, Friedman, & Kruglanski, 2003). In addition to lateral connections of goals at similar levels of abstraction, goals are hierarchically organized (Carver & Scheier, 1982). This means that some goals can function as means for higher-order goals. Furthermore, one goal has typically multiple means (i.e., equifinality) and a means may serve to attain multiple goals (i.e., multifinality; Kruglanski & Köpetz, 2009).

We propose that moral identity goals are characterized by their overlap with other identity goals. Being moral is an essential self-definition for most people (Strohminger & Nichols, 2014) but the meaning of this self-definition is only roughly defined by some core values. Here we adopt a self-complexity perspective (Hannah et al., 2011). Many identity goals contain a moral dimension. For a good father, being caring and supportive of his children are a moral imperative. For a good scientist, scientific rigor and integrity are important moral values. For a fan of a sports team, loyalty towards one’s team is at the core of the sports fan’s morality. These various identity goals can be held by the same person in different contexts. Importantly, although all of these identity goals have morality as one of its dimension, what it means to be moral depends on the domain in question. Thus, depending on which identity goal is activated, different norms and values become salient as means to symbolize the identity of being moral. Instead of a unified global moral self, we propose that a moral identity goal is distributed across our various self-definitions. It is a high-level goal that is bestowed with content by the other identity goals it overlaps with. We concur with Hannah et al. (2011) who argued that the question for the moral self is “Which identity is activated and what is the unique moral makeup of that identity” (p. 673). In addition to moral values, other aspects of the moral self, including cognition, memory content, sensitivity to certain cues, and self-regulatory resources are likely to depend on other active identities. For example, a person might think of moral matters in abstract terms in the professional context and in concrete terms as a private person. This same person may resist temptations of illicit conduct in the workplace but lack this self-regulatory capacity in close relationships.

Cross-situational consistency within this framework is a function of personal unity (Hannah et al., 2011), where unity refers to the overlap of moral
content across identity domains. Generally, if behavior or motives are linked to multiple goals, high motivational stability can be expected (Ford, 1992). For example, a person for whom being caring is an important value in their personal life (e.g., as a parent), in their professional life (e.g., as a nurse), and in other domains (e.g., towards animals) is likely to hold being caring as one of their core values and orient their moral conduct around being a caring person. Regardless of whether a person has high personal unity, there is at least one stable aspect of the moral identity goal: its content, even if it varies between situations, is thought of by the person in normative, often universalized terms.

**Motivational Dynamics**

Moral identity motivates behavior (Blasi, 2004a). There is by now an impressive body of research demonstrating that active moral identities can lead to consistent moral behavior (Mullen & Monin, 2016). The self-regulation of moral identity however does not work in isolation from the rest of a person’s conduct. In addition to activating moral identities, past behavior can also have an effect on future motivation. There is a growing literature about moral compensation. Moral compensation refers to the tendency to relax one’s moral standards after demonstrating moral identity and to increase moral efforts after ethical shortcomings (Merritt, Effron, & Monin, 2010). The ironic effect can be observed that experiencing negative moral feedback or committing immoral actions leads to a stronger tendency to behave morally (i.e. moral cleansing), whereas positive moral feedback or demonstrating being a moral person leads to relaxing one’s moral standards (i.e. moral licensing; Blanken, van de Ven, & Zeelenberg, 2015; Effron & Conway, 2015).

For example, Monin and Miller (2001) have demonstrated that after people have symbolized being nonprejudiced, they were more likely to behave in ways that could be construed as prejudiced. After rejecting sexist statements or selecting a woman for a stereotype-neutral job, participants showed a greater preference for men over women as candidates for a stereotypically male job. Similarly, after selecting an African American candidate for a neutral position, participants were more likely to endorse ethnicity as a factor for a job described as stereotypically suited for white people. (Effron, Cameron, & Monin, 2009) replicated this effect, showing that endorsing then-presidential nominee Barack Obama in the 2008 election increased the likelihood of favoring Whites over Blacks especially for individuals with high racial
prejudice. Beyond the realm of prejudice, moral licensing has been demonstrated in a wide array of domains including honesty (Mazar, Amir, & Ariely, 2008), prosocial intentions (Jordan, Mullen, & Murnighan, 2011), consumer choice (Mazar & Zhong, 2010), and donations (Sachdeva, Iliev, & Medin, 2009).

When individuals have the feeling that their moral self is compromised, they tend to restore moral identity via moral cleansing. Interestingly, because being morally pure is metaphorically linked to being physically clean (Zhong & Liljenquist, 2006; Zhong, Strejcek, & Sivanathan, 2010), compensatory behavior can take the form of prosocial behavior, such as donations to charity (Mulder & Aquino, 2013) but also literal acts of physical cleansing (Schnall, Benton, & Harvey, 2008) and attitudes towards cleaning products (M. Gollwitzer & Melzer, 2012).

Longoni, Gollwitzer, and Oettingen (2014) have demonstrated with a wide array of measures that compensation effects regarding moral (pro-environmental) identity impacts behavior, cognition, and perception. In their studies, they had their participants perform an online shopping task at the end of which bogus feedback about the participants’ green behavior was provided. The authors showed that negative feedback increased recycling behavior in an unrelated task after the feedback, whereas positive feedback led to a decrease in recycling. Furthermore, the cognitive availability of constructs related to the green identity was increased after positive feedback and decreased after negative feedback. Finally, the same pattern of results was observed on the perceptual level: participants who were reminded of failure to measure up to their green identity perceived patches of the color green to be greener and participants who felt secure in their green identity perceived them to be less green.

**Motivational dynamics and the Self Model.** The challenge of motivational dynamics for moral identity theories boils down to the question: When can we expect moral consistency and when can we expect moral compensation? Moral compensation is beyond the explanatory reach of the Self Model. The model makes no predictions regarding the effect of past behavior on present choices. Motivational processes in the model are driven by the need for self-consistency (Blasi, 2005). It has been pointed out that moral consistency effects are more prevalent after one thinks about past moral actions in abstract terms indicating commitment to being moral, whereas licensing was observed more after thinking in concrete terms indicating progress towards actually being moral (Mullen & Monin,
This explanation departs from the Self Model and clearly relies on processes that are components of goal striving.

**Motivational dynamics and Social Cognitive Theory.** In the context of Social Cognitive Theory, Lapsley (2016) argued that people have different lay theories about moral character (Dweck & Leggett, 1988). Some individuals have a fixed view of moral identity (entity theorists), while others believe in the malleability and potential for improvement of one’s moral identity (incremental theorists). Entity theories are associated with performance goals and incremental theories are associated with learning goals. Lapsley speculated that moral cleansing might result from individual’s learning goals directed at moral identity and moral licensing from performance goals.

**Motivational dynamics and the goal perspective.** There are explicit models of moral compensation (e.g., acquisition of moral credentials; Monin & Miller, 2001). These are mostly in line with the goal perspective laid out in this article. Past behavior is an indicator of one’s goal progress and motivation to act on a goal is stronger for incomplete goals than for complete ones (Koo & Fishbach, 2008). Symbolic self-completion theory (Wicklund & Gollwitzer, 1982) makes this explicit for identity goals. There are two prerequisites for an individual to act on an identity goal. First, the agent must be committed to the goal. Second, the agent must be lacking in symbols that indicate having attained the identity goal in question. Although people strive to uphold a moral identity, moral behavior is often costly. Thus if people feel secure (complete) in their moral identity they have fewer incentives to behave in a moral way (Sachdeva et al., 2009).

If moral identity is conceptualized as a goal rather than a mere schema, it becomes clear why frequency and recency of moral identity activation sometimes increase moral behavior and sometimes decrease it. The cognitive and motivational consequences of goals are responsive to the state of goal pursuit. Individuals are tuned towards redirecting their resources from fulfilled goals to unfulfilled goals. In contrast to complete goals, incomplete goals lead to a stronger urge to act, higher attention to goal-relevant information, and inhibition of processing goal-irrelevant or distracting information (Kruglanski & Köpetz, 2009). Thus, if moral identity is made salient by indicators of moral identity completeness, this should lead to a decrease in moral motivation. Indicators of goal commitment and goal incompleteness should however both increase moral motivation. Indeed, this seems to be a process underlying moral
Summary and Conclusion: The Goal Perspective

In the present article, we presented a goal perspective on moral identity as a structural and motivational model for the moral self. To account for the variability and impact of moral identity, such a model needs to support the strong agentic features of the moral self that are the focus of other models (e.g., the Self Model; Blasi, 2005). Agency is intimately related to goals. Moskowitz and Grant (2009) defined goals as “the most basic element of control” (p. 2). Adopting this definition, moral identity must either be tied to goals or it must be a goal itself. Treating moral identity as a goal offers a simple model of the moral self, that can explain behavioral and cognitive effects using the tools of motivational psychology tested in other domains. To be viable, a goal perspective of the moral self must 1) be compatible with a pluralistic notion of human moral values to reflect individual differences in moral content, 2) explain the interrelationship of other self-definitions and the moral self, and 3) explain the motivational basis of moral compensation and moral consistency.

The configuration of goal systems differs across individuals. While there is an overlap in the makeup of goals, especially for people who share the same culture, goal associations (between desired end states, alternative goals, means, and other goal-relevant information) are idiosyncratic (Kruglanski & Köpetz, 2009). Cultural differences in moral norms underlie a variety of moral practices and values (Graham et al., 2018; Haidt et al., 1993) that are the basis for the appraisal of the desirability of goals. Culture is one of the main factors that lead to individual differences, but other influences like socio-economic status (Haidt et al., 1993) and personal developmental background (Camacho, Higgins, & Luger, 2003; Higgins & Silberman, 1998) shape both moral values and goal processes as well.

We propose that identity goals are stored in memory structures similar to those proposed by Social Cognitive Theory (Lapsley, 2016). Importantly, as a special class of self-schemata, identity goals always have a motivational force that is central to them. People strive towards being a moral person and this striving is linked to memories of past experience, prototypical situations in which morality is important, emotional responses, behavioral patterns, and cognized moral judgments and deliberation. Via spreading activation, any part of this network could potentially
activate the moral identity goal. Activation of the goal is especially likely when an individual is disrupted in their goal striving and enters a self-evaluative state (i.e., incompleteness; Wicklund & Gollwitzer, 1982).

Symbolic self-completion theory (Wicklund & Gollwitzer, 1982) provides a theoretical foundation that can be used to describe the moral self as an identity goal. People strive towards being a moral person and the commitment to this self-definition determines how central morality is in their life. Higher commitment to the moral identity goal is predictive of persistent moral behavior over time, especially when the focus of observation is a long-term timeframe. This hypothesis is supported by a recent review of moral licensing and consistency effect (Mullen & Monin, 2016). People exhibit consistent moral behavior if they are in an abstract frame of mind and are tuned towards long-term goals. In the short run however moral licensing can be expected. This can be explained in terms of goal completeness versus incompleteness. The urge to work towards a moral identity goal is stronger if the agent is in a state of goal incompleteness and weaker in a state of completeness. Commitment to the moral identity goal, thus, determines the general overall long-term effort put into the moral self-definition. On the other hand, experienced completeness vs. incompleteness in situ determines the moment to moment fluctuations in motivational pressure in people who are committed to a moral identity goal.

Treating moral identity as a goal leads to further specific predictions derived from the goal concept. Goal commitment depends on the desirability of the goal (e.g., transmitted by cultural values), but also on respective self-efficacy beliefs (Gollwitzer & Oettingen, 2012). The goal of being moral should be stronger for those who believe they can act on this goal. A qualitative analysis of interviews with rescuers, bystanders, and Nazis during the holocaust underlines the importance of personal efficacy for moral identity (Monroe, 2008). Bystanders reported lower efficacy beliefs than both rescuers and Nazis. Rescuers and Nazis also differed in their conceptualization of efficacy; in contrast to rescuers, the Nazis construed self-efficacy not as a product of individual independence but as an attunement to historical forces. Such qualitative analyses of historical cases demonstrate the real world consequences of self-efficacy for the construction of moral identity. Future research needs to complement such studies by shedding further light on the underlying processes.

Goal framing is another variable that likely affects moral identity striving. Moral identity goals construed as approach goals (e.g., “I want to be a moral person.”)
can be expected to entail different values and be pursued differently from moral identity avoidance goal (e.g., “I don’t want to be an immoral person.”). Indeed, a general orientation towards moral approach or avoidance seems to be predictive of moral attitudes and political orientation (Janoff-Bulman, Sheikh, & Baldacci, 2008). Regarding identity construction, preferences for approach goals correlate to identity goals based on desired \textit{ideal selves}, whereas preferences for avoidance goals correlate to identity goals based on \textit{ought selves} which one is compelled to pursue because of external demands. A third commonly used distinction in goal framing is between learning and performance goals (Molden & Dweck, 2006). Individuals thinking of moral character as stable may pursue moral identity goals differently (as a performance goal) than those thinking of moral character as malleable (as a learning goal). Exploring different framings of moral identity goals can be a promising direction for future research and possible interventions.

In our discussion, we have not addressed the development of the moral identity goals. A detailed analysis of the developmental steps and properties of this construct would be beyond the scope of this article. It has been suggested (Wicklund & Gollwitzer, 1982) that self-defining goals are acquired when individuals pursue an external, instrumental goal and the goal-directed action is interrupted. Agents would enter a tension state in which they experience a readiness to perform the goal-directed behavior. This perceived readiness is the first step towards a self-defining goal. If the agent can perform a substitute action, and cognitively links the substitute to the original goal, the network of goal, behavior, and substitute form a proto-identity goal that is perceived as a general readiness to act towards this goal. The other component necessary for the formation of an identity goal, according to symbolic self-completion theory is that other people recognize substitute behavior as appropriate symbols (i.e., social reality). It is likely that moral identity goals share features with this general account of identity goal development, but as in the case of a moral identity goal a host of multiple goals are involved the process is likely to be more complex and more empirical work is needed to explicate it.

Measurement of moral identity goals will be challenging if one attempts to capture the multidimensional, overlapping nature of the construct. A promising approach has been proposed by Hannah et al. (2011) based on the assessment of self-complexity (Woolfolk et al., 2004). Instead of traditional questionnaires, the authors propose to use a trait sorting task, in which participants match a set of moral traits to a
set of identities (either created by the participant or provided by the researcher). Traits can be attributed to more than one identity. Such a task would allow researchers to measure which moral traits a participant endorses in a specific identity (e.g., “I, as a father, am caring”). Beyond that, the overlap of identity goals (i.e., how many traits are shared between two identity goals) and the complexity of the identity goal system should be mathematically modelled (Luo, Watkins, & Lam, 2009).

In conclusion, in this article we have reviewed major theories of moral identity. We extend these theories by incorporating a motivation science and self-complexity approach into one integrative theoretical framework. More specifically, we introduced a goal perspective to describe the moral self. In this view, moral identity goals are hierarchically high level, highly overlapping goals. Three propositions are central to this conception of the moral self. First, the moral self is for most people a self-defining goal associated with processes and phenomena typical for goal striving, like self-regulation, compensation, and commitment. Second, the moral identity goal overlaps and interacts with other goals and these other goals can determine the situational content of moral identity. Third and finally, a substantial amount of individual variation in overlapping goals and goal content can be expected. Research on moral identity can benefit from considering the moral self as a goal because the goal perspective offers a robust model for the motivational mechanisms underlying moral identity. But research on identity goals can also benefit from considering the moral self. So far, there has been no systematic work on identity goal networks. Moral identity is perhaps special, because of its numerous associations with other goals. It might, therefore, be an ideal test case to explore identity goal interactions and networks of identity goals.
Research Paper II
The Pursuit of Multiple Simultaneous Identity Goals Depends on the Multifinality of Available Means

Johannes T. Doerflinger¹, Erkan Poyraz¹ & Peter M. Gollwitzer¹²

¹University of Konstanz, Germany ²New York University, USA
Abstract

The present article aims to test under what circumstances the pursuit of an identity goal is hindered by additional active identity goals and under what circumstances identity goal synergy occurs. We hypothesize that interference versus synergy of identity goals depends on the compatibility of active goals and available means. An identity goal that is incompatible with the means at hand can be expected to interfere with the use of these means. In contrast, we predict that multiple identity goals that are supported by the same available means jointly increase effort expended on these means. These hypotheses were explored in two studies. Experiment 1 was designed to test the interaction of two identity goals in a case where the available means was only compatible with one of the goals. Specifically, we investigated whether effort towards the moral identity subgoal of a professional identity goal was hampered by a simultaneously active competence subgoal. Our results suggest that this is indeed the case when no means to satisfy both identity goals are at hand. Participants with an active moral identity goal donated more money to a charitable organization than participants with an inactive moral identity goal. However, when a competence identity goal was active at the same time as a moral identity goal, donations were at the same level as for participants with an inactive moral identity goal. Can goal striving benefit from multiple simultaneously active identity goals if these goals can be pursued with the same means? This question was addressed in Experiment 2 with cultural identity goals of bicultural individuals. Identity goals were activated for no culture, one culture, or both. As a means for both identity goals, the participants could then symbolize their identity by doing voluntary tasks. Both identity goals increased symbolization effort and the most effort was observed for participants with both identity goals active. Taken together, the studies suggest that the pursuit of identity goals can be impeded by and benefit from additional active identity goals depending on the compatibility with available means.

Keywords: Symbolic self-completion theory, multifinality, goal systems, moral identity
The Pursuit of Multiple Simultaneous Identity Goals Depends on the Multifinality of Available Means

The self-identity of individuals is usually composed of several sub-identities (Linville, 1987). In the present article, we explore in two experiments how the configuration of identity networks within an individual influences motivational processes stemming from desired selves. We conceptualize identities as desired future selves an individual is committed to attaining (i.e., identity goals; Wicklund & Gollwitzer, 1982). The central question in the present work is: When does simultaneous identity goal pursuit lead to decreased goal-directed effort and when does it lead to increased effort?

Past research has shown that individuals who are disrupted in their striving for an identity goal increase their effort towards that goal and attempt to symbolize having the desired identity (e.g., Marquardt et al., 2016), a process labeled symbolic self-completion. A disruption of identity goal striving leads to identity goal incompleteness, an unpleasant tension state that motivates agents to seek out symbols of the identity goal and thereby decrease psychological tension. Once enough symbols are acquired and the tension subsides the agent is in a state of identity goal completeness, and effort in striving towards the identity goal is decreased (Gollwitzer & Kirchhof, 1998; Gollwitzer & Wicklund, 1985a; Wicklund & Gollwitzer, 1982).

Thus far, research on symbolic self-completion has only investigated identity goals in isolation. However, individuals typically hold multiple goals at the same time and goals are connected in network-like structures (Shah & Kruglanski, 2000). The present studies are among the first to investigate the interaction of multiple simultaneously active identity goals. We propose that striving towards an identity goal can be impaired by additional active identity goals if the means at hand are incompatible with the additional goals. However, if the available means are compatible with all active identity goals, then all active goals should contribute to the effort invested in the available means. In the present studies, we focus on moral and cultural identity in Studies 1 and 2, respectively.

Identity Goals

Goals allow agents to exert control over their actions, direct mental processes, and interact with the environment in a meaningful way (Moskowitz & Grant, 2009). A goal can be thought of as a mental representation of a desired future
state that the agent intends to realize (Oettingen et al., 2001). This mental representation is stored in memory as an organized knowledge structure connecting the desired end state, means, opportunities to act, and other goal-relevant information (Kruglanski & Köpetz, 2009). Having a goal means that the agent is committed to working towards that goal, which in turn energizes the agent for goal-directed action (Locke & Latham, 2002, 2013).

**Symbolic Self-Completion Theory**

Goals related to the identity of the agent are labeled identity goals and can be distinguished from instrumental goals. Identity goals are aimed at attaining a lasting quality that indicates possession of the desired identity. Instrumental goals are aimed at achieving an external state outside of the agent. External goals are complete and come to an end once the desired state is reached. After that moment they no longer have motivational relevance. In contrast, identity goals do not have a clear endpoint (Wicklund & Gollwitzer, 1982). People who are committed to an identity goal seek out indicators of possessing the identity (i.e., symbols). Symbols can be physical objects associated with the desired identity (Wicklund & Gollwitzer, 1982), displays of academic titles (Harmon-Jones, Schmeichel, & Harmon-Jones, 2009), self-presentation in social-networks (Toma & Hancock, 2013), self-descriptions as possessing the identity (Gollwitzer et al., 1982), social interactions such as persuasion, efforts to acquire the goal relevant skills (Wicklund & Gollwitzer, 1982), or publicly stating the intention to behave as someone possessing the desired identity (Gollwitzer, Sheeran, Michalski, & Seifert, 2009).

If symbols of an identity goal are recognized by others (i.e., they have social reality), the agent enters a state of identity goal completeness (Gollwitzer, 1986; Gollwitzer et al., 2009). Being complete in one's identity lessens the tension driving symbolization efforts. Thus, less compensatory behavior can usually be observed for people with a complete identity (Wicklund & Gollwitzer, 1982). Losing or failing to acquire an identity-relevant symbol leads to a sense of incompleteness, a psychological tension (Gollwitzer et al., 1982) that motivates individuals to compensate by symbolizing the identity. According to symbolic self-completion theory, the symbols of an identity are interchangeable. An agent could thus use any indicator of possessing the incomplete identity to self-symbolize. For example, a person with the identity goal of being a cook who burns a meal accidentally would
feel incomplete as a cook. He could compensate by mentioning the great meals he has prepared in the past, walking around in an apron in front of his guests, or use any other symbol that would signify that he possesses the “cook” identity and thereby restore his identity completeness. If losing an identity symbol leads to incompleteness, alternative symbols can be used to restore a sense of identity completeness (Gollwitzer et al., 1999). Symbolization works only in the domain of the identity goal in question, not on the global level (like the restoration of self-esteem). Thus, although symbols are interchangeable they must be goal-relevant. Unrelated self-affirmation will not lead to identity goal completeness (Gollwitzer et al., 2013).

The effects of Identity goal incompleteness are wide-ranging. For example, Longoni et al. (2014) have shown that identity incompleteness affects behavior, basic cognition, and even visual perception. The authors invited individuals with a high commitment to being environmentally friendly (green identity) to their studies and induced identity goal incompleteness versus completeness with bogus feedback regarding the participants’ shopping behavior. Afterward, participants with an incomplete green identity recycled more and participants with a complete green identity recycled less compared to participants in a control condition (Study 1). Additionally, a complete green identity decreased performance and an incomplete green identity increased performance on a lexical decision task testing the availability of goal-related content (Study 2). Finally, individuals with an incomplete green identity perceived patches of color as greener than control participants, and individuals with a complete green identity perceived them as less green (Study 3).

Furthermore, active identity goals can override other concerns such as taking others’ perspective into account in social interactions (Gollwitzer & Wicklund, 1985b). Along this line, Marquardt et al. (2016) have shown that an incomplete professional identity goal can lead to symbolization with immoral means. Importantly for the present research, Marquardt et al. provided means that served the professional identity goals of their participants but not their moral identity goals. They did not activate the moral identity goal of their participants.
**Goal Systems**

If multiple goals are taken into account, it is important to consider how the goals relate to available means and to each other. According to goal systems theory (Kruglanski et al., 2018; Shah et al., 2003), goals are represented as associative cognitive networks (i.e., goal systems). The core feature of a goal system is that goals are interconnected with other goals and with means. The connecting links can be facilitatory (e.g., encountering viable means activates goals; Shah & Kruglanski, 2003) or inhibitory (e.g., inhibition of conflicting goals, goal shielding; Shah et al., 2002). Goals are conceptualized as hierarchically organized in Goal Systems Theory (Shah et al., 2003), similar to other hierarchical models of goals (e.g., control theory; Carver & Scheier, 1982). Lower order goals are connected as means to higher-order goals. In addition, goals can be associated laterally on the same hierarchical level (see Figure 1). Consistent with this framework, we argue that identity goals represent a high level in the goal hierarchy and direct behavioral intentions represent a relatively low level. Links between different hierarchical levels are typically excitatory, while connections at the same level are mainly inhibitory, especially for competing goals (Shah et al., 2003).

![Figure 1](image)

**Figure 1.** An example of a goal system. A higher order goal is served by four lower order subgoals. Subgoal 1 and 2 have a shared means. Subgoal 3 has a facilitatory link with Subgoal 2 and an inhibitory link with Subgoal 4.

**Multifinality.** Multiple means can be associated with the same goal (i.e., equifinality) and a means can serve multiple goals (i.e., multifinality). The number of equifinal means is negatively correlated with the associative strength between any one means and the corresponding goal. Similarly, the number of goals associated with a
multifinal means correlates negatively with the strength of the link between said means and any one goal (Kruglanski et al., 2018; Y. Zhang et al., 2007). Interconnectedness, equifinality, and multifinality are cognitive foundations of goals and their functions are in line with basic cognitive mechanisms. In contrast to other cognitive structures, however, goals have motivational properties. Goals carry commitment and commitment to a goal can transfer to means associated with the goals (Kruglanski & Shah, 2000). Multifinal means advance multiple goals. Thus, the preference for these means depends on the value of all associated goals. If all else is equal, the subjective value of multifinal means should be higher than the subjective value of unifinal means serving a subset of the goals associated with the multifinal means (Kruglanski et al., 2013). Which means an individual chooses depends on additional factors besides the values of associated goals. For example, self-regulatory mode (Orehek, Mauro, Kruglanski, & van der Bles, 2012), the feasibility of identifying multifinal means, and differences in incentives for the individual goals (Köpetz et al., 2011) moderate the preference for means. Importantly, when goals are simultaneously active, they affect each other’s pursuit. Agents prefer means to a focal goal that also support additional background goals, even if the background goals are implicit (Chun et al., 2011). In line with general goal theories (Martin & Tesser, 2009), this preference depends on goal activation. Chun et al. (2011, Study 4) have shown, that the preference for multifinal goals disappears after background goals have been satisfied (i.e., deactivated) by alternative means.

Identity Goals in Interaction

Thus far, research on identity goals has focused on the pursuit of single goals. We propose that just as simultaneously active instrumental goals can facilitate and inhibit each other, so can identity goals. Identity goals are high-level goals (Carver & Scheier, 1982) but the hierarchical relation between identity goals is less defined. Some identity goals may operate on the same level (e.g., competence and morality goals), others may be in a one-directional hierarchy (e.g., the basic need for competence is superordinate to the identity goal of being proficient in math), yet other identity goals may partially overlap and serve as means for each other (e.g., through shared sub-goals). For overlapping identity goals, lower-level multifinal goals are viable means. This is not the case for goals without overlap. We predict that
overlapping active identity goals will increase effort towards shared sub-goals, while non-overlapping goals will inhibit each other.

Present Research

In the present studies, we explore the simultaneous pursuit of multiple identity goals. We predict that the activation of two identity goals at the same time will only lead to an increase of symbolization effort if both identity goals share available means for completion. If the means at hand can only serve to symbolize one active identity goal but not the other, then activation of a second identity goal should interfere with symbolization of the first one (Experiment 1). In contrast, if attainment of two identity goals can be symbolized via the same means while the agent is incomplete in both domains, symbolization effort should be greater than if the individual is incomplete in a single identity domain associated with that symbol (Experiment 2).

In the first experiment, we explore interference with identity goal symbolization by a second active identity goal. Research on symbolic self-completion has demonstrated that individuals who suffer a setback in their pursuit of an identity enter a state of identity goal incompleteness. Being incomplete in one’s identity motivates symbolization of the identity to reenter a state of identity goal completeness (Moskowitz et al., 2011; Wicklund & Gollwitzer, 1982). If the individual is incomplete regarding more than one identity (i.e., multiple identity goals are active) and no multifinal means are available to symbolize the incomplete identities simultaneously, then the identity goals likely interfere with each other because each individual identity goal requires cognitive and motivational resources. This inhibition should cease once competing identity goals have been completed through alternative means. We tested this hypothesis in the identity domain of “becoming a psychotherapist”. We chose to investigate the professional identity goal of becoming a psychotherapist because many students in the field of psychology hold this goal and it contains both competence and moral identity subgoals. While this professional identity requires factual knowledge and skills in clinical psychology, a good psychotherapist is required to live up to certain moral standards. One of these standards is being nonjudgmental of people with mental illnesses. In the first experiment, we pitted the moral identity goal of being unprejudiced against people with mental illness (moral identity goal) against a professional identity goal of being
knowledgeable about mental illness (competence identity goal). The available means for symbolization were only relevant for morality (donating money to charity). Thus, having an active competence goal in addition to the morality goal should reduce symbolization of being moral. The goal system investigated in Experiment 1 is depicted in Figure 2a.

**Figure 2.** Goal systems investigated in Research Paper II. a) the goal System in Experiment 1. Donating served as a means for the moral identity goal but not for the competence identity goal. We expect an active competence identity goal to interfere with symbolization of the moral identity goal. b) the goal system in Experiment 2. Both the Turkish identity goal and the German identity goal could be symbolized by the same means (voluntary tasks). In this case, we expect goal synergy.

In the second experiment, we test the additive motivational effect of two identity goals that can be symbolized by the same means. If the available means are multifinal and serve both active identity goals, then a second incomplete identity goal can be expected to energize further efforts towards symbolization beyond the motivational push of the first incomplete identity. This configuration of identity goals is shown in Figure 2b. We investigate this hypothesis in the domain of the cultural identity of bicultural individuals with a background in both the German as well as the Turkish culture. People with roots in two cultures can identify with both and often have to manage two cultural identities at the same time (Benet-Martínez & Haritatos, 2005). While in some situations the cultural identities are in conflict, in others they may be compatible and symbolized with the same means. If the available means is multifinal in the service of both cultural identity goals, then we predict that the goals
should not interfere with each other. Rather, symbolization efforts should be increased in this case.

**Experiment 1: Unifinal Moral Symbolization is Hindered by Competing Goals**

The first experiment was framed as a study on the suitability of psychology students as future psychotherapists. The participants performed a task supposedly measuring their general aptitude for the profession and their implicit stereotypes (Blair, 2001) towards individuals with mental illnesses. The participants received bogus feedback indicating that they were unsuitable to become psychotherapists because they showed an implicit bias against people with mental illness conditions. This feedback poses a threat to the participants’ identity goal of becoming a psychotherapist and leads to a state of identity goal incompleteness. Importantly, the threat has both moral as well as professional implications. After incompleteness in the identity goals had been established, some of the participants had the opportunity to self-symbolize, either by demonstrating moral behavior that did not symbolize professional qualifications (i.e., rating statements about everyday moral behavior) or by demonstrating professional competence that was unrelated to morality (i.e., answering knowledge questions). The opportunity to symbolize restores identity goal completeness, thus specific identity subgoals can be deactivated. This procedure resulted in three experimental groups: participants with both the moral identity goal and the competence identity goal active (moral incomplete, competence incomplete), participants an active moral identity goal and an inactive competence identity goal (moral incomplete, competence complete), and participants with an inactive moral identity goal and an active competence identity goal (moral complete, competence incomplete). At the end of the experiment, participants had the opportunity to donate money to a charitable organization that was relevant to the field of psychotherapy. Donating money serves as a symbol of moral identity. Previous studies have shown that morally incomplete agents use donations to compensate for moral shortcomings (Sachdeva et al., 2009). While donations are a viable symbol for moral identity, donating money does not symbolize one’s competence. Thus, having an active competence goal in addition to a moral identity goal should reduce donations. Only if the competence goal is satisfied by other means should this goal-interference effect disappear (see Figure 2a for a visualization of this goal system). Both the moral identity goal and the competence goal are subgoals of the psychotherapist identity
goal. Thus, we predict goal completion effects to be stronger for individuals with a strong commitment to becoming a psychotherapist compared to individuals with a weak commitment.

Moral identity has long been described as a bridge between moral intentions and moral behavior (Blasi, 1980; Shao et al., 2008). Being moral is perhaps the most essential characteristic of human identity (Strohminger & Nichols, 2014) and morality has even been described as a basic psychological need (Prentice et al., 2018). Past research has conceptualized moral identity as a self-schema in which moral values, knowledge, and behavioral intentions are organized. The more accessible the moral self-schema is in general, the more central is morality to a person, which can be measured as an individual difference (Aquino & Reed, 2002; Lapsley, 2016). Moral identity centrality is treated as a trait-like property that affects behavior in interaction with the respective situation (Aquino et al., 2009). Studies have shown that individuals high in moral identity centrality tend to donate more than individuals low in moral identity centrality (Aquino & Reed, 2002), especially after behaving in an immoral manner (Mulder & Aquino, 2013). Thus, we control for moral identity centrality.

We expect participants who did not have the opportunity to symbolize their identity goal and participants who symbolized morality in a different domain to donate the smallest amount. Participants who had no opportunity to symbolize are likely to experience interference from the incomplete competence identity goal. Participants who have already symbolized being moral in another domain would not need to complete their moral identity. Participants who had the opportunity to compensate for their professional incompleteness by demonstrating their knowledge in the field should perceive becoming a psychotherapist as a feasible goal. These participants are complete in regard to the competence goal. Because they are, however, still incomplete in the moral domain, these participants are expected to self-symbolize morality. Thus, participants who were professionally complete but morally incomplete are expected to donate more than the other two groups. Differences should be more pronounced for participants who are committed to their professional identity goal.
Methods

Participants and design. Identity completeness was manipulated in a between-subjects design and the participants were randomly assigned to one of three experimental conditions: 1) no symbolizing (moral and competence identity incomplete), 2) moral symbolizing (moral identity complete, competence identity incomplete), and 3) competence symbolizing (moral identity incomplete, competence identity complete). One hundred and thirteen psychology students completed the experiment. Of these, six participants were suspicious about the feedback offered and indicated that they did not take the experiment seriously. These participants were excluded from the analysis, resulting in a sample of 107 participants (94 female) with an average age of 21.3 (SD = 3.6, Range 18 – 43). A power analysis with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) showed that with this sample size, medium to large effects can be detected with a probability of .80. The participants took part for course credit and a financial compensation of 2 €, which they could either donate to a charitable organization or keep for themselves at the end of the experiment.

Procedure. Upon signing up for an appointment in the lab, participants were asked to complete a brief online questionnaire consisting of a German translation of the moral identity scale (Aquino & Reed, 2002) and demographic questions. The moral identity scale is designed to assess moral identity centrality on two dimensions: internalization (i.e., the degree to which morality is an integral part of the self) and symbolization (i.e., the degree to which an individual is inclined to demonstrate their morality to others). We assume that even if morality overlaps with other identity goals, a global sense of the moral self should affect behavior in most domains. Specifically, moral identity internalization has been found to correlate with donation behavior (Aquino & Reed, 2002).

The participants were told that they would receive 2 € in addition to course credit as compensation for taking part in the online survey. One day to two weeks after the participants filled out the online questionnaire, the second part of the experiment took place in the lab. Once in the lab, the participants were told that the purpose of the Experiment was to investigate how apt psychology students at the University of Konstanz were to become psychotherapists. Commitment to this identity goal was measured on two 9-point Likert Scales (“How important is it to you to become a psychotherapist?”; “How bad would it be if you could not become a
psychotherapist?”). The scales were significantly correlated ($r = .464, p < .001$) and their mean was used as an overall measure of commitment to the psychotherapist identity goal.

At the beginning of the experiment, identity incompleteness was induced for all participants. This was done by having participants take a modified version of the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998), a tool designed to measure implicit associations such as biases and stereotypes against certain groups. For our Experiment, we created a version of the IAT designed to measure attitudes towards individuals with mental illnesses. Irrespective of their actual performance, all participants received feedback in form of a diagram displaying low performance in the test compared to the average psychotherapist, and a message on the computer screen indicating that they had demonstrated a low aptitude for the psychotherapist profession.

After the bogus negative feedback, the participants were randomly assigned to one of the experimental conditions. In all conditions, the participants were asked to provide answers to a series of questions and in all conditions, they received positive feedback on their performance, regardless of their actual answers. The topic of the questions varied to represent different domains of self-symbolizing behavior. In the no symbolization condition, the participants could complete a knowledge questionnaire about marine mammals. This task was deemed neither relevant for the competence nor the moral aspect of a psychotherapist. In the moral symbolization condition, the participants completed items from the *civic moral disengagement scale* (Caprara, Fida, Vecchione, Tramontano, & Barbaranelli, 2009), a questionnaire covering moral behavior in various everyday situations. Importantly, the content of the questionnaire fell into the moral domain but was unrelated to psychotherapy. In the competence symbolization condition, the participants were asked to answer knowledge questions from the domain of clinical psychology. This task covered the competence but not the moral aspect of the psychotherapist identity.

Upon completion of the questionnaires, participants were told that they had reached the end of the Experiment and they were reminded that they would get 2 € for their participation in the online survey. At this point, they were offered the opportunity to donate any portion of that money to *vivo international*, a charitable organization that offers help to traumatized individuals and training for psychotherapists in trauma-therapy. The organization was chosen because it
represents a moral cause highly relevant to the domain of psychotherapy. After the donations were recorded, the participants were paid the respective amount and thoroughly debriefed. The debriefing was conducted in writing and verbally, especially emphasizing that all of the received feedback was bogus and that nothing had been found out about the participants’ suitability as psychotherapists whatsoever.

**Results**

An ANCOVA with the identity incompleteness conditions as the predictor, commitment, as well as moral identity internalization as covariates, and the donated money as the criterion was overall significant ($F(4, 102) = 4.740, p < .001, \eta^2_p = .157$) and revealed significant effects of both identity incompleteness ($F(2, 102) = 4.009, p = .021, \eta^2_p = .066$) and moral identity internalization ($F(1, 102) = 13.171, p < .001, \eta^2_p = .103$) on the amount of money participants donated to charity. No significant effect of commitment was observed ($F(1,102) = 1.328, p = .252$). Planned contrast tests using Helmert contrasts compared the professional symbolization group to the other two groups and the moral symbolization to the no symbolization group. The tests revealed that donations in the professional compensation group were significantly higher than in the other experimental conditions ($t(103) = 2.67, p = .009$), whereas no difference was found between the moral compensation and the no compensation group ($t(103) = 0.18, p = .857$). Figure 3 illustrates this pattern.

![Figure 3. Experiment 1: Donations in € by experimental conditions. Error bars represent the 95% confidence interval.](image-url)
We predicted stronger effects of goal completion for participants with high commitment compared to low commitment. To test this hypothesis, a general linear model (GLM) with the experimental condition, commitment, and their interaction as predictors and donations as the criterion was calculated. The overall model was significant \( F(5,101) = 2.736, p = .023, \eta^2_p = .119 \) The main effect of the experimental condition was marginally significant, \( F(2,101) = 2.879, p = .061, \eta^2_p = .054 \). Commitment had no significant main effect, \( F(1,101) = 0.011, p = .918, \eta^2_p < .001 \). Importantly, the interaction of the experimental condition and commitment was significant, \( F(2,101) = 4.095, p = .019, \eta^2_p = .075 \). The interaction pattern is visualized in Figure 4. As can be seen in Figure 4, the overall effect of identity completeness vs. incompleteness was most pronounced for participants who were highly committed to being a psychotherapist. Committed participants with a complete professional identity and an incomplete moral identity donated more than the other groups, whereas the committed participants with both identities incomplete donated the least amount of money.

Figure 4. Experiment 1: Donations in € by experimental conditions and commitment to being a psychotherapist. Shaded areas represent the 95% confidence interval.
Discussion

In our first experiment, an active competence identity goal interfered with the pursuit of the moral identity goal. The available means served only the moral identity goal. A threat to the aspired-to identity of becoming a psychotherapist including both a moral as well as a competence identity subgoal led to less symbolization if participants did not have the opportunity to symbolize their competence in the profession. This was evidenced by lower donations of participants who could only complete a task that was unrelated to professional competence and morality. Participants who symbolized their morality in an unrelated domain donated similarly low amounts. In comparison, participants who answered questions about clinical psychology — thereby symbolizing their competence identity goal but not their moral identity goal — donated subsequently more money to charity.

Previous studies have shown that moral identity internalization predicts moral behavior (Aquino & Reed, 2002; Reed & Aquino, 2003; Sage et al., 2006). These results were conceptually replicated in our experiment. Participants’ moral identity internalization was positively related to the amount of money donated to charity. In the present case, behavior was clearly influenced not only by self-schemata but also by goal striving processes. A new contribution of the present study relevant for moral identity research and identity theories, in general, is the exploration of two simultaneously active identity goals. Our findings demonstrate that the pursuit of moral identity and other identity goals is interdependent. Moral compensation research predicts that incompleteness in the moral domain should increase moral behavior (Brañas-Garza, Bucheli, Espinosa, & García-Muñoz, 2013; Mulder & Aquino, 2013). While this was the case for participants with no active competence goal, participants for whom both competence and moral identity goals were active did not increase their donations. We interpret these findings as evidence for the interplay of simultaneously active identity goals. Further support for this interpretation is provided by the interaction of goal commitment and identity goal incompleteness. Individuals who were uncommitted to the superordinate goal of being a psychotherapist did not exhibit compensatory symbolization effects. In fact, there was no difference between identity completeness conditions for uncommitted individuals. In contrast, individuals who were committed to the superordinate identity goal symbolized their incomplete moral identity, but only when their competence identity goal was not interfering with their moral identity goal.
Alternative explanations for the results of Experiment 1 are possible. This is mostly due to the fact that the experiment has no condition in which both identity goals are complete. We would predict that if both the competence and the moral identity goal were satisfied, little motivation for symbolization would exist. Thus, if our interpretation is correct, participants complete in a professional and moral identity would relax their moral standards. In the present case, this condition was omitted to save resources, as the predicted results would be the same as in the moral complete condition and the moral/professional incomplete condition. For the sake of clarity, future studies on this subject should include such a condition. Furthermore, future studies could most likely simplify the method to achieve the desired pattern of goal activation. Instead of inducing incompleteness for all participants and then offering a chance to self-symbolize as an experimental manipulation, identity goals could simply be activated by targeted incompleteness inductions. In sum, a future conceptual replication and extension of Experiment 1 would benefit from a 2 by 2 design in which incompleteness in two identity goals is manipulated independently.

We propose that the pattern of results in our study is not unique to the moral realm, but generalizes to other identity domains. Future research should investigate identity goal interference of different types of identity goals. We further predict that whether active identity goals interfere with each other or supplement each other depends on the available means for symbolization. Research on goal systems has shown that people with multiple active goals prefer multifinal means to unifinal means in order to avoid goal conflict (Köpetz et al., 2011). If, as in the present case, no multifinal means are available, additional active goals beyond the first can be expected to interfere with striving for this goal. Even if multiple active goals do show some overlap with each other, they are competing for motivational and attentional resources. Thus, active goals are in conflict unless multifinal means can be used to resolve this conflict (Kruglanski et al., 2018; Shah et al., 2003).

Our findings have important implications for policy making and interventions as they highlight that moral identity goals are contextualized. This is reflected in two important properties of moral identity goals that affected behavior in the current study. First, whether an individual seizes a means to pursue a moral identity goal depends on the compatibility of the means and other active goals. Individuals who are morally incomplete and at the same time incomplete in other self-relevant domains (e.g., vocational competence) may be overwhelmed by the demands
of simultaneous pursuit of these goals, especially if the available means serve only the moral identity goal. Organizations that want to implement ethical training for their members should keep this in mind because the study suggests that such interventions could be ineffective if they activate other identity goals besides the moral identity goal. If ethical interventions are part of a larger program (e.g., employee training), it could be worthwhile to provide the participants of the intervention an opportunity to symbolize other relevant identities beforehand so that they can concentrate their resources on moral identity goals. Second, moral identity can at least in some instances function as a subgoal of other identity goals. In our study, this is reflected by the fact that the effects of goal activation depended on the participants’ commitment to the identity goal of becoming a psychotherapist. Stronger effects of goal activation were observed for highly committed participants. We can conclude that for these participants morality was integrated into their professional identity goal. Future research should test to what degree content and activation of moral identity goals depend on other active identity goals. What it means to be moral may be dependent on the various identity goals with which morality is connected.

In Experiment 1 the participants could only symbolize being moral with unifinal means (i.e., means that did not serve the competence identity goal). It is likely that multiple active goals do not interfere with each other if the available means are compatible with all active goals. In Experiment 2, we test whether two incomplete identity goals that can be symbolized via the same means will increase symbolization efforts.

Experiment 2: The Additive Effect of Cultural Identity Goals on Multifinal Means

In the second experiment, the effect was investigated of two simultaneously incomplete identity goals on a multifinal means that could be used to symbolize both goals. We hypothesize that if multifinal means are available, then additional active identity goals can increase symbolization efforts because the overall motivation is increased by the active goals. While multiple goals are competing for resources and are potentially incompatible, if no multifinal means are available (Shah et al., 2003) the agent can complete multiple goals with the same means if such multifinal means are at hand, thus goal-synergy can be expected.
This hypothesis was tested in the context of the cultural identity of bicultural individuals. Huynh, Nguyen, and Benet-Martínez (2011) defined bicultural individuals as “those who have been exposed to and have internalized two cultures” (p. 828). These can be immigrants, refugees, ethnic minorities, or simply individuals who had extensive contact with two cultures and have integrated both cultures in their self-concept. From an identity goal perspective, bicultural individuals have two cultural identity goals. In the present work, we tested identity goal synergy of cultural identity goals in individuals with both a German and a Turkish cultural background. We specifically predict that individuals who are incomplete in both cultural identity goals put more effort into a multifinal means in service of both goals than individuals who are only incomplete in one cultural identity goal. Individuals who are complete in both identity goals should demonstrate the least symbolization effort. Symbolization was assessed as voluntary completion of additional tasks at the end of the experiment. See Figure 2b for a depiction of the goal system in Experiment 2.

Furthermore, incomplete cultural identity goals should have a distinct effect on the subjective feeling of incompleteness, measured as a discrepancy between current and general subjective belongingness to the respective culture, and the desire to reduce this discrepancy. Thus, we predict that incompleteness induction leads to a feeling of lacking the respective cultural identity and a desire to belong to the culture independently for each cultural identity goal. We measure these subjective feelings separately for the German and the Turkish culture.

Methods

Participants and Design. One hundred and fourteen individuals with a German-Turkish bicultural background participated in the experiment (38 female, age $M = 29$, $SD = 10.1$, range: 16 – 60) for a compensation of 5 €. The experiment had a 2 (German identity: complete vs. incomplete) by 2 (Turkish identity: complete vs. incomplete) between-subjects design. A power analysis with G*Power (Faul et al., 2007) showed that with this sample size, medium to medium to large effects can be detected with a probability of .90.

The experiment was advertised in university groups for Turkish students at two German Universities, at several bicultural events in southern Germany organized by local communities, at a Turkish sports club in Germany, in schools, and in a mosque. For practicality most experimental sessions were conducted outside of the
lab at locations easily reachable by the participants (e.g., in the classroom of a local school). All experimental sessions were conducted in a separated room where participants could work undisturbed on the experiment with an experimenter present. The experimenter was himself a German-Turkish bicultural student.

**Procedure.** The experiment was presented to participants on tablet computers. At the beginning of the experiment, the participants were randomly assigned to experimental conditions. Then, they could select whether they wanted to perform the experiment in German or Turkish. Language selection was offered because — although all participants were bicultural — it was expected that many participants were only fluent in one of the two languages. Being able to speak the corresponding language is one marker of possessing a cultural identity; however, it is not the only one. We decided not to limit the experiment to only bilingual participants.

After the language selection, participants were asked to indicate their feeling of belongingness to the German and the Turkish culture (“I feel like I belong to the German (Turkish) culture.”) on two visual analog scales with the endpoints “not at all” and “very much”. The order of these questions was randomized. This measure served as a baseline for cultural belongingness and allowed us to test whether experimental groups differed systematically in their sense of belonging before the incompleteness manipulation.

Next, the incompleteness and completeness conditions were established. As an incompleteness induction, participants had to write a short text about a personal experience in which they felt like they did not belong to the respective culture. The participants were asked to be as detailed in their descriptions as possible and describe the location and events of the experience as well as their feelings and thoughts in the situation. Short vignettes of typical experiences of bicultural individuals were provided as examples to help participants come up with their own experiences. The same procedure was used to induce identity goal completeness, except that the participants were asked to write about a situation where they felt like they did belong to the respective culture. Each participant had to write one text about an experience with the German culture and one text about an experience with the Turkish culture. The order of the texts was randomized.

Directly after the incompleteness and completeness induction tasks, participants were asked to what degree they felt right now, at the moment, whether
they belonged to the respective cultures (“I feel right now like I belong to the German (Turkish) culture.”) and how much they would like to belong to the respective cultures (“I would like to belong to the German (Turkish) culture.”). The responses could be given on two visual analog scales with the endpoints “not at all” and “very much”. The difference scores of the baseline and the current belongingness rating after the incompleteness tasks served as a manipulation check. If the manipulation was successful, participants in the incomplete conditions should indicate a decrease in their sense of belonging for the incomplete cultural identity goal. The difference between current belongingness and the desire to belong to the respective cultures is a secondary indicator of motivation and serves as a dependent variable.

In the next step, symbolization via unifinal means was assessed. A list of activities was shown, consisting of six domains based on culture dimensions found in earlier work on cultural comparison (Hofstede, 1983, 2011; Trompenaars & Hampden-Turner, 2011). The domains were: language, orientation towards tradition and future, self-restraint, individualism vs. collectivism, uncertainty, and religion. These dimensions were selected based on a pretest with an independent sample of 58 German-Turkish bicultural individuals. For each domain three items were generated, one representing the German culture, one representing the Turkish culture, and one neutral item. For example, in the domain “language” the items were: “I would like to listen to German music or watch a German movie.”, “I would like to listen to Turkish music or watch a Turkish movie.”, and “I would like to listen to the radio, or go to the movie theatre.”. The participants were asked to select four activities out of the six domains that they would like to perform.

Finally, participants were presented with the opportunity to symbolize their cultural identity with multifinal means. At the end of the experiment, the participants were told they could answer optional questions on a voluntary basis. Answering these questions had no benefit beyond self-symbolizing for the participants. Questions were presented in blocks of four questions per page with six blocks of questions. The questions were based on the cultural dimension developed by Hofstede (1983, 2011). Self-descriptions on culturally relevant dimensions could be used to symbolize both a German and a Turkish identity. The participants could choose before they answered any questions and at the end of each block to skip to the end of the experiment. If they wanted to answer more questions, they had to actively select this option before proceeding to the next block. This was designed to make the task more tedious and
increase the incentive to skip to the end. The number of completed blocks serves as our indicator of symbolization effort. If the participants had skipped to the end or answered all questions, demographic information was assessed including age, sex, nationality, migration status, and place of birth. Afterward, they were thoroughly debriefed, thanked, and paid for participating.

Results

A priori differences. The selected language had no effect on our dependent variable \( F(1, 112) = 0.333, p = .565 \) and will not be included in further analyses. Belongingness to the German and Turkish culture before the incompleteness and completeness induction did not differ between the experimental conditions \( (F_s < 1.240, ps > .300) \). A paired samples T-test resulted in no significant difference between a priori belongingness to the German \( (M = 76.74, SD = 20.55) \) and Turkish \( (M = 72.68, SD = 31.90) \) cultures, \( t(113) = 0.524, p = .602 \).

Change in belongingness and desire to belong. First, the differences between the belongingness ratings before and after the experimental manipulation were calculated. These difference scores represent the change in the sense of belonging. Two two-factorial ANOVAs were calculated with the experimental conditions and their interaction as predictors and change in belongingness to the cultures as the dependent variables. The change in belongingness was significantly predicted by the models, both for the Turkish culture \( F(3,110) = 26.864, p < .001, \eta_p^2 = .423 \) and the German culture \( F(3,110) = 20.312, p < .001, \eta_p^2 = .356 \). In the model predicting change in belongingness to the Turkish culture, the main effect of Turkish identity incompleteness was significant, \( F(1, 110) = 75.440, p < .001, \eta_p^2 = .407 \). The main effect of German identity incompleteness was not significant, \( F(1,110) = 1.093, p = .298, \eta_p^2 = .010 \). In addition, the interaction of both incompleteness conditions was significant, \( F(1,110) = 5.170, p = .025, \eta_p^2 = .045 \). In the model predicting the change in belongingness to the German culture, the main effect of German identity incompleteness was significant, \( F(1, 110) = 60.458, p < .001, \eta_p^2 = .355 \). Neither the main effect of Turkish identity incompleteness \( F(1,110) = 0.291, p = .591, \eta_p^2 = .003 \) nor the interaction were significant \( F(1,110) = 0.249, p = .619, \eta_p^2 = .002 \). Figures 5 and 6 depict the pattern of means in the experimental conditions. As can be seen in, the sense of belonging to the Turkish culture decreased for participants in the Turkish
incomplete conditions (Figure 5), whereas the sense of belonging to the German culture decreased for participants in the German incomplete conditions (Figure 6). The effect of incompleteness in the Turkish identity on the sense of belonging to the Turkish culture was more pronounced for individuals in the German identity complete condition.

Figure 5. Experiment 2: Change in the sense of belonging to the Turkish culture after the experimental manipulation. Error bars represent the 95% confidence interval.

Figure 6. Experiment 2: Change in the sense of belonging to the German culture after the experimental manipulation. Error bars represent the 95% confidence interval.
In a next step, difference scores for the current belongingness and the desire to belong were calculated. These scores represent a discrepancy between the desire and the current state. The discrepancy scores were used as criteria in a two-factorial ANOVA with the experimental conditions and their interaction as predictors. The models were significant for the Turkish ($F(3,110) = 4.449, p = .005, \eta^2_p = .108$) and the German ($F(3,110) = 6.784, p < .001, \eta^2_p = .156$) culture. The GLM predicting the discrepancy for the Turkish identity revealed a significant main effect of Turkish identity incompleteness ($F(1,110) = 12.219, p = .001, \eta^2_p = .100$). Neither the main effect of German identity incompleteness ($F(1,110) = 0.003, p = .957$) nor the interaction ($F(1,110) = 1.392, p = .241$) were significant. In the GLM predicting the discrepancy for the German identity, the main effect of German identity incompleteness was significant, $F(1,110) = 15.910, p < .001, \eta^2_p = .127$. In addition, Turkish identity incompleteness had a significant main effect, $F(1,110) = 4.009, p < .048, \eta^2_p = .035$. The interaction did not reach significance ($F(1,110) = 0.753, p = .387, \eta^2_p = .007$). The pattern of means can be seen in Figures 7 and 8. The discrepancy of the desire to belong to the Turkish culture and the current sense of belonging was higher in the Turkish incomplete conditions than in the Turkish complete conditions (Figure 7), and the discrepancy for the German identity was higher in the German incomplete conditions (Figure 8).

![Figure 7](image_url)

**Figure 7.** Experiment 2: Difference between the desire to belong the Turkish culture and current sense of belonging. Error bars represent the 95% confidence interval.
Figure 8. Experiment 2: Difference between the desire to belong the German culture and current sense of belonging. Error bars represent the 95% confidence interval.

**Symbolization effort.** The experimental conditions had no significant effect on the selection of unifinal symbols, $F_s < 1.99, ps > .16$, indicating that the groups did not differ in the number of typically Turkish and typically German actions they check on the list. For multifinal symbols, a two-factorial ANOVA was estimated predicting the number of the voluntary task with the experimental conditions and their interaction. The model was overall significant, $F(3,110) = 17.433, p = .010, \eta^2_p = .098$. Both German ($F(1,110) = 4.326, p = .040, \eta^2_p = .038$) and Turkish ($F(1,110) = 7.567, p = .007, \eta^2_p = .064$) identity goal incompleteness had significant main effects on the number of voluntary tasks worked on at the end of the experiment. The interaction of the variables had no significant effect, $F(1,110) = 0.285, p = .595, \eta^2_p = .003$. Figure 9 depicts the group means. As can be seen in the figure, incompleteness in both identity goals increased the number of voluntary tasks.
Discussion

The main results of the experiment are consistent with the symbolic self-completion literature (Gollwitzer & Kirchhof, 1998; Gollwitzer & Wicklund, 1985a; Wicklund & Gollwitzer, 1982) — incompleteness in an identity goal leads to compensatory symbolization efforts. The manipulation checks clearly show that reflecting on negative cultural experiences led to a culture-specific lower sense of belongingness. Individuals incomplete in their Turkish identity goal felt as if they did not belong to the Turkish culture and individuals incomplete in their German identity goal felt less belongingness to the German culture. The discrepancy between current belongingness and the desire to belong was also culture-specific. This is strong evidence for two distinct cultural identity goals (or two distinct subgoals of a higher order belongingness goal).

Importantly, the participants seized multifinal means of symbolization that were compatible with both cultural identity goals. Consequently, once activated via incompleteness, the German and Turkish identity goals did not inhibit one another. Instead, they had an additive effect on symbolization. As can be seen in Figure 9, individuals incomplete in both identity goals did the highest number of voluntary tasks.

Figure 9. Experiment 2: Symbolization effort, indicated by the number of voluntary tasks. Error bars represent the 95% confidence interval.
tasks, followed by individuals incomplete in only one identity goal. Those with both identity goals completed did the least amount of voluntary tasks.

Incompleteness did not affect items checked on a list of preferences. This variable was designed to measure unifinal means. However, there are two possible explanations for the observed lack of an effect. First, it is possible that unifinal means are rejected when multiple goals are salient. The activities the participants could select were presented in a list in which groups of three items belonged to the same domain of activity with one item corresponding to German culture, one item corresponding to Turkish culture, and one culturally neutral item (e.g., listening to German music, listening to Turkish music, and listening to the radio). It is possible that the participants perceived these items to be mutually exclusive. If this was the case, then selecting a typically German item implies rejecting a typically Turkish item. However, if this was the case, one would expect participants who were only incomplete in one domain to choose more items belonging to that domain. This was not the case in our sample. Second, the participants might not have seen the list as suitable means of symbolization. Effective self-symbolization requires social reality (Gollwitzer et al., 2013; Gollwitzer & Wicklund, 1985a). Because the experiment was anonymous, the participants’ choices of preferences were not registered by others. In contrast, effort in the multifinal task was readily observable — the more tasks the participants worked on, the longer they took and this was observed by other participants and the experimenter. Therefore, the voluntary tasks are a more effective identity symbol. We hypothesize that individuals prefer multifinal symbols when multiple goals are active. However, this conclusion cannot be clearly drawn from our results because alternative explanations remain. Future studies should explore preferences for unifinal and multifinal symbols under identity completeness and incompleteness. Based on the current findings, social reality should be taken into account.

The analyses of the participants’ sense of cultural belongingness and the desire to belong shed further light on the identity goal processes involved. While the identity goals had an additive effect on the mutually shared multifinal means, their effect on the sense of belongingness and the desire to belong was identity specific and largely independent for both goals. These independent effects reflect the motivational consequences of identity goal striving predicted by symbolic self-completion theory (Wicklund & Gollwitzer, 1982). The current feeling of not belonging to one’s culture...
upon entering identity goal incompleteness evident in our data corresponds to the self-evaluative negative affective state predicted by symbolic self-completion theory. After setbacks, individuals should be motivated and energized to restore their incomplete identities. This motivation was evident in our results as the discrepancy between the current feeling of belongingness and the desire to belong. Furthermore, according to symbolic self-completion theory, these processes should be specific to the incomplete identity goals. The distinct patterns found for German and Turkish cultural identity support this notion. This suggests that for the participants German and Turkish identity were indeed separate goals and while the underlying processes ran in parallel, they both affected the same multifinal means. Thus, grounded in distinguishable motivational processes, the identity goals worked in synergy.

The findings presented here are important for current debates about the integration of immigrants into a host culture. Our experiment provides evidence for synergy between two cultural identity goals. The goals did not interfere with each other in the pursuit of a means serving both identity goals. Instead, both goals contributed to the symbolization effort. Yet at the same time, the identity goals were distinguishable in their effect on the subjective experience of belongingness. This indicates that multiple cultural identity goals can coexist in harmony in the same person, provided that the situation allows for multifinal symbolization.

**General Discussion**

The gist of the two studies presented in this article is that the interplay of simultaneously active identity goals depends on the available means. If, on the one hand, means are not multifinal, multiple identity goals are most likely conflicting and performance, resource allocation, and effort is hampered. If, on the other hand, multifinal means are at hand, goal conflict can be avoided. Resources can be effectively focused on a shared means. This interpretation is also in line with the goal systems approach (Kruglanski et al., 2018), according to which goals are interconnected to each other and to means for the goals. Although the association of multifinal means and their goals is weaker than the association of unifinal means and their goals (Kruglanski et al., 2013), multifinal means tend to be preferred when multiple goals are active (Köpetz et al., 2011). This is most likely because multifinal means maximize the potential value of an action since their value depends on all goals served by these means (Kruglanski et al., 2013).
Köpetz et al. (2011) have identified the difficulty of detecting multifinal means as a predictor for selecting them. In our present studies, the means were externally provided and we measured effort, not the choice of means. Future studies should investigate the effect of identity goal systems on the choice of self-completion symbols. A possible experiment investigating this could use a design similar to our second experiment. Identity completeness versus incompleteness would be induced in two identity goals. To measure the choice of means, instead of one means available for symbolization participants would have the option to choose a multifinal means, or a unifinal means. We predict in line with Köpetz et al. (2011) that multifinal means will be chosen over unifinal means, but only when both identity goals are incomplete. If one of the identity goals is completed, then a unifinal means for the other will be chosen.

Identity goals represent relatively high-level goals in a goal hierarchy (Carver & Scheier, 1982). In this article, we explored the interaction of multiple goals at a similar hierarchical level. The interaction is likely to be markedly different for lower level goals. According to Shah and Kruglanski (2003), the connections of goals at the same level are mainly inhibitory. This notion is compatible with our results. Two identity goals at the same level inhibited each other. Only when a shared means (which can be thought of as a lower level intention) was introduced did the goals work in accord. Although all identity goals are high-level goals in comparison to instrumental goals and behavioral intentions, we propose that the interrelation between identity goals is best described as their degree of overlap rather than a simple association in a hierarchical model. As the present studies suggest, this overlap would depend greatly on situational factors such as availability of means and associated lower level goals.

Limitations

The focus of the present experiments was on the interaction of multiple identity goals. Consequently, we manipulated goal activation within the studies. Experiment 1 shows identity goal interaction under the condition that only unifinal means are available, and Experiment 2 shows the interaction when multifinal means are available. We do not have a direct comparison between these cases within one experiment. Future studies should experimentally manipulate the availability of multifinal and unifinal means within one experiment and measure effort as a
consequence of goal incompleteness. This will be challenging because the effort in the multifinal means would have to be comparable to the effort in the unifinal means. For such an experiment, suitable identity goals with distinct but comparable means and comparable shared means would have to be identified.

A further limitation of our studies is that we only assessed the subjectively experienced incompleteness in Experiment 2. While the behavioral effects of goal activation are the main topic for the present work, subjective experience of the participants can provide further insight into the underlying processes. Indeed, we found in Experiment 2 that the participants experienced the activated identity goals as two motivationally distinct goals. We expect that a similar pattern would arise if one were to assess the experienced goal activation in a goal system like that of our first experiment (i.e., the moral and the competence identity goals are experienced as two separate goals). Future work investigating the dynamics of identity goals may adopt a method similar to our experiment to assess identity goal incompleteness (i.e., subjectively experienced identity goal activation).

**Conclusion**

The present experiments provide evidence that the multifinality of available means for symbolizations determines whether multiple identity goals support or hinder each other. Whenever we have more than one active identity goal, we are confronted with the challenge of managing these goals. Multifinal means, in particular, are suitable to navigate effectively in these situations. If no multifinal means is available, we must prioritize one identity goal over the others. This challenge of goal prioritization individuals face in such a situation is similar to the challenge of goal selection. Self-regulatory strategies might help to overcome a possible decision impasse. In particular, a strategy labeled mental contrasting (Oettingen et al., 2001; Oettingen & Stephens, 2009) could be useful. In mental contrasting, a decision maker pits a desired future against the present reality. By activating expectations about success or failure, commitment is regulated. If the goal is seen as feasible, commitment will increase. If it is seen as unachievable, commitment will decrease or the agent will disengage from the goal. Mental contrasting is a useful self-regulatory strategy for goal selection and commitment. Future studies should test the merits of mental contrasting for managing multiple identity goals.
Finally, our experiments provide inside into identity goal striving that is relevant for social policy and interventions. We investigated moral identity goals in interaction with competence identity goals. A situation where individuals face the challenge of being a moral person and demonstrating their competence at the same time is not unusual. A study by Marquardt et al. (2016) has shown that individuals with active professional identity goals are willing to endorse immoral actions and describe themselves in less moral terms if immorality is seen as instrumental for professional identity. In Marquart et al.’s study, moral identity goals were not activated. The present evidence suggests that even if being moral is currently a concern for an individual, this goal competes with other goals. Thus, for the organizational context we recommend that supervisors avoid mixing feedback concerning employees’ ethical behavior and competence.

Our second experiment dealt with cultural identity. The core finding here is that synergy between two separate cultural identity goals emerged in the use of a shared means. Future research should examine to what degree this situational factor determines the identity conflict that bicultural individuals may experience (Benet-Martínez & Haritatos, 2005). Understanding in what type of situations cultural identity goals are in conflict and when they are in harmony can contribute to successful cultural integration.
Research Paper III

Controlling Emphasis Effects in Moral Judgments

Johannes T. Doerflinger\textsuperscript{1} & Peter M. Gollwitzer\textsuperscript{12}

\textsuperscript{1}University of Konstanz, Germany \textsuperscript{2}New York University, USA
Abstract

We demonstrate in three experiments that emphasizing content in a moral dilemma by presenting it in emotionally expressive language influences moral judgments. Emphasis leads to less consequentialist judgments if the emphasis is on negative consequences of a consequentialist choice, but it leads to more consequentialist judgments if the emphasis is on the negative consequences of a deontological choice. Importantly, action phase-related mindsets can be used to control this emphasis bias. The predecisional deliberative mindset is associated with undirected attention and processing of information, while the preactional implemental mindset is associated with focused attention and processing of information. Individuals’ attention and processing in the implemental mindset is selectively directed at goal relevant content. Therefore, we expect individuals in the implemental mindset to be less susceptible to the emphasis bias than individuals in the deliberative mindset. The results of Experiments 1 & 2 suggest that effects of emphasis are more pronounced when individuals are in a deliberative mindset and they are less pronounced when individuals are in an implemental mindset. Evidence for the focus of attention and processing in an implemental mindset is provided in Experiment 2. We show in an eye-tracking task that visual attention is indeed more focused — in particular on goal-directed means — in the implemental mindset and undirected in the deliberative mindset. Thorough reflection is tested as an alternative way of controlling emphasis effects. In Experiment 3, participants were instructed to plan to reflect thoroughly or make a spontaneous moral judgment. Emphasis effects occurred irrespective of whether participants planned thorough reflection or spontaneity. These findings suggest that the occurrence of an emphasis bias does not depend on how much we think about a moral dilemma, but rather how we think about it. Focused attention and processing reduce the emphasis bias, whereas undirected processing increases it.

Keywords: Mindset theory of action phases, trolley-type dilemmas, moral judgment
Controlling Emphasis Effects in Moral Judgments

In 2005, the German government created the Aviation Security Act (Ger. “Luftsicherheitsgesetz”) as a response to the September 11 terrorist attacks. The Act would have allowed the military to shoot down commercial airplanes if they had been hijacked by terrorists to be used as weapons. The Act was controversial from the beginning and ended up at the Federal Constitutional Court, which ultimately annulled the Aviation Security Act because it was irreconcilable with the constitutional guarantee of unconditionally protected human dignity. The court argued that it was unlawful to sacrifice innocent lives even to save others and that human lives cannot be weighed in numbers (Spendel, 2006).

Beyond the legal argument, this example is a moral dilemma where a consequentialist option based on the outcome (to save many lives by killing a small number of people) is in conflict with deontological ethics of principle (killing is wrong, no matter what). Ideally, agents faced with such moral dilemmas would make their decisions based on the facts and their moral convictions. However, typically many other influences affect our moral judgments as well (Christensen, Flexas, Calabrese, Gut, & Gomila, 2014). The language used to describe a moral dilemma can nudge our reactions (Schaich Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006). If the dilemma was presented with dramatic emphasis on the horrifying death of the passengers in the airplane, the response is likely to be different from the reaction to a neutral presentation, even if the factual content was the same in both cases. Past research has associated emotional processes with deontological judgments, and cognitive control and reasoning with consequentialist judgments (Greene et al., 2004; Moore, Lee, Clark, & Conway, 2011; Shenhav & Greene, 2014).

We argue that emotionally expressive language can bias moral judgments both ways. If the circumstances surrounding the Aviation Security Act were presented with emphasis on the consequences of the potential victims of a terrorist attack, moral judgments of the case are likely to be skewed in the direction of consequentialist responses. We refer to this as an emphasis bias, because the linguistic presentation emphasizing one side or the other in a scenario does not alter the core content or structure of a dilemma. Our aim in the current article is to demonstrate the emphasis bias and show how the bias can be controlled by changing one’s mode of processing.

Focusing attention and processing on essential content should shield individuals from distractions and biases. In particular, we turn to cognitive
orientations that are part of the mindsets activated during goal pursuit. According to 
*mindset theory of action phases* (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016),
four successive tasks have to be fulfilled in four action phases in the pursuit of a goal: 
Selecting a goal (*predecisional phase*), choosing a strategy how to implement the 
chosen goal (*preactional phase*), acting on the strategy (*actional phase*), and 
evaluating the goal striving efforts (*postactional phase*). Each phase comes with a 
cognitive tuning (i.e., *mindset*) to meet the phase-specific demands. These mindsets 
 affect attention and information processing.

In the current work, we focus on the first two phases: the predecisional and 
the preactional phase. In the predecisional phase, the agent is usually in a *deliberative 
mindset* which facilitates broad, open-minded processing of information, even if it is 
incidental or task-irrelevant. In contrast, the prevalent mindset in the preactional 
phase is the *implemental mindset* which facilitates focused, closed-minded processing 
and shields the agent from distractions by irrelevant information (Fujita et al., 2007). 
We show that agents in a deliberative mindset are more prone to the emphasis bias 
than agents in an implemental mindset. This is likely due to the fact that in the 
implemental mindset attention is more focused and information processing is more 
selective than in the deliberative mindset. In support of this line of reasoning, eye-
tracking data revealed that participants in an implemental mindset focused their visual 
attention on action-oriented means, whereas visual attention in the deliberative 
mindset was evenly spread.

In the following sections, we will first summarize the relevant literature on 
moral judgment and lay out our arguments why emphasis should be able to influence 
moral judgments towards deontological as well as consequentialist responses. Then 
we will introduce mindsets as a self-regulatory mechanism used in our experiments to 
change the mode of information processing (Gollwitzer, 1990, 2012; Gollwitzer & 
Keller, 2016) and discuss how we expect this mechanism to moderate the influence of 
emphasis on moral judgments.
Moral Judgment and Moral Dilemmas

The case of the German Aviation Security Act bears many similarities to a prototypical thought experiment that has gained wide attention in moral psychology: the trolley-problem (Foot, 1983; Thomson, 1985). The standard version of the trolley problem is a hypothetical moral dilemma in which a runaway trolley is approaching a group of five people on the track. If the protagonist does not interfere, the trolley will hit and kill the group of five people. The only way to avoid the deaths of these people is to sacrifice the life of a single person. One can construct similar stories with the same structure and we will refer to this class of dilemmas as trolley-type dilemmas.

Trolley-type dilemmas are defined as a story in which the protagonist must choose between two options. One, a consequentialist option, typically sacrificing the life of one person or a small group in order to save a larger group; and two, a deontological option, typically refraining from killing a person or small group of people. Since the publication of an influential article on the neural basis of moral judgment (Greene et al., 2001), trolley-type dilemmas have become a widely used tool for researchers interested in moral cognition.

Emotional Deontology and Reasoned Consequentialism?

Moral judgments of trolley-type dilemmas have been described as the product of cognitive, reflective processes driving a consequentialist assessment and emotional, intuitive processes associated with deontological preferences (Cushman et al., 2010; Greene & Haidt, 2002; Greene et al., 2001; Shenhav & Greene, 2014). We will summarize research in support of this framework, but also note that in some cases the lines between emotion and deontological judgments, on the one hand, and reasoning and consequentialist judgments, on the other hand, are blurred. Verbal emphasis, the central phenomenon of the present paper, is one of these cases. We propose that emotion-driven processes can lead to less consequentialist responses if an anti-consequentialist emphasis is added to a dilemma, and to more consequentialist responses if an anti-deontological emphasis is added.

There is converging evidence for the notion that emotional responses support the formation of deontological judgments. Individuals presented with personal moral dilemmas, a class of dilemmas that evoke strong negative emotions, make consistently more deontological judgments than individuals presented with impersonal, less emotional dilemmas (Greene, 2009a; Greene et al., 2001; Hauser,
Cushman, Young, Jin, & Mikhail, 2007; McDonald, Defever, & Navarrete, 2017; Moore et al., 2011). A plausible explanation for this effect is that humans have an aversive emotional response to inflicting harm. A strong stress response correlates with less consequentialist responses in trolley-type dilemmas, and thinking about inflicting harm or simulating harmful actions increases the stress response (Cushman, Gray, Gaffey, & Mendes, 2012). Downregulating the emotional response to moral dilemmas (Lee & Gino, 2015) or externally inducing positive affect with a different task to counteract the emotional alarm response in moral judgments increases the consequentialist judgments in personal moral dilemmas (Valdesolo & DeSteno, 2006). This effect may not generalize to all positive emotions. Specifically, humor increases consequentialist preferences while elevation, another positive emotion, decreases it (Strohminger, Lewis, & Meyer, 2011).

As the studies above show, affect contributes to moral judgments. Some researchers even argue that moral judgments are primarily emotional assessments and reasoning serves to come up with post-hoc rationalizations for these assessments (Haidt, 2001). However, experimentally inducing reflective reasoning has been shown to increase consequentialist judgments (Paxton, Ungar, & Greene, 2012), whereas putting participants under cognitive load, which effectively interferes with effortful reasoning, has been shown to slow down consequentialist but not deontological judgments (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008). Furthermore, higher working memory capacity — which is necessary for reasoning — is positively related to consequentialist judgments (Moore, Clark, & Kane, 2008). Attributing moral judgments solely to affect misses part of the picture, just as earlier theories which define morality as purely cognitive (e.g., Kohlberg & Candee, 1984) are incomplete.

According to neuroimaging studies (Hutcherson et al., 2015; Shenhav & Greene, 2014), the processes are integrated such that when a person contemplates a trolley-type dilemma two types of assessments are made. The emotional system arrives at an assessment about which option feels worse. This assessment is rapid and automatic. The reflective system produces an assessment about which option will yield better consequences. This assessment requires cognitive resources, effort, and accurate processing of available information. An overall moral judgment is produced by integrating the emotional and the consequentialist assessments.
It is important to note that there is evidence for controlled reasoning in deontological judgments and emotional processes involved in consequentialist judgments. Some moral principles supporting deontological judgments (e.g., harmful actions are worse than harmful omission) are consciously accessible for moral reasoning (Cushman, Young, & Hauser, 2006). Furthermore, individuals who endorse a deontological response are less confident about their choice and have slower reaction times when they respond to moral dilemmas that involve a conflict between deontological and consequentialist options than when they respond to dilemmas without a conflict (i.e., the non-deontological option does not yield a better outcome). These results indicate that individuals who endorse deontological responses do take the consequentialist option into account. Although emotional reactions to moral dilemmas correlate with deontological judgments, this is not a clear one-to-one mapping. Szekely and Miu (2015, Study 1) assessed the emotional experience of participants who worked on trolley-type dilemmas. In support of the “emotional deontology” hypothesis, they found that participants reported experiencing affective responses more frequently after a deontological choice than after a consequentialist choice. However, when emotional experiences were categorized into discreet classes of emotions, a distinct emotional response could be identified for both consequentialist and deontological choices. After deontological responses a higher degree of fear and disgust was reported, and after consequentialist responses the participants indicated more often that they experienced regret.

Taking these studies into account, it is reasonable to assume that both reasoning and emotions influence deontological as well as consequentialist preferences. Moretto, Làdavas, Mattioli, and Di Pellegrino (2010) offered an argument that explains why emotional reactions tend to promote a preference for deontological options. In the dilemmas most commonly used in moral psychological research, consequentialist options are tagged with negative emotions which individuals tend to avoid. We propose that such an emotional tag could also be attached to deontological options. This, in turn, can be expected to increase relative preferences for consequentialist options. Importantly, this argument rests on a core feature of emotions: emotions are directed at a target (Colman, 2008). Adding a verbal emphasis of consequences that discourages endorsement of deontological or consequentialist judgments respectively is one method of marking the options with an emotional tag.
The Emphasis Bias in Moral Judgments

Schaich Borg et al. (2006) argued that dramatic language could potentially bias moral judgments. In a study in which the neural correlates of moral judgments were assessed, they systematically varied depending on dilemmas being presented in either plain or colorful language. In this study the effects of language on moral judgments were small. However, they did find that the neural response to dilemmas depended partly on the linguistic style of presentation. Further evidence for the influence of language in moral judgments is provided by framing studies. The framing effect (Tversky & Kahneman, 1981) refers to the fact that people tend to judge factually identical situations differently, solely depending on the way the situations are presented. If harmful actions in dilemmas are described as “killing”, rather than “saving”, then the endorsement of these actions is reduced (Petrinovich & O'Neill, 1996). In line with these studies, Nichols and Knobe (2007) have shown that subjects attribute more moral responsibility in third-party moral judgments if scenarios are presented in an emotionally salient style as compared to abstract formulations. Reading scenarios in a foreign language reduces the emotional response. Consequently, individuals endorse consequentialist options in foreign languages more than in their native language (Costa et al., 2014). In sum, language affects moral judgments of trolley-type dilemmas and this effect seems to stem at least in part from the saliency of emotions.

Although there is suggestive evidence for the impact of emotionally expressive language on moral judgments, some studies have only found modest effects (Schaich Borg et al., 2006). This can in part be attributed to the fact that the linguistic style is a very subtle manipulation of emotional framing. However, an additional puzzle piece may lie in the nature of emotions. Emotions are, in contrast to moods, by definition evaluative and directed at an object (Colman, 2008). Personal moral dilemmas are told in a way that makes respondents feel bad about the harmful consequentialist options (Shenhav & Greene, 2014). The emotional emphasis in these dilemmas is directed at the consequentialist option. It is therefore possible that the target of negative affect in personal dilemmas (and moral dilemmas more generally) influences moral judgments to favor one option over the other. Such effects would be in addition to the general affinity of intuitive, emotional processes for deontological options and the affinity of reflective reasoning for consequentialist options.
We argue thus that there is no strong link between emotional expressive language per se and non-consequentialist preferences. Rather, we propose that how emotional expressive language influences moral judgments depends on what the emotions are directed at. If, on the one hand, the harm inflicted in consequentialist options is emphasized, this will elicit negative affect directed at the consequentialist option and thus reduce consequentialist preferences. If, on the other hand, the harm that results from inaction in the non-consequentialist option is emphasized, then negative emotion directed at the non-consequentialist option will arise. In this case, emotionally expressive language can, in fact, be expected to increase consequentialist preferences. We thus hypothesize that emotionally expressive language can both increase and decrease the preference for consequentialist judgments depending on the target of the emotion. We refer to this effect as emphasis bias.

**Mindset Theory of Action Phases**

Human judgment and behavior rarely occur in a motivational vacuum and motivated behavior can be understood in the framework of goal pursuit (Moskowitz & Grant, 2009). According to Mindset Theory of Action Phases (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016), individuals process information differently, depending on their current stage of goal pursuit. Such differences in information processing are likely to influence how we react to factual information and emotional framing in moral dilemmas. In Mindset Theory of Action Phases (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016; Heckhausen & Gollwitzer, 1987), goal pursuit is subdivided into four consecutive phases. In each phase, the agent deals with a specific challenge of goal pursuit and a phase-specific cognitive orientation (i.e., mindset) that promotes successfully overcoming the challenges of the current task. In the present experiments, we focus on the mindsets that are active in the first two phases, the deliberative mindset, and the implemental mindset. The deliberative mindset is active in the predecisional phase, in which goal selection takes place. After a goal decision has been made, people enter the preactional phase. In this phase, the task at hand is to implement appropriate actions to reach the chosen goal and to initiate action. This is aided by an implemental mindset.

The mindset in each phase is helpful for overcoming the specific challenge of that phase. One can distinguish mindsets from task sets. Task sets are a form of cognitive tuning that is specific and limited to a given task. For example, while
writing a poet could be tuned towards forming sentences in rhymes. In contrast, mindsets represent a general cognitive orientation (e.g., focusing on task-relevant information) that can carry over from a task that elicited them (e.g., planning out how to move to a new city) to a new, unrelated task (e.g., organizing a birthday party). The carry-over effect of mindsets makes them an attractive tool to investigate the influence of action phase-related cognitive procedures on various tasks, judgments, and behaviors (Gollwitzer & Keller, 2016). The breadth of information processing differs between the deliberative and implemental mindsets. Information processing and attention are broad in the deliberative mindset and narrow in the implemental mindset (Fujita et al., 2007). This can be expected to affect the susceptibility to the emphasis bias.

**Open-Mindedness and Focus – Deliberative vs. Implemental Mindsets**

In a deliberative mindset, the agent is cognitively tuned towards impartially processing of feasibility and desirability concerns. In addition, deliberative mindsets are associated with broad, open-minded processing of information. Because the agent has not yet committed to a goal, making good decisions would benefit from considering all available information (Gollwitzer & Keller, 2016). While open-mindedness aids decision making in goal selection, it can be disadvantageous when the agent is already committed to a goal and seeks ways to implement it. When a goal is set, ignoring goal-irrelevant stimuli and information supports successful goal attainment. Information processing in the preactional phase, when the agent is in an implemental mindset, is thus more selective. Only goal-relevant information should be processed, and the agent can be expected to focus on the means for goal attainment (Gollwitzer, 1990). These mindset effects have been demonstrated to affect information processing, memory, and attention.

In a series of studies, Fujita et al. (2007) first induced deliberative vs. implemental mindsets. Afterward, the participants worked on a mental concentration task. During some trials of the task, task-irrelevant, incidental words were displayed. The authors assessed how accurately the participants remembered the irrelevant stimuli in a recognition memory task. Participants who were in a deliberative mindset had significantly better recognition rates for the irrelevant information than participants in an implemental mindset. Additionally, response latencies in the recognition task were faster for participants in the deliberative mindset. The studies
show that openness to incidental information is higher in the deliberative than in the implemental mindset. Büttner et al. (2014) argued that the results of Fujita et al. (2007) have two potential sources. It is possible that the deliberative mindset, relative to the implemental mindset, increases the breadth of attention. Alternatively, memory processes may be enhanced in the deliberative mindset. To demonstrate that mindsets affect attention, Büttner and his colleagues conducted three experiments in which they measured the breadth of visual attention after a mindset induction. They found that participants in a deliberative mindset perceived lines in an optical illusion task as longer, compared to participants in an implemental mindset (Studies 1 & 2) — an effect that indicates a wider breadth of attention (Predebon, 2004). Furthermore, the authors demonstrated in an eye-tracking paradigm that participants in a deliberative mindset evenly explored the whole scene in pictures whereas participants in an implemental mindset spent more time looking at foreground objects (Study 3). Taken together, these studies provide evidence for the open-minded and broad distribution of attention in the deliberative mindset and a focused, narrow distribution of attention in the implemental mindset. We hypothesize that this feature of the two mindsets explored in the present studies moderates the emphasis bias. Processing in the deliberative mindset is relatively undirected. Consequently, we expect individuals in a deliberative mindset to be more prone to the emphasis bias. In contrast, processing in the implemental mindset is more focused on goal-directed means. This focus should be a protective factor against the emphasis bias.

**Present Research**

The aim of the present experiments is to test whether emotionally expressive language can bias moral judgments in a directional way. Former research has established a link between emotional responses and deontological judgments (Cushman et al., 2010; Shenhav & Greene, 2014). We concur that consequentialist judgments likely depend generally more on a cognitive response and capacities of reasoning, while deontological judgments are more likely to be a product of intuitive and affective processes. However, we hypothesize that emotional framing via expressive language can go both ways. Emphasizing harm that would occur if the consequentialist option is chosen directs negative emotions at that option and should reduce consequentialist judgments. However, if a dilemma is presented that is similar in content, but the harm that would occur if the deontological option is chosen is
emphasized, moral judgments are likely biased towards consequentialism. We demonstrate this emphasis bias in three experiments.

In addition to providing evidence for the emphasis bias in trolley-type dilemmas, we test how the emphasis effect can be controlled by regulating one’s mode of thought. Two questions are addressed with this investigation. First, how does the emphasis bias depend on active mindsets that affect the breadth of attentional focus and processing? Second, does thinking more thoroughly about a dilemma help to resist emphasis effects? While the former is about qualitative changes (undirected vs. focused attention and processing), the latter deals with a quantitative variation (increasing time and effort).

Mindsets alter the way information is processed and attention is directed (Büttner et al., 2014; Fujita et al., 2007). In particular, having a deliberative mindset results in open minded-processing of all available information and undirected spreading of attention. Having an implemental mindset leads to closed-minded, focused processing of only goal-relevant information and an attentional focus on goal-directed means. We expect these states to moderate the emphasis bias. Because individuals in the deliberative mindset are open to information, even if it is task-irrelevant, they are more likely to be susceptible to the emphasis bias. In contrast, individuals’ focus in the implemental mindset should decrease the emphasis bias. (Fujita et al., 2007) have suggested that the selective information processing in the implemental mindset contributes to a phenomenon referred to as goal shielding (Shah et al., 2002). Goal shielding is the process by which active goals are protected from competing intentions. Competing intentions are inhibited to reduce interference with active goals. If individuals in the implemental mindset perceive the emphasis information as disrupting their goal of making sound moral judgments, then they can be expected to focus on the factual content of the dilemmas instead.

In addition, we explore whether regulating the time and effort invested in a moral judgment qualifies as an alternative strategy to control emphasis effects. If the emphasis bias relied solely on intuitive processes, increasing the time to make a moral judgment could reduce the bias. Past research has shown that planning to adopt a reflective (i.e., slow) mode of processing can improve decision quality in situations where intuitive biases lead otherwise to suboptimal spontaneous preferences (Bieleke, Gollwitzer, Oettingen, & Fischbacher, 2016; Doerflinger et al., 2017). Consequently, one might expect a weaker emphasis effect after individuals plan to reflect thoroughly
on their moral judgments and a stronger emphasis effect after planning to make spontaneous judgments. Our data do not support this hypothesis. Instead, we observed the emphasis bias both for participants who planned to thoroughly reflect and participants who planned to make spontaneous judgments.

**Expressive Language and the Emphasis Bias**

A set of stories was created as stimuli for the present experiments that met the criteria of a trolley-type dilemma (i.e., if the agent does not interfere, then a group of people will die; the only way to avoid this is to sacrifice a single person or a smaller group of people). The stories are based on the materials used by Greene et al. (2004) and Moore et al. (2008) as well as additional dilemmas created for the purpose of this study. For each dilemma, a neutral version, a version with emphasis on the harmful consequences in case of the consequentialist option (*anti-consequentialist version, AC*), and a version with emphasis on the harmful consequences in case of the deontological option (*anti-deontological version, AD*) were created. All factual information about the dilemmas was provided in the neutral version. For AC and AD versions a short paragraph was added that provided emotional but no factual information about the dilemma. For example, for the standard trolley dilemma, this results in the following variants.

[All versions] You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workers. On the tracks extending to the right is a single railway worker. If you do nothing, the trolley will proceed to the left, causing the deaths of the five workers.

*[AD version]* In that case these five workers would see the approaching trolley unable to flee from their dire situation before being hit by the trolley. They would die an agonizing death.

[All versions] The only way to avoid the deaths of these workers is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single worker.

*[AC version]* In that case the single worker would see the approaching trolley unable to flee from his dire situation before being hit by the trolley. He would die an agonizing death.
Eight moral dilemmas with this structure were created for Experiments 1 & 3. One dilemma (organ transplant) elicited extreme responses with the median endorsement of consequentialism at the minimum of the scale in both experiments. Excluding this item from the analysis does not alter our observed pattern of results and significance of the effects. Because we had no prior hypothesis regarding the item and we analyzed our data in mixed models with random intercepts for each item, we decided to keep the item in our dataset. For Experiment 2 the dilemmas were refined and additional items were added. In Experiment 2, pictures representing the dilemmas were used in addition to written text.

**Experiment 1: Controlling the Emphasis Bias with Mindsets**

In the first experiment, we explored the moderating effect of mindsets (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016) on the emphasis bias. According to mindset theory of action phases, the phases of an action are associated with distinct cognitive orientations. These mindsets can carry over from one task to the next and can influence information processing on the subsequent task. Of particular interest to our experiments were the deliberative mindset, which is associated with broad, open-minded information processing, and the implemental mindset, which is associated with more focused, closed-minded information processing. We predict that an implemental mindset reduces the emphasis bias and a deliberative mindset increases it. To test this hypothesis, an online experiment was conducted in which participants first worked on a mindset induction task and then provided moral judgments of trolley-type dilemmas.

**Method**

**Participants and design.** The experiment had a 3 (emphasis: AC vs. no emphasis vs. AD) × 3 (mindset: deliberative vs. control vs. implemental) between subjects design and was conducted as an online experiment on the platform prolific.ac (Palan & Schitter, 2018). Assuming that mindset effects are small to medium-sized and responses to different trolley type dilemmas are moderately correlated, a power analysis with G*Power was conducted (Faul et al., 2007). To achieve a statistical power of $1-\beta = .95$, 288 participants would be needed. We recruited 306 participants on prolific.ac. Only participants who did not take part in earlier moral dilemma studies from our lab with an acceptance rate higher than 90% were admissible for the
experiment. This second restriction was put in place to ensure that the participants took the experiment seriously. The participants were paid 2 £ for taking part in the experiment. One participant entered seemingly random letters in the mindset induction task, another participant failed an attention check, and a third participant indicated in a debriefing question that they did not take the experiment seriously. These three participants were excluded from the data analysis. Consequently, our sample size is 303 (125 female, 171 male, 7 other/did not indicate) with a mean age of 26 ($SD = 9.4$, age range: 16 – 52).

**Procedure.** Mindsets were evoked with a standard procedure often used in mindset studies (Gollwitzer & Keller, 2016). To induce a deliberative mindset, participants were instructed to think of a personal problem they had not yet decided to act upon. They were asked to think of a complex problem. Examples for such problems were choosing your field of study, deciding whether to move, or deciding whether to get in touch with someone new. After choosing their personal problem, participants were asked to think of three immediate positive consequences and three immediate negative consequences of choosing to take action. In addition to naming the consequences, participants were asked to estimate their probability in percent. This was repeated for long-term consequences. After deliberating about the consequences of taking action, participants were asked to repeat this process by thinking about the consequences of not taking action. To induce an implemental mindset, participants were asked to think about a personal project on which they had already decided to act upon, but had not started yet. This should be a complex problem. Examples of such problems were changing your field of study, moving out of an apartment or getting in touch with someone new. After choosing a project, participants were asked to name five steps that were necessary to complete the project. In addition, they were asked to indicate when, where and how they would act on each of these steps. In the control condition participants were asked to indicate when they woke up, when they had breakfast, when they had lunch, and when they had dinner for the last seven days.

**Manipulation check.** Participants in the two mindset conditions were asked to mark on a visual analog scale representing a timeline where they saw themselves in relation to the decision they worked on during the mindset task. The scale had “making a decision” at its center. Participants in the deliberative mindset condition should place their mark on the left side of the decision (before making a decision) and
participants in the implemental mindset condition should place their mark on the right side of the decision (after making a decision). The scale was coded ranging from 1 on the left side to 100 on the right side.

**Moral dilemma task.** After the manipulation check, the participants were presented with eight moral dilemmas in random order. The participants only saw the version of the moral dilemmas from their emphasis condition. Each moral dilemma was presented on a separate page and on the bottom of the page was a 5-point rating scale on which the participants could indicate the moral wrongness of the consequentialist option (e.g., “It is morally wrong to hit the switch.”) ranging from “I agree” to “I disagree”. Next, the moral dilemmas were again presented in randomized order and the participants were asked to indicate how emotional and how difficult each of the dilemmas was.

At the end of the experiment, demographic variables were assessed. Among the demographic questions was an attention check. This question was a multiple-choice question in a drop-down format, which could be answered: “left”, “right”, “bottom”, and “top”. The participants were instructed on the same page to select the option “bottom”. Finally, the participants were asked to indicate what they thought the purpose of the present experiment to be.

**Results**

Mixed models in all three experiments were computed with the lme4 package for R (Bates et al., 2015). For mixed linear models, test statistics were approximated using the Satterthwaite approximations for degrees of freedom in the lmerTest package for R (Kuznetsova, Brockhoff, & Christensen, 2017).

**Manipulation checks.** Participants in the deliberative mindset condition indicated on average that they saw themselves as predecisional on the timeline (\(M = 46.36, SD = 22.77\)) and participants in the implemental mindset condition indicated that they saw themselves as postdecisional on the timeline (\(M = 58.06, SD = 22.27\)). An independent samples \(t\)-test revealed that this difference was significant, \(t(201.34) = 3.72, p < .001\). A mixed linear model with the emotionality ratings for each dilemma as the criterion, the emphasis condition as predictor, and random intercepts for participants and dilemmas revealed that the AC dilemmas (\(M = 3.69, SD = 1.25\), \(t(302.53) = 2.90, p = .004\)) and the AD dilemmas (\(M = 3.55, SD = 1.31\), \(t(302.53) = 1.99, p = .048\)) were rated as more emotional than the no emphasis dilemmas.
(M = 3.27, SD = 1.39). Emotionality ratings for the AD and AC dilemmas did not differ significantly, t(302.53) = 1.05, p = .295. Difficulty rating did not differ between conditions, ts < 1.32, ps > .18.

**Moral judgments.** The experimental condition variables and their interaction were entered as predictors in a linear mixed model with the moral judgments for each dilemma as the criterion and random intercepts for participants and dilemmas. The main effect of mindset was not significant, t(302.13) = -0.65, p = .519. The emphasis conditions had a significant main effect on moral judgments, t(302.13) = 2.21, p = .028. This main effect was qualified by a marginally significant interaction of mindset and emphasis, t(302.13) = 1.87, p = .063. For more information on the model see Table 1. The mean moral dilemma ratings are visualized in Figure 10.

**Table 1.** Mixed linear model estimating moral judgments in Experiment 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.04</td>
<td>0.19</td>
<td>8.85</td>
<td>16.28</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mindset&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.04</td>
<td>0.06</td>
<td>302.13</td>
<td>-0.65</td>
<td>.519</td>
</tr>
<tr>
<td>Emphasis&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.12</td>
<td>0.06</td>
<td>302.13</td>
<td>2.21</td>
<td>.028</td>
</tr>
<tr>
<td>Mindset&lt;sup&gt;a&lt;/sup&gt; × Emphasis&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.12</td>
<td>0.06</td>
<td>302.13</td>
<td>1.87</td>
<td>.063</td>
</tr>
</tbody>
</table>

*Note.*  
<sup>a</sup> implemental = -1, control = 0, deliberative = 1,  
<sup>b</sup> AC = -1, no emphasis = 0, AD = 1
Figure 10. Mean endorsement of the consequentialist options as a function of emphasis and mindset condition in Experiment 1, error bars represent 95% confidence interval.

Discussion

In Experiment 1 we demonstrated the emphasis bias effect. Participants endorsed the consequentialist option more in dilemmas where an anti-deontological emphasis was added and endorsed the deontological option more in dilemmas where an anti-consequentialist emphasis was added. This finding underscores that the use of expressive language can influence moral judgments. It is an important extension of previous research on this topic. Past studies have shown that framing a dilemma in terms of killing versus saving people influences moral judgments (Petrinovich & O’Neill, 1996); further, more moral responsibility is attributed to the protagonist of moral dilemmas in third-party moral judgments if the dilemmas are presented in colorful, emotional language compared to plain language. Our study goes beyond previous work by using verbal emphasis to target the specific alternatives in trolley-type dilemmas. This is done by attaching an emotional tag to either the deontological or the consequentialist option. Our present experiment suggests that the effects of expressive language go beyond a simple one to one mapping of emotion and deontology. Instead, how emphasis affects our moral judgments depends on what
content is emphasized. This way, emotion-driven processes can increase consequentialist judgments. Because this effect does not rely on any factual content but solely on emphasis, this can be interpreted as a bias.

It is important to note, however, that the emphasis effects were moderated by the induced mindsets. As can be seen in Figure 10, the emphasis conditions do not differ for participants in an implemental mindset, whereas the clearest effect can be observed for participants in a deliberative mindset. Thus, the occurrence of an emphasis bias was controlled by the prevalent mindset. We hypothesize that the moderation is due to an attentional focus and more selective processing in the implemental mindset and more open-mindedness, even to irrelevant information such as the emphasis in the deliberative mindset.

An important caveat is that the interaction effect was only marginally significant. Thus, replication is needed. If the emphasis by mindset moderation can be reliably replicated, the cognitive mechanisms underlying this effect need to be explored. We propose that emphasis effects are eliminated in the implemental mindset because attention and processing in the implemental mindset are focused on goal-relevant content. This feature should shield agents from distractions and irrelevant information. In contrast, undirected attention and open-minded processing in the deliberative mindset likely increase the influence of emphasis on moral judgment. Experiment 2 was designed to replicate the moderating effect of mindset on the emphasis bias and to assess attentional focus in the two mindsets investigated in the present work.

**Experiment 2: Replication and an Exploration of Visual Attention in Different Mindsets**

So far, we have demonstrated the emphasis bias. Additionally, there is tentative evidence for a moderating role of mindsets on this effect. Here we aim to conceptually replicate this interaction and to further explore the role of mindsets on attention. One limitation of Experiment 1 is that the mindset manipulation was a relatively artificial task that may lack external validity. To address this potential limitation, in Experiment 2 a new method to induce mindsets was used. Participants were given short monologues from Shakespeare’s play Hamlet, in which the protagonist was either deliberating about a choice he has not yet made, or laying out a
plan to achieve a goal he was committed to. The participants were instructed to identify themselves with the protagonist and analyze his thought processes.

A major addition in Experiment 2 is that we measure visual attention while participants look at schematic images depicting the scenario in the trolley-type dilemmas. Deliberative and implemental mindsets have been shown to affect attention differently (Büttner et al., 2014; Fujita et al., 2007). In a deliberative mindset, attention is relatively undirected, whereas an implemental mindset is focused. This may contribute to the open-mindedness vs. closed-mindedness feature of mindsets and the interaction with emphasis. In the present study, we go one step further. We propose that attention in an implemental mindset (relative to a deliberative mindset) is not just more focused in general, but rather more focused specifically on goal-directed means. In Study 3 of Büttner et al. (2014), mindset effects on attentional breadth were measured in an eye-tracking task. The authors first established a deliberative vs. implemental mindset and then instructed the participants to evaluate pictures. The pictures contained a focal foreground object and a complex background. The focal object and the background were defined as areas of interest (AOIs). Gaze time for participants in the deliberative mindset was longer on the background objects compared to participants in the implemental mindset. In contrast, participants in the implemental mindset spent more time looking at the foreground objects. This study points to overall differences in focus and attention between the mindsets investigated. Because the scenes used as stimuli depicted only static objects without narrative content or significance, the study could not show whether attention in the implemental mindset is focused on specific content such as goal-directed means and action-relevant content.

The systematic structure of trolley-type dilemmas is a suitable environment to test whether attention in the implemental mindset is indeed focused on goal-directed means. Trolley-type dilemmas can be constructed such that each dilemma has a goal-directed means that can be visualized (e.g., a button or a lever). The means of the potential action represent content that is important for the implementation (action vs. inaction) of the response. Consequently, we expect individuals in an implemental mindset to focus more on these elements, compared to individuals in a deliberative mindset when they look at a visual depiction of a trolley-type dilemma they are currently judging. This should result in more fixations on means in the implemental mindset condition. In sum, Experiment 2 has two aims. First, it serves as a conceptual
replication of the emphasis by mindset interaction observed in Experiment 1. Second, we aim to show that attention in an implemental mindset is more focused on content that is related to action implementation as compared to attention in a deliberative mindset.

Method

Participants and design. The experiment had a $2 \times 2$ (emphasis: AC vs. AD) × 2 (mindset: deliberative vs. implemental) between-subjects design and was conducted as a laboratory experiment. Ten dilemmas were created for repeated measures of moral judgment. We aimed at a power of $1 - \beta = .80$. A power analysis with G*Power (Faul et al., 2007) resulted in a target sample size of 100 participants. We recruited 108 German speaking participants (93 female, mean age 22, $SD = 4.08$, age range: 16 – 41) at the University of Konstanz. The participants received 7 € or course credit as compensation for taking part in the experiment.

Procedure. The experiment was advertised as an experiment about literary interpretation and moral judgment. In each session up to four participants took part. Upon entering the laboratory, the participants were placed in individual cubicles and randomly assigned to the experimental conditions. The experimenter was blind to conditions and all tasks and instructions were computerized.

Mindset manipulation. All participants were told that the first task was to interpret a monologue from the play Hamlet. They received a short description of the plot of the play, the scene in which their assigned monologue featured, and a summary of the content of the monologue. Participants in the deliberative mindset were asked to work on a monologue from the first scene of the third act of the play (“To be or not to be”). In this monologue, Hamlet ponders the question of whether to take action against the villain of the play, to take his own life, or not to act at all. A central characteristic of the monologue is that Hamlet is indecisive and has not made up his mind — features of the predecisional phase associated with a deliberative mindset. The monologue ends before he reaches a decision. Participants in the implemental mindset condition received a soliloquy from act 2, scene 2 (“The play’s the thing”) in which Hamlet devises a plan to determine the guilt or innocence of his uncle whom he suspects to be the murderer of his father. In the soliloquy, he lays out the steps of the plan and it ends with his strong determination to act, which is characteristic of an implemental mindset. The participants in both mindset conditions
were asked to analyze the texts from Hamlet's perspective and to imagine what he would be thinking and feeling. Guiding questions based on standard mindset manipulations were provided to the participants. After completing the literature task, the participants were asked to indicate on a visual analog scale — representing a timeline with a decision at its center — where they thought Hamlet was in relation to making a decision. This is parallel to the manipulation check in the standard mindset task. Because the literature task is likely to be rather demanding, we also assessed how difficult the participants thought it was to take Hamlet’s perspective (“It was easy for me to take Hamlet’s perspective.”, German: “Es fiel mir leicht, mich in Hamlet hineinzuversetzen.”) on a six-point rating scale ranging from “I agree” to “I disagree”. It is possible that the mindset by emphasis interaction is weaker for participants who had difficulties with the task.

**Materials and technical setup.** After the mindset task, the participants were asked to rate the moral permissibility of ten trolley-type dilemmas. The stories were partly based on materials used in Experiment 1, but additional dilemmas were added to increase the number of trials. The dilemmas in Experiment 2 all had the properties of trolley-type dilemmas and were created with two emphasis variants. An additional requirement was that the decision situation in the dilemmas could be visualized in a schematic picture. In the pictures, the means for the consequentialist action (e.g., a lever), the people who would be harmed in case the consequentialist option was chosen, and the people who would be harmed in case the deontological option was chosen were depicted. Additional background (e.g., rails) was added to create a coherent scene. Ten dilemmas and corresponding pictures were created that met these requirements. For an example of a picture depicting the standard trolley case, see Figure 11.

The moral judgment task was implemented in PsychoPy (Peirce, 2007). The task made use of eye-tracking to assess gaze data during picture presentation and participants controlled the task with their eye movements. Stimuli were presented on a 19-inch monitor with a screen resolution of 1280 × 1024 pixels. Eye-tracking data were collected using a Gazepoint GP3 eye tracker with a sampling rate of 60 Hz. The eye tracker was placed at the bottom of the screen and participants sat at a distance of approx. 70 cm from the screen. The eye tracker was approx. 50 cm below the participants’ eye-level. To minimize head movements, chin rests were provided. The
eye tracker was connected to the experiment software with the iohub event monitoring framework for PsychoPy.

![Diagram](image1.png)

**Figure 11.** Example of a picture used for the moral dilemma task in Experiment 2. Trees were added to balance the visual saliency of the single individual on the right side and the group on the left.

**Moral judgment task.** At the beginning of the task, the participants received headphones and the chinrests were adjusted. Headphones were needed later on because the experiment included an auditory signal. The eye-trackers were calibrated with the Gazepoint Control software. Next, the moral judgment task was explained to the participants and the participants were familiarized with the task procedure. They were asked to take the moral dilemmas seriously even if they might seem unrealistic. The experiment had 10 trials. The moral dilemma texts were presented with an emphasis manipulation that was identical to the manipulation in Experiment 1. The participants were instructed to read the moral dilemmas. At the end of each dilemma the question “Is it morally wrong to [perform the consequentialist action]?” was asked. Next, a fixation cross was shown in the lower section of the screen for 500 ms. Then, a picture representing the moral dilemma was shown. The participants were instructed to look at the picture for at least 15 seconds. A sound notified the participants when this time had elapsed. After 15 seconds, the participants could indicate their moral judgments by looking at the words “Yes” or “No” in the lower corners of the screen. Figure 12 depicts the sequence in a trial.
Figure 12. Sequence of screens in a trial in Experiment 2.

The position of responses (right corner vs. left corner) and the orientation of the pictures (original vs. mirror image) were counterbalanced between participants. The eye-trackers were recalibrated after five trials. Moral judgment responses and gaze data during the picture presentation served as dependent variables. Each picture showed a means to perform an action, a small group of people (or a single person), and a larger group of people. AOIs were defined with an approx. 30-pixel border around these objects. After completing the moral judgment task, the participants provided demographic information, were thanked, paid, and thoroughly debriefed.

Results

Manipulation check and perceived task difficulty. The participants in the deliberative mindset condition rated the protagonist in the literature task as more predecisional ($M = 40.63$, $SD = 13.53$), and the participants in the implemental mindset rated him as more postdecisional ($M = 59.31$, $SD = 22.36$). This difference was highly significant, $t(89.14) = 5.27$, $p < .001$. Perceived difficulty of the literature task did not differ significantly between the mindset conditions, $t(107) = -0.54$, $p = .588$. However, a linear regression revealed a significant interaction effect of perceived difficulty and mindset condition on the manipulation check variable, $F(1,103) = 4.04$, $p = .047$, indicating that participants who had difficulties with the task tended to rate the monologues in the opposite direction. Thus, we decided to control for perceived task difficulty in our main analysis.

Moral judgments. To test the combined effect of mindsets and emphasis on moral judgments, a mixed linear logit model with the moral judgment decisions as the criterion, emphasis, mindset, and perceived difficulty in the mindset task, as well as the interaction terms as predictors and random intercepts for participants and moral
dilemmas, was computed. There were significant main effects of emphasis ($z = -2.10, p = .036$), mindset condition ($z = -2.01, p = .044$), and task difficulty ($z = -2.27, p = .023$). In addition, the interaction of emphasis and difficulty ($z = 2.20, p = .028$) was significant. Importantly, the interaction of mindset condition and emphasis was significant ($z = 2.18, p = .029$). This effect was qualified by a significant three-way interaction of mindset, emphasis, and task-difficulty in the opposite direction ($z = -1.97, p = .048$), indicating a weaker effect for participants who had more difficulties with the task. For more detail on the model see Table 2a. The proportion of consequentialist choices is shown in Figure 13.

**Table 2.** Mixed linear logit models estimating moral judgments and focus on means in Experiment 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>SE</th>
<th>$z$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Moral judgments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.42</td>
<td>0.76</td>
<td>3.20</td>
<td>.001</td>
</tr>
<tr>
<td>Mindset $^a$</td>
<td>-2.11</td>
<td>1.05</td>
<td>-2.02</td>
<td>.044</td>
</tr>
<tr>
<td>Emphasis $^b$</td>
<td>-2.08</td>
<td>0.99</td>
<td>-2.10</td>
<td>.036</td>
</tr>
<tr>
<td>Difficulty</td>
<td>-0.45</td>
<td>0.20</td>
<td>2.27</td>
<td>.044</td>
</tr>
<tr>
<td>Mindset $^a$ × Emphasis $^b$</td>
<td>3.27</td>
<td>1.50</td>
<td>2.18</td>
<td>.029</td>
</tr>
<tr>
<td>Difficulty × Emphasis $^b$</td>
<td>0.59</td>
<td>0.27</td>
<td>2.20</td>
<td>.028</td>
</tr>
<tr>
<td>Mindset $^a$ × Difficulty</td>
<td>0.52</td>
<td>0.29</td>
<td>1.82</td>
<td>.069</td>
</tr>
<tr>
<td>Mindset $^a$ × Emphasis $^b$ × Difficulty</td>
<td>-0.80</td>
<td>0.41</td>
<td>-1.97</td>
<td>.048</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>SE</th>
<th>$z$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>b) Proportion of fixations on the means</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.16</td>
<td>0.30</td>
<td>-7.17</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mindset $^a$</td>
<td>-0.24</td>
<td>0.12</td>
<td>-2.04</td>
<td>.042</td>
</tr>
<tr>
<td>Emphasis $^b$</td>
<td>-0.04</td>
<td>0.11</td>
<td>-0.38</td>
<td>.705</td>
</tr>
<tr>
<td>Mindset $^a$ × Emphasis $^b$</td>
<td>0.17</td>
<td>0.16</td>
<td>1.07</td>
<td>.285</td>
</tr>
</tbody>
</table>

*Note. $^a$ deliberative = 1, implemental = -1. $^b$ AC = -1, AD = 1*
Figure 13. Proportion of endorsement of the consequentialist options by experimental conditions in Experiment 2.

**Gaze data.** The eye-tracking data were prepared for analysis using the eyetrackingR (Dink & Ferguson, 2015) and the saccades (von der Malsburg, 2015) packages for R. Data from two participants were discarded because of low data quality (high amount of track loss). We calculated fixations during the passive picture presentation phase for each trial with the built-in function of the saccades package. Figure 14 shows the distribution and density of fixations for each dilemma separate for the deliberative and the implemental mindset condition. Two-dimensional kernel density is an indicator for the clustering of data points in a two-dimensional space. In the case of fixations, higher density signifies that the gaze was more concentrated and fixations were closer to each other. Thus, this serves as an indicator of focus. As can be seen in Figure 14, the maximum density averaged over trials was higher in the implemental mindset ($M = 6.397 \times 10^{-6}$, $SD = 0.874 \times 10^{-6}$) than in the deliberative mindset ($M = 4.803 \times 10^{-6}$, $SD = 1.541 \times 10^{-6}$).

A mixed linear logit model over all fixations was computed to test whether the higher fixation density in the implemental mindset was a result of focusing on a particular area of interest (i.e., the means). The experimental conditions and their interaction served as predictors, random intercepts were included for participants and trials. Mindset-task difficulty had no effect on fixations ($z < 0.98$, $p > .323$), thus this control variable was not included in the model. For a summary, see Table 2b. In the model the main effect of mindset condition was significant, $z = -2.04$, $p = .042$, ...
indicating a higher proportion of fixations on the means for participants in the implemental mindset condition ($M = .14$) than participants in the deliberative mindset condition ($M = .12$). The experimental conditions had no significant effect on fixations on the group harmed in case the deontological option was chosen ($z_s < 0.68, ps > .49$) and the group harmed in case the consequentialist option was chosen ($z_s < 0.59, ps > .55$).

**Discussion**

As can be seen in Figure 13, the emphasis effect was moderated by the currently active mindset. While participants in the deliberative mindset condition chose congruently with the emphasis manipulation, this bias was absent for participants in the implemental mindset condition. This is a conceptual replication of the main finding of Experiment 1 and further confirms that mindsets are an effective regulating mechanism for emphasis effects.

We argue that the implemental mindset counteracts the emphasis bias because individuals in an implemental mindset are more focused on goal-relevant information. In line with Mindset Theory of Action Phases (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016), we argue that selective attention in an implemental mindset is directed primarily at goal-relevant means. Our eye-tracking data support this hypothesis. As can be seen in Figure 14, the overall density of fixations is higher in the implemental mindset condition than in the deliberative mindset condition. This indicates that the fixations in the deliberative mindset were more broadly distributed over the entire screen, while fixations in the implemental mindset were more focused on particular areas. These findings line up with Büttner et al. (2014) who have shown that visual attention is narrower in the implemental mindset than in the deliberative mindset. Here we went one step further and tested what individuals in an implemental mindset focused on while looking at a scene with goal-relevant, meaningful content. The proportion of fixations on means for actions was higher for participants in the implemental mindset. A plausible explanation for this pattern is that an implemental mindset tunes attention and processing towards goal-directed means. This focus effect is likely the cause for the moderating role of mindsets on the emphasis bias in moral judgments. Individuals in the deliberative mindset are more open to incidental information (e.g., emphasis), whereas individuals in the implemental mindset focus on the core content of the task at hand (e.g., dilemma story, action-relevant means).
Figure 14. Distribution and density of fixations for each dilemma separated by mindset conditions in Experiment 2. Dots represent individual fixations; colors indicate the density of fixations.
Experiment 3: Testing Thorough Reflection as an Alternative Control Mechanism

In the third experiment, we explored whether a simple induction of thorough reflective thought qualifies as an alternative method to control the emphasis bias. To induce different modes of thought we used *implementation intentions*, a self-regulatory strategy that has been used to activate intuitive and reflective thinking in past studies (Bieleke et al., 2016; Doerflinger et al., 2017). Generally, implementation intentions are if-then plans that link a goal-directed response (e.g., a mode of thinking) to a situational cue (e.g., reading a moral dilemma). They have the format: “If I encounter situation S, then I will initiate response R.” (Gollwitzer, 1999, 2014). If emphasis influences moral judgments because individuals spent too little time and effort elaborating the available information, then planning to think thoroughly about a dilemma should result in a weaker emphasis effect, and planning to make spontaneous judgments should result in a stronger effect of emphasis. Importantly, in the present experiment the participants were instructed to plan to think more versus less thoroughly without any specification concerning what particular content they should think about.

The participants first provided information about their general inclination towards reflective thinking (i.e., need for cognition; Cacioppo, Petty, & Feng Kao, 1984) and impulsivity. These variables reflect the disposition for reflective reasoning and intuitive decision making on the trait level and could potentially affect moral judgments. Next, an intuitive vs. reflective mode of thought was induced. Finally, the participants were asked to provide moral judgments about trolley-type dilemmas that had either no emphasis added, had an anti-deontological emphasis (AD), or dilemmas with an anti-consequentialist emphasis (AC).

**Method**

**Participants and design.** The experiment had 3 (emphasis: AC vs. no emphasis vs. AD) × 2 (mode of thought: intuitive vs. reflective) between-subjects design and was conducted as a laboratory experiment. Implementation intentions typically have medium to large effects (Gollwitzer & Sheeran, 2006). Assuming overall medium-sized effects and a moderate correlation between responses to different dilemmas, we estimated with the G*Power software (Faul et al., 2007) that a sample size of 186 participants would be needed to achieve a statistical power greater...
than $1 - \beta = .95$. 212 German speaking participants (172 female) with a mean age of 22 (age range 17 – 56, $SD = 3.8$) were recruited at the University of Konstanz. The main dependent variable was moral judgments of the trolley type dilemmas. Need for cognition (Cacioppo et al., 1984) and general impulsivity were assessed as covariates.

**Procedure.** Upon entering the laboratory, the participants were randomly assigned to a mode of thought and an emphasis condition. Each session started with the measurement of the covariates, which was followed by the mode of thought induction and finally the moral judgment task. Need for cognition was assessed with the German 19-item version of the need for cognition scale (Preckel, 2014). Impulsivity was assessed with a go-stop task based on Dougherty, Mathias, Marsh, and Jagar (2005). The participants worked on the go-stop task for 150 trials. In the task, participants had to react to strings of numbers presented on a screen in black on a grey background. If two consecutive numbers in a trial were identical, the participants had to respond with a keypress. However, if the second number was colored red, they should refrain from reacting. In the task, the percentage of false positive responses signifies impulsivity.

Next, all participants received a short, written description of the assigned implementation intention and the instruction to adopt it, visualize it, and write it down three times on a sheet of paper. Participants in the intuitive mode of thought condition were asked to adopt the plan: “If I make a judgment about the short stories, then I decide immediately and spontaneously!” (German: “Wenn ich die Kurzgeschichten beurteile, dann entscheide ich sofort spontan!”), and participants in the reflective mode of thought conditions were asked to adopt the plan: “If I make a judgment about the short stories, then I deliberate thoroughly!” (German: “Wenn ich die Kurzgeschichten beurteile, dann denke ich gründlich nach!”).

After the mode of thought induction, the participants worked on the moral dilemma task. The task was implemented in the software PsychoPy (Peirce, 2007). Each participant was asked to rate eight trolley-type dilemmas. The same dilemmas were used as in Experiment 1. The participants saw only dilemmas corresponding to their emphasis condition. In the AC condition, dilemmas were presented with an emphasis on the harm that would come to the single person or smaller group, if the consequentialist option was chosen; in the no emphasis condition no emphasis was added; and in the AD condition the dilemmas were presented with an emphasis on the harm that would come to the larger group if the deontological option was chosen. The
dilemmas were first presented in random order. The participants were instructed to press a key as soon as they had completed reading a dilemma to get to the next screen on which they could rate the moral wrongness of the consequentialist option (e.g., “It is morally wrong to hit the switch.”) on a six-point scale ranging from “I agree” to “I disagree”. We recorded moral judgments and response times. Response times were measured as the time between reading the dilemma and answering the moral judgment questions. After providing moral judgments for each dilemma, the dilemmas were again presented in randomized order and the participants were asked to indicate on six-point scales ranging from “I agree” to “I disagree” how emotional they thought the dilemmas were (“The story was very emotional.”) and how difficult it was to make a moral judgment (“Making a decision in the dilemma was very difficult.”). Emotionality and difficulty ratings were assessed after the moral judgments to avoid a possible influence on moral judgments of explicitly considering emotionality and difficulty of the dilemmas.

At the end of the experiment, the participants were asked to provide demographic information, then they were debriefed about the purpose of the experiment, thanked, and reimbursed with 8 € as a compensation for taking part in the experiment.

Results

Manipulation check. If the mode of thought manipulation was successful, then participants in an intuitive mode of thought should be faster than participants in a reflective mode of thought. A general linear mixed model was calculated to test differences in reaction times. The mode of thought condition, the emphasis condition, and their interaction were entered as predictors. Reaction times for moral judgments of each dilemma were the criterion. Random intercepts were estimated for each participant and each dilemma. There was a significant main effect of mode of thought, $t(207.68) = 2.195, p = .004$, indicating faster reactions ($M = 8.71s, SD = 9.19s$) by participants in an intuitive mode of thought than by participants in a reflective mode of thought ($M = 6.93s, SD = 5.12s$). Neither the emphasis condition nor the interaction of emphasis and mode of thought significantly influenced reaction times ($ts < 1.26, ps > .21$). Although the emphasis manipulation was subtle, we expect emotionality ratings in the AD and the AC conditions to be higher than in the no emphasis condition. Although there was a trend in this direction with higher emotionality
ratings in the emphasis conditions ($M = 3.62, SD = 1.15$) compared to the no emphasis condition ($M = 3.50, SD = 1.18$), this effect did not reach significance, $t(208.45) = 1.34, p = .183$. Perceived difficulty did not differ between experimental conditions, $t < 1.48, ps > .141$.

**Moral dilemma judgments.** Moral judgments were predicted based on a linear mixed model with the experimental conditions and their interaction as predictors and random intercepts for participants and dilemmas. General need for cognition and impulsivity did not significantly affect moral judgments ($ts < 1.13, ps > .26$) and are consequently not included as covariates in the model reported here. There was a highly significant main effect of the emphasis conditions on moral judgments, $t(208.21) = 4.45, p < .001$. The main effect of mode of thought, $t(207.00) = 0.90, p = .370$, and the interaction of emphasis and mode of thought, $t(208.21) = 1.34, p = .182$, were not significant. See Table 3 for more detail on the model. As can be seen in Figure 15, the endorsement of the consequentialist option was lower in the AC condition than in the no emphasis condition and higher in the AD condition.

### Table 3. Mixed linear models estimating moral judgments in Experiment 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.33</td>
<td>0.26</td>
<td>8.39</td>
<td>13.55</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mode of thought$^a$</td>
<td>0.04</td>
<td>0.05</td>
<td>210.00</td>
<td>0.90</td>
<td>.370</td>
</tr>
<tr>
<td>Emphasis$^b$</td>
<td>0.27</td>
<td>0.06</td>
<td>211.34</td>
<td>4.47</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mode of thought$^a$ × Emphasis$^b$</td>
<td>0.08</td>
<td>0.06</td>
<td>211.30</td>
<td>-1.34</td>
<td>.180</td>
</tr>
</tbody>
</table>

*Note.* $^a$ reflective = 1, intuitive = -1, $^b$ AC = -1, no emphasis = 0, AD = 1
Discussion

The emphasis bias effect was again replicated in Experiment 3. Importantly, emphasizing the negative consequences that would occur if a deontological option was chosen biased the participants to make more consequentialist moral judgments and emphasizing the negative consequences of a consequentialist choice tilted moral judgments towards deontology. Both an anti-deontological emphasis as well as an anti-consequentialist emphasis could influence moral judgments in a predictable pattern. This bias occurred independently of whether participants planned to make intuitive or reflective decisions.

Our reaction time data suggest that planning to rely on intuition vs. reflective reasoning worked. Participants who planned to make spontaneous judgments were faster to respond than participants who planned to make reflective judgments. The mode of thought had no effect on moral judgments. The moral judgments of participants who planned to engage in more thorough reflection did not differ from those of the participants who planned to make spontaneous judgments. Importantly, both groups were equally susceptible to the emphasis bias. This can be interpreted as evidence that regulating how much we think about a moral dilemma has little effect
on the influence of emphasis. Increasing reflective reasoning can be expected to be especially adaptive in situations where a bias arises because agents process too little information and the agent desires to overcome this bias (Doerflinger et al., 2017). The emphasis added in our dilemmas, although not crucial to the content of the dilemmas, is additional information. Thus, the influence of emphasis on moral judgments is not due to too little processing but rather a result of processing too much information.

**General Discussion**

We have shown in three experiments that verbally emphasizing consequences in trolley-type dilemmas can bias moral judgments to be both more consequentialist if the negative consequences of a deontological choice are emphasized, and to be more deontological if the negative consequences of a consequentialist choice are emphasized. This is an important addition to the existing literature on moral judgment and emotions (e.g., Greene, 2010) because it highlights that emotional responses are directed at an object. Taking the target of emotions into account should lead to more precise predictions of moral judgments. Furthermore, the results suggest that emotional processes can in fact bias individuals to be more consequentialist. Even if consequentialist judgments rely generally more on cognitive processes and deontological judgments rely typically more on emotional processes, emotional presentation can be used to sway moral judgments either way.

However, the emphasis bias depends on the cognitive orientations induced by action phases. Being in a deliberative mindset coincides with a widening of attention and processing, which opens the agent to influences by non-relevant information (e.g., emphasis). Being in a deliberative mindset is adaptive when the agent needs to take in as much information as possible to make an informed decision (Gollwitzer, 2012; Gollwitzer & Keller, 2016). Once a goal is set (e.g., “I want to make an unbiased moral judgment.”), a cognitive tuning towards indiscriminate information processing and undirected attention is no longer adaptive. Instead, the implemental mindset shields agents from distractions and non-relevant information. These processes modulate the influence of incidental emphasis in a moral dilemma. Selective attention and processing in the implemental mindset make the agent less prone to be influenced by emphasis.

Our data suggest that the emphasis bias is unaffected by planning to make moral judgments in a more reflective manner. Both participants who planned to
reflect thoroughly and participants who planned to make spontaneous judgments were affected by emphasis. Because the emphasis bias relies at least partly on emotion, which is rooted in the hot system, one might expect that reflective reasoning should reduce the bias. This hypothesis is not supported by our experiments. One possible explanation is that the emphasis bias is composed of a cognitive as well as an emotional component. Trolley-type dilemmas are relatively complex stimuli, and processing the emphasis added by expressive language requires at least some degree of reflective reasoning. Thus, the bias might not be exclusively an emotional phenomenon. Additionally, simply increasing the amount of reflection and thinking is an adaptive strategy if one faces problems where otherwise too little information is processed. The scenarios in the present experiments do not map onto such a case. Thinking more about a decision may very well lead the agent to ruminate on biasing information. The sobering conclusion of this finding is that thinking harder about a moral dilemma does not immunize us against arbitrary biases. If a decision maker wants to avoid such biases, a more viable strategy than regulating how much we think about a dilemma would be to regulate how we think about it.

Conclusion

In sum, the present paper makes two major contributions to research on moral dilemma judgments and a third major contribution to research on goal-striving. First, we demonstrated that emotional emphasis can be directed at an anti-consequentialist target and an anti-deontological target. Such a targeted emphasis results in a bias against the targeted option. This finding is important because it implies that a negative emotional tag is added to the targeted option. Consequently, directed affective processes can promote consequentialist as well as deontological preferences. Second, cognitive orientations affect how individuals respond to moral dilemmas with an added emphasis. Mindsets alter the selectivity of attention and processing. Emphasis effects are increased if attention and processing are undirected (in the deliberative mindset) and decreased if the agent is focused on task-relevant content (in the implemental mindset). Third, we demonstrate with an eye-tracking experiment that attention in the deliberative mindset is indeed more broadly distributed, whereas attention in the implemental mindset is more focused. In line with Mindset Theory of Action Phases (Gollwitzer, 1990, 2012; Gollwitzer & Keller, 2016), attention in the implemental mindset is focused on goal-directed means.
Motivational science can benefit from investigating moral judgments. Moral issues are those that humans care deeply about (Greene, 2014) and moral dilemmas are complex decision cases — yet moral dilemmas, like trolley-type dilemmas, are highly structured and represent a promising testing ground for motivational theories. Our findings obtained with eye-tracking in the context of moral dilemmas can be expected to generalize to other areas of judgment and decision making. In fact, it would be surprising if the means-focus we observed for participants in the implemental mindset was unique to the moral realm. Instead, we hypothesize that the implemental mindset generally supports attention and processing in service of selected goals while simultaneously shielding the goals from information that could hamper goal striving. In conclusion, motivational science offers a rich arsenal of methods and theory to further the understanding of moral judgment, and moral psychology might be a fruitful area to explore general motivational and volitional mechanisms.
References


References


Doerflinger, J. T., Martiny-Huenger, T., & Gollwitzer, P. M. (2017). Planning to deliberate thoroughly: If-then planned deliberation increases the adjustment of


Gollwitzer, P. M. (1981). *The social reality of self-symbolizing: Winning completeness through others*. (Doctoral dissertation), University of Texas at Austin,


References


Kruglanski, A. W., & Köpetz, C. (2009). What is so special (and nonspecial) about goals. In G. B. Moskowitz & H. Grant (Eds.), *The psychology of goals* (pp. 27–55). New York: Guilford Press.


Nucci, L. P. (2002). Because it is the right thing to do. Human Development, 45, 125–129. doi:10.1159/000048158


References


Record of Achievement

All the presented research papers were the product of cooperation with other psychological researchers inside and outside the University of Konstanz. I am thankful for the effort, ideas, and knowledge any of my coauthors brought into our research projects. Generally, the order of authors indicates the contribution an author had on the research paper. Specifically, I now want to list my contribution to each research paper.

**Research Paper I**

I contributed to the following: search, selection, and review of the studies, interpretation and theoretical refinement, and writing the manuscript.

**Research Paper II**

I contributed to the following: design of all experiments, programming of the task paradigms, creating experimental material, data collection and analysis, interpretation and theoretical refinement, and writing the manuscript.

**Research Paper III**

I contributed to the following: design of all experiments, programming of the task paradigm, creating experimental material, data collection and analysis, interpretation and theoretical refinement, and writing the manuscript.