PUSHING, COASTING, DISENGAGING:

A MOTIVATIONAL FRAMEWORK OF SOCIAL COMPARISON

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Promotionsordnung vom 18.12.2018

vorlegt von

Katharina Helene Diel

aus

Osnabrück

Vorgelegt am 12. Juli 2021

Erstgutachter: Prof. Dr. Wilhelm Hofmann

Zweitgutachterin: PD Dr. Simone Dohle

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität Köln im November 2021 angenommen.

Datum der mündlichen Prüfung: 26.11.2021

To Naomi,

may your life be full of inspiring and attainable role models.

ERKLÄRUNG

Chapter 2 beruht auf folgendem Manuskript:

Diel, K., & Hofmann, W. (2019). Inspired to perspire: The predicted interplay of social comparison direction and standard extremity in the context of challenging exercising goals. *Social Cognition*. *37*(3), 247-265, http://dx.doi.org/10.1521/soco.2019.37.3.247

Der zweite Autor hat die erste Idee entwickelt. Die Weiterentwicklung und die Idee zum Studiendesign stammen von mir. Ich habe die Daten erhoben, die Daten analysiert und das Manuskript geschrieben, welches durch wertvolle Beiträge des zweiten Autors profitiert hat.

Chapter 3 beruht auf folgendem Manuskript:

Diel, K., Grelle, S., Hofmann, W. (2021). A motivational framework of social comparison. *Journal of Personality and Social Psychology*. *120*(6), 1415-1420. http://dx.doi.org/10.1037/pspa0000204

Die Idee wurde von mir zusammen mit dem dritten Autor entwickelt. Ich habe die Daten erhoben, die Daten analysiert und das Manuskript geschrieben. Der dritte Autor hat zu jedem Schritt wertvolle Vorschläge beigetragen.

Chapter 4 beruht auf folgendem Manuskript:

Diel*, K., Broeker*, L., Raab, M., & Hofmann, W. (2021). Motivational and Emotional Effects of Social Comparison in Sports. *Psychology of Sport and Exercise. 57,* 102048. https://doi.org/10.1016/j.psychsport.2021.102048

*geteilte Erstautorenschaft

Die Idee wurde von mir entwickelt. Die Weiterentwicklung der Idee profitierte von Beiträgen des vierten Autors. Zusammen mit der zweiten Autorin habe ich die Daten erhoben, die Daten analysiert

und das Manuskript geschrieben. Der dritte und vierte Autor haben zu jedem Schritt wertvolle Vorschläge beigetragen.

Chapter 5 beruht auf folgendem Manuskript:

Fleischmann, A., Lammers, J., Diel, K., Hofmann, W., & Galinsky, A.D. (2021). More threatening and more diagnostic: How moral comparisons differ from social comparisons. *Journal of Personality and Social Psychology, 121*(5), 1057-1078. http://dx.doi.org/10.1037/pspi0000361

Die erste Autorin hat die Idee entwickelt, die Daten analysiert und das Manuskript geschrieben. Eine Studie vom Manuskript wurde von mir mitentwickelt. Die Daten zu dieser Studie habe ich erhoben und analysiert und die Studie wurde von mir als Teil des Manuskriptes aufgeschrieben. Der zweite, vierte und fünfte Autor haben zu jedem Schritt wertvolle Vorschläge beigetragen.

Abstract

People often strive towards self-improvement, and whether they are currently motivated or lacking motivation can hinge on their social environment, namely on the people around them. Research on social comparison has emphasised the inspirational nature of upward comparisons, but has not covered the whole spectrum of possible comparison standards in terms of their motivational impact. The current dissertation integrates social comparison with motivational principles from self-regulation research and investigates how discrepancy assessments between the self and a standard of comparison influence people's motivation and emotion. Chapter 1 outlines the new framework of motivation by integrating literature on social comparison and self-regulation. In Chapter 2, a line of experimental studies shows how motivational patterns provoked by moderate (vs. extreme) upward (vs. downward) standards differ from classic assimilation and contrast judgments in the social comparison literature. Chapter 3 moves into the field, and demonstrates how discrepancy assessments between the self and a standard influence motivation and associated emotion, where a negative discrepancy (upward comparison) is associated with increased effort investment (pushing), but with decreased effort investment if the discrepancy becomes too large (disengagement). Positive discrepancies, on the other hand, are related to less effort investment as the standard has been surpassed (coasting). Chapter 4 complements the former chapter with a behavioral effort measure that assesses performance improvement of sports students after they have reported on their social comparisons during the semester. Lastly, Chapter 5 illustrates how the domain of morality deviates from the motivational framework as people draw motivation from downward and avoid upward comparison to positively distinguish their moral self from others. Finally, the last chapter offers concluding thoughts concerning implications, future research, and limitations of the current research.

Keywords: social comparison, self-regulation, motivation, emotion, morality

Deutsche Zusammenfassung

Menschen streben häufig nach Selbstverbesserung, und ob sie gerade motiviert oder unmotiviert sind, kann von ihrem sozialen Umfeld abhängen. Die Forschung zu sozialen Vergleichen hat den inspirierenden Charakter von Aufwärtsvergleichen hervorgehoben, aber nicht das gesamte Spektrum möglicher Vergleichsmaßstäbe in Bezug auf ihre motivationale Wirkung erfasst. Die vorliegende Dissertation integriert soziale Vergleichsprozesse mit motivationalen Prinzipien aus der Selbstregulationsforschung und untersucht, wie Diskrepanzbewertungen zwischen dem Selbst und einem Vergleichsstandard, die Motivation und Emotionen von Menschen beeinflussen. Kapitel 1 beschreibt einen neuen Motivationsrahmen durch Integration der Literatur zu sozialen Vergleichen und zur Selbstregulation. In Kapitel 2 zeigen eine Reihe von experimentellen Studien, wie sich die Motivationsmuster, die durch moderate (vs. extreme) Aufwärts- (vs. Abwärts-) Vergleiche hervorgerufen werden, von den klassischen Assimilations- und Kontrasturteilen der Literatur zu sozialen Vergleichen unterscheiden. Kapitel 3 führt in das Feld ein und zeigt, wie Diskrepanzbewertungen zwischen dem Selbst und einem Standard, die Motivation und die damit verbundenen Emotionen beeinflussen, wobei eine negative Diskrepanz (Aufwärtsvergleich) mit einer erhöhten Anstrengungsinvestition (Pushing), aber mit einer verringerten Anstrengungsinvestition verbunden ist, wenn die Diskrepanz zu groß wird (Disengagement). Positive Diskrepanzen hingegen sind mit geringerer Anstrengungsinvestition verbunden, da der Standard übertroffen wurde (Coasting). Kapitel 4 ergänzt das vorherige Kapitel mit einem verhaltensbezogenen Anstrengungsmaß, das die Leistungsverbesserung von Sportstudierenden bewertet, nachdem sie über ihre sozialen Vergleiche während des Semesters berichtet haben. Schließlich illustriert Kapitel 5, wie der Bereich der Moral vom motivationalen Rahmen abweicht, da Menschen ihre Motivation aus dem Abwärtsvergleich ziehen und den Aufwärtsvergleich vermeiden, um ihr moralisches Selbst positiv von anderen abzuheben. Schließlich bietet das letzte Kapitel abschließende Gedanken zu Implikationen, zukünftiger Forschung und Einschränkungen der aktuellen Forschung.

Schlagwörter: Soziale Vergleiche, Selbstregulation, Motivation, Emotionen, Moral

ACKNOWLEDGEMENTS

This dissertation could not have been written without all the support I received from many people over the years. First of all, thank you Wilhelm, for offering me the position as your PhD student and for giving me the great opportunities to research and travel. You gave me a place to grow and I have learned a lot from you over the years. If you ever need a DJ for future garden parties, you know where to find me.

Thank you Simone, for accepting to be my second supervisor and for all your help and advice in the last years. Thank you for always having an open door and for all great conversations whether about research or sparkly pink shoe mispurchases.

To Matt, my mentor and unofficial third supervisor: You have always been an inspiration to me and I learned so much from you (Frisbee skills included). Thank you for motivating me and believing in me, when I needed it the most and for all the shared humor.

To my office mates, Lena and Mike: Thank you for your daily support, coffee breaks, fun conferences and amazing Adventskalender. Thank you Lena, for your friendship and the most delicious office cakes. Maybe I'm a cake-person after all. Thank you Mike for always checking my references and for totally non-judgmental conversations behind the closed door.

To my friend and colleague Adam: Thank you for always having an open door and open ear, your late at night Kölsche Leeder voice messages, and for always being available at Karneval, when everyone else is already exhausted.

To Sonja, Lucy, Stephen, Elke and Heike: Thank you for being an amazing Team in Bochum and for all your support and fun Fun Days.

To Julia and Max, my favorite Hiwis: Thank you for your great work, all your support with my research and Kegel- and Flunky Ball competitions (you didn't stand a chance).

To Rita and Lea: Both of you became very close friends, thank you for always being there for me and handing me a beer or Vinho Verde when needed.

To the Random Group, for always giving advice when needed and for numerous fun nights out.

Danken möchte ich auch meiner Familie und meinen Freunden, die mich nicht nur die letzten Jahre, sondern schon mein ganzes Leben unterstützen.

Sophie, danke, dass du mir immer so viel Selbstvertrauen geschenkt hast. Du bist und bleibst mein großes Vorbild und ohne dich wäre ich nicht der Mensch, der ich heute bin. Neben einer ganze Menge Blödsinn, haben wir doch auch schon einiges auf die Beine gestellt in den letzten 20 Jahren. Da können wir auch mal stolz sein!

Kristina und Caro, danke für die beste Studienzeit, unsere Power-Lernsessions und Abende im *Golden Fust*. Danke für eure endlose Unterstützung, ohne die ich heute nicht da wäre, wo ich bin.

Ankris, Verena, Ina, Julia, ihr bierigen Kinder: Danke, dass ihr mir immer die Freizeit versüßt.

Danke an meine Familie, an meine großen Schwestern Miri, Dany und Andrea, meine großen Neffen, Lukas, Simon und Maxi und die kleinen Quatschköpfe, Helene, Felix, Lilli und Julius. Danke, dass ihr immer für mich da seid und mich zum Lachen bringt.

Lieber Papsi, du hast mir vorgelebt und gezeigt, dass man mit harter Arbeit, Ehrgeiz und ein wenig Leidenschaft für den Beruf Großes erreichen kann. Danke, dass du mir das alles ermöglichst hast!

Liebe Mamsi, die Danksagungen an dich würden ein ganzes Buch füllen, deswegen versuche ich mich kurzufassen: Danke, für deine endlos große Liebe und Unterstützung und das Gefühl sowieso immer stolz zu sein, egal welchen Weg ich im Leben einschlage.

Trudi, John und Snow: Danke für eure Kuscheleien, Liebe und Gesellschaft im Homeoffice. Ihr seid so flauschig…

Felix, du kennst alle meine Höhen und Tiefen und ohne dich, deine Liebe und Unterstützung hätte ich das hier nicht geschafft. Danke, dass du auch mich zum kleinen Nerd gemacht hast und einfach weißt, wann es Zeit ist, das Feierabendbier zu öffnen.

Naomi, danke, dass es dich gibt und du mich täglich zum Lachen bringst. Du bist der beste Grund Feierabend zu machen.

DANKE!

Chapter I: General Introduction

"Comparison is the thief of joy" – a quote that may sound familiar and conveys the message that the comparison to others will do more harm than good. Who has not experienced situations where the comparison to other people dragged us down: Everyone else appears to be more successful, healthier or happier, while ourselves seem to struggle with what we had intended to do or to be. Indeed, a large body of research has shown that the comparison to more successful others affects the self negatively, for instance, when observing people who seem to have a better social life (Davidai et al., 2020; Davidai & Deri, 2019; Deri et al., 2017), show better academic performances (Rogers & Feller, 2016), or are better athletes (Marsh et al., 1997). Those detrimental effects of social comparisons are especially prevalent on social media (Midgley et al., 2020; Nesi & Prinstein, 2015), for instance when people compare appearances and looks (de Vries & Kühne, 2015; Fardouly et al., 2017), or their roles of being parents (Coyne et al., 2017; Sidani et al., 2020). Downsides of social media even worsened during the Covid pandemic when social isolation led people to spend even more time online (Ruggieri et al., 2020). Frequent comparisons to others increase feelings of envy, guilt and regret, are accompanied by negative behaviors, such as lying (White et al., 2006) and can even lead to mental health disorders, such as depression (Liu et al., 2017). Given all this evidence, should one follow the advice to stop comparing altogether? Or can people actually benefit from comparisons, such as when role models inspire and guide people towards a better version of themselves (Lockwood & Kunda, 1997c; Luong et al., 2020)? Especially since people strive towards all kinds of long-term self-improvement goals (e.g. money savings, job promotion, or a healthy lifestyle), they are subject to challenges that come with goal pursuit, such as overcoming short-term rewards and temptations (e.g. spending money for short term pleasure, binging on unhealthy food; Hofmann et al., 2009, 2011; Kotabe & Hofmann, 2015). Hence, positive role models may facilitate their path to success. However, not all comparisons give a push in the desired direction as people may be apt to compare to non-relevant others. In this thesis, I investigate a wide spectrum of possible social comparison standards and their motivational impact on people by taking a self-regulatory approach to social comparison processes. To do so, I first propose a new framework that combines social comparison processes and self-regulation research to better capture motivational ups and downs (Chapter 1). In Chapter 2, a line

of empirical studies demonstrates to what extent motivational patterns differ from pure social-cognitive outcomes (i.e. judgements) triggered by different comparison standards. In Chapter 3 and 4, I investigate how discrepancy assessments between the self and a comparison standard influence people's motivation and emotion (Chapter 3), and behavior in everyday life (Chapter 4). In Chapter 5, I illustrate how certain comparison domains, such as morality, may prompt different motivational patterns from the proposed framework. And finally, in Chapter 6, I discuss implications of the findings, future research and limitations of the current research.

1.1 Social Comparison Motives

More than ten percent of our daily thoughts resolve around comparative thinking (Summerville & Roese, 2008). In this regard, intentionally refraining from daily social comparisons in order to avoid negative consequences might not be that simple, because a great deal of comparisons happen spontaneously and outside people's conscious awareness (Mussweiler et al., 2004; Ohmann et al., 2016). People compare to others even if the comparison itself is irrelevant to the individual (Tversky & Kahneman, 1974). First and foremost, social comparison guides people to evaluate their standing in the world. Every day, people may seek feedback on their abilities, performances, and appearances; and in order to get an accurate picture of one's current standing in the absence of an objective standard, people compare their abilities and achievements to their social environment (Festinger, 1954; Taylor, Wayment, & Carrillo, 1996). Consider a person named Norah - before deciding to take the next career step, she has most likely oriented herself toward others in order to evaluate her professional strengths and weaknesses. For most precise self-views, people engage in comparisons with others who are perceived as similar on an evaluated dimension (lateral comparison; Festinger, 1954; Wheeler, 1966). Meaning, if Norah wants to accurately assess her professional performance, she will most likely compare to others from her own profession, her co-workers for instance, to evaluate her current skills and achievements. However, for most diagnostic standards, people may compare to others who are similar not only on the critical, but also on related dimensions (Goethals & Darley, 1977). So, if Norah was a data scientist, it might be diagnostic for Norah to compare to other data scientists in her company or elsewhere, but even more diagnostic if she compared to other data scientists who were in her own age group or who had an equal number of years of professional experience. In sum, social

comparison is a fast and efficient tool to gain evaluative self-knowledge (Corcoran & Mussweiler, 2009; Mussweiler & Epstude, 2009) and therefore is an indispensable part of everyday life. With regard to long-term goals, it is crucial to initially assess one's current standing in order to set an appropriate goal state (Carver & Scheier, 1982, 1990). For example, a person might want to evaluate their current fitness level first in order to determine how far they deviate from their desired fitness state.

Although lateral comparisons serve one of the main motives of social comparison, they are by far not the most prevalent ones (Gerber et al., 2017). Even more frequently, people compare to others who are worse off (downward comparison) to fulfil their need of self-enhancement: Comparing to inferior others can give a boost in self-esteem and positive affect (Morse & Gergen, 1970; Taylor & Lobel, 1989; Wills, 1981). According to downward comparison theory, a current threat to one's self-esteem even increases the need to compare downwards (Hakmiller, 1966; Wills, 1981). For instance, to overcome a setback at work, comparing to others who are even worse off might decrease negative and increase positive feelings. On the other hand, upward comparison – the comparison to others who are better off, is associated with self-improvement. Comparing to someone who is more successful can inspire and motivate to reach personally important goals (Taylor & Lobel, 1989). If Norah is uncertain about whether she should start her own business, comparing herself to others who have already successfully gone down this path can motivate her. The encouraging effects of upward comparison look promising. However, at the same time, upward comparison can increase negative affect and threaten the self-esteem (Morse & Gergen, 1970), which more closely aligns with the examples of social comparison's detrimental effects described previously. Despite triggering occasional negative outcomes, people have a general preference for upward comparison (Gerber et al., 2017), which could indicate that the encouraging effects of upward comparison might somewhat outweigh the negative effects. In this regard, it is important to note that in order to benefit from upward comparison, people need to assimilate with successful others. As shown by experimental studies, participants choose to compare with other individuals who performed only slightly better on a performance task (Collins, 2000; Thornton & Arrowood, 1966; Wheeler, 1966), which allows for assimilation to those higher standards (Collins, 1996, 2000; Gruder, 1971). To understand under what circumstances people

assimilate with an upward standard or rather contrast away, it is crucial to dive deeper into cognitive processes of social comparison.

1.2 Cognitive theories of comparison processes

The outcome of a comparison does not only depend on whether a person compares upward or downward, but also whether this person assimilates with the standard or perceives themselves as discrepant from the standard. In other words, if a target is confronted with a comparison standard, the target either perceives her or himself and the standard as one unit or as two separate units (Schwarz & Bless, 2007). For instance, if an athlete makes use of available social comparison information (e.g. comparing to a successful athlete), that information can be either incorporated into the self, leading to the process of *assimilation* ("I am also athletic") or can be excluded from the self, causing the self to be distant from the social comparison information, leading to *contrast* ("My athletic abilities are weak"). According to the selective-accessibility model (SAM; Mussweiler, 2003), if a target evaluates a comparison standard and spontaneously perceives this person as similar, the target engages in similarity testing – integrating all information that makes the target and standard similar – whereas perceiving the other as dissimilar will prompt dissimilarity testing, integrating all information that makes the target and standard dissimilar. Similarity testing leads to assimilation and dissimilarity testing to contrast. Importantly, similarity testing and thus assimilation become more likely if the standard is perceived as moderate or attainable compared to extreme or unattainable (Lockwood & Kunda, 1997, 2000; Mussweiler, 2003). For example, assimilation with another more successful athlete is more likely if that athlete is a nearby neighbor or a frequent training partner than if that athlete is Usain Bolt—an Olympic sprinter. However, the comparison outcome of similarity testing serves first and foremost the purpose of solidifying self-perceptions (e.g. "how athletic am I in relation to X", Strack & Mussweiler, 1997) but does not necessarily translate into motivational outcomes concerning personally endorsed goals. While the motivational effects of the different upward modes of upward comparisons may be well inferred from the cognitive processes, it is less apparent for downward comparisons. For instance, our target athlete might be motivated by an attainable upward standard (by means of *assimilation*), but she might experience a decrease in motivation when comparing to the *unattainable* success of another athlete (*contrast*). On the other hand, consider

Norah, who compares her professional skills to the company's young intern (i.e. an extreme downward comparison). She will most likely experience a contrast effect when comparing her professionalism to the intern's greenness; but, in terms of self-improvement motivation, she will most likely *not* benefit from the comparison that deviates too much from her actual goal of a progressing career. The motivational outcomes are even less clear when considering comparisons to a moderately downward standard. Past research has shown that the comparison to an inferior but still somewhat similar standard can be perceived as a threat to the self and in turn, can increase prevention motivation in order to avoid a negative future self (Lockwood et al., 2002; Shakya et al., 2015). However, past research has also revealed that downward comparison is less likely to boost self-improvement motivation compared to upward comparison (Lockwood et al., 2002). For instance, only healthier but not unhealthier comparison standards pushed people towards an improved lifestyle (Lockwood et al., 2005). In sum, judgement of assimilation and contrast regarding a standard of reference clearly serve to shape self-perceptions, but these patterns of comparison do not consistently translate into motivational self-improvement outcomes. So to better understand the motivational nature of different upward and downward comparisons, it may be worthwhile to unite the literatures on social comparison processes and self-regulation.

1.3 A motivational framework

The merging of motivational principles from self-regulation research with social comparison processes should shed light on how upward or downward standards bias affective outcomes (i.e. motivation and emotion) in addition to pure comparative judgments of similarity and contrast. In this process, the three motives of social comparison are treated as dynamic and functional states that can serve but also temporarily hinder goal pursuit. In reference to Carver and Scheier (1982) who have already touched on this connection before, social comparison may work like the use of personal standards as references to guide self-regulatory behavior. According to their model, the process of self-regulation in the pursuit of specific goals works as a feedback loop where an individual compares her current standing to a to-be-attained standard and then assesses the discrepancy between both (Carver & Scheier, 1981, 1982). Depending on whether the discrepancy appears to be positive (the current standing overshoots the reference standard) or negative (the current standing undershoots the

reference standard), predictable actions then follow. With the integration of comparison processes, each social comparison motive can mirror different steps from the self-regulation feedback loop. First of all, self-evaluation accompanied by lateral comparisons serve the goal of gathering information about a person's present state and abilities. In the presence of a comparison standard, the current state is evaluated relative to the standard with the outcome of possible discrepancy assessments (stage: *self-assessment*). One possible outcome is the detection of a negative discrepancy, where the individual is inferior to the standard. A negative discrepancy is often associated with the motive of self-improvement and hence, the aim to reduce the discrepancy (i.e. the investment of effort; *pushing*; Fulford et al., 2010). However, an alternative scenario is that a standard overshoots the individual by a large degree. When a discrepancy becomes too large, which makes the standard seem out of reach, the individual may experience a drop in motivation (i.e. a reduction of effort; *disengagement*). From an emotional perspective, negative discrepancies first of all lead to feelings of ambivalence (Carver & Scheier, 1981, 1990, 2011) as the person is inspired and hopeful on the one side ("I can also reach the outcome."), but doubtful and uncertain on the other side ("Can I attain the same outcome?"). Positive affect, however, becomes increasingly unlikely with more extreme upward standards. Hence, once the motivational peak of upward comparison is reached, a negative discrepancy results in an unreachable standard perception and with accompanied negative affect and a threat to self-esteem (Carver & Scheier, 1990, 2011; Morse & Gergen, 1970). A third outcome is the detection of a positive discrepancy, where the individual is superior to the standard. The associated goal of superiority is to improve or maintain self-worth. A positive discrepancy triggers positive feelings, such as happiness and a boost in self-esteem (Carver & Scheier, 1990, 2011). With regard to self-regulation, motivation and effort investment are temporarily set to pause (coasting; Carver & Scheier, 1990, 2011). That means, according to the present model, even a moderate downward comparison would not increase improvement motivation but on the contrary would stifle further effort investment as a person feels that her relative standing is sufficient and thus "safe". The relationship between comparison direction and motivation in terms of effort investment is displayed in Figure 1 as a non-linear pattern: Effort investment increases with upward standards, reaches its maximum with moderate upward standards, before it continues to decrease again with more extreme upward standards (Brehm & Self, 1989;

Carver & Scheier, 1981, 2011; Klinger, 1975). With moderate upward standards, the individual feels inspired and encouraged (self-improvement or *pushing* motivation) until the standard becomes increasingly unattainable and motivation drops ("I better give up", *disengagement*). On the other side, effort decreases with downward standards where the individual feels increasingly better about their superiority in emotional terms (*self-enhancement*) but also in motivational terms ("I am doing just fine", *coasting*).

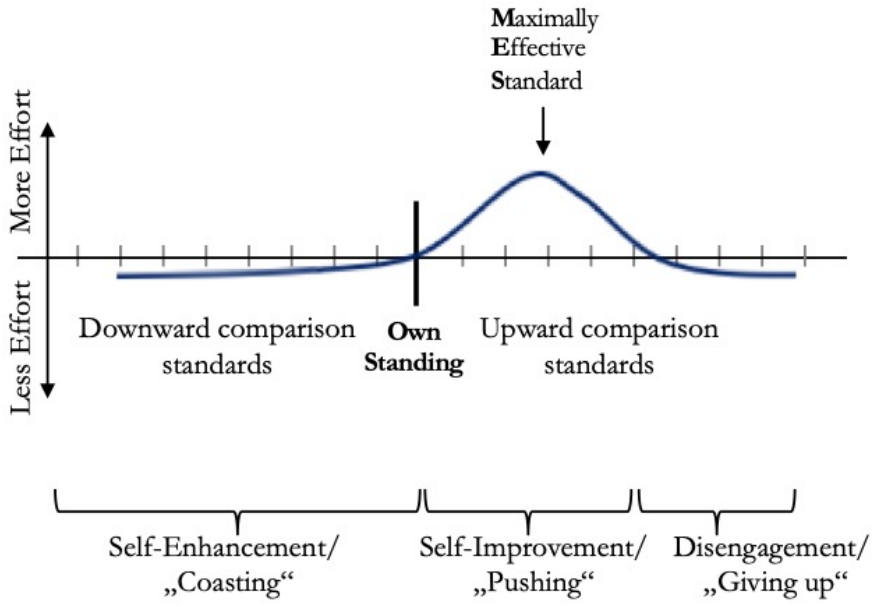


Figure 1. Hypothesized mapping of discrepancy assessment between own standing and relatively better or worse social comparison standards, change in effort investment (curve), and key motivational states of self-improvement ("pushing"), self-enhancement ("coasting"), and disengagement ("giving up").

In sum, on their way to goal pursuit, people not only compare their current standing to their ultimate goal-state, but also seek feedback from their social environment. People's motivation does not steadily increases until a goal is reached, but rather fluctuates from one day to another (Milyavskaya & Werner, 2018). Hence, social comparison motives adapt dynamically to the current states and thereby guide self-regulation and goal pursuit. Certainly, a person not only draws motivation from her social environment, but may also be discouraged at times or motivated to rest temporarily. In this way, however, a person can decide for herself at that moment what goals she wants to pursue and with how much effort to invest. This framework gives new insights into how people track their long-term goals by help of their social environment. Simultaneously, the framework

spells out under what circumstances people may benefit from social comparisons either in a motivational or emotional sense and may imply which social comparison standards a person can safely orient to and which should better be avoided due to their detrimental effects.

1.4 Social Comparison In Morality

People are biased towards upward comparison and should turn toward (moderate) superior others when striving towards self-improvement. There are, of course, exceptions to the rule and domains exist where people more strongly benefit from downward comparisons, namely the domain of morality. For instance, besides her career, Norah wants to live an environmentally friendly life, but sometimes fails to act according to her moral standards (e.g. by being a frequent flyer). In other words, people consider themselves as moral individuals but still frequently fail to act upon their moral views (e.g. failing to resist environmentally unfriendly purchases; Nielsen & Hofmann, 2021). According to the above motivational framework, people would then turn towards comparison standards that already display the desired behaviour (i.e. more moral people) as a source of inspiration. However, when making moral comparisons, people are motivated to present themselves in the best possible light that helps them to look moral (Hauke & Abele, 2020; Monin, 2007; Ybarra et al., 2012) and hence, may show a general tendency to avoid upward comparisons altogether and rather draw moral affirmation from downward comparison.

One reason why morality is so distinct is that morality builds a crucial part of people's identities and is unique to human beings (Goodwin et al., 2013; Monin, 2007; Strohminger & Nichols, 2014). Therefore, threat (e.g. the comparison to more moral standards) is perceived as more severe than in other domains and in turn, people are motivated to prevent moral upward comparisons. To follow my example of environmental sustainability, when eating meat, people evaluate vegetarians more negatively if those vegetarians frame their decisions to not eat meat morally (Cramwinckel et al., 2013). Hence, instead of upward comparisons, moral comparisons are characterized by a bias towards downward comparison, especially after threat to the self. That threat triggers downward comparison is shown by previous research. For instance, students who performed poorly on an exam, avoided comparisons to more successful others and rather compared to even less successful others in order to protect their academic self-concept (Pyszczynski et al., 1985). Thus, the motive of self-enhancement

might be more pronounced among moral comparisons compared to other domains as people are constantly motivated to maintain a positive self-image and appear a morally good person.

Furthermore, moral comparisons in contrast to other comparisons are less sensitive to whether a standard is perceived as similar or dissimilar: Within moral domains, even an extreme (i.e. dissimilar) downward standard is perceived as relevant and appropriate for self-enhancement purposes. For example, whereas a very unathletic person is unlikely to be a relevant standard of comparison for an ambitious athlete, on the contrary, in the moral sphere an extremely negative exemplar can be considered a suitable comparison to fulfil the motive of self-enhancement. Similarly, although extreme upward comparisons that are usually considered as less relevant and diagnostic, and thus lead people to disengage from those comparisons, an extremely *moral* exemplar can still pose a threat. Since people assume that their own moral standards are correct (Graham et al., 2009, 2013; Haidt et al., 2009; Haidt & Graham, 2007; Haidt & Joseph, 2008; Koleva et al., 2012) and upward standards pose a threat to their own beliefs, people try to prevent upward comparisons altogether in the moral domain. In addition, people protect their own moral values, which are non-negotiable with other values people may hold and that is why people respond particularly negative towards other people's values that appear to be better (Atran et al., 2007; Tetlock, 2003; Tetlock et al., 2000). Thus, even upward standards that are distant rather than close, and extreme rather than moderate and hence less relevant to the self (Corcoran et al., 2011; Festinger, 1954; Lockwood & Kunda, 1997c; Major et al., 1991; Tesser, 1988, 1991) are still posing a threat to one's moral self and are to be avoided. In sum, in the domain of morality people prefer downward comparison for moral affirmations. Upward comparisons, on the other hand, pose a threat to the moral self, even if the example is rather extreme.

1.5 The Current Research

In the current research, I propose a new framework that integrates research from both social comparison and self-regulation to provide a better understanding of how people orient themselves to their social environment in order to meet their long-term goals. This involves people's motivations but also emotions. This framework will also answer which comparative strategies can positively support the achievement of goals and which – in a motivational sense – have the opposite effect. Linking self-regulation research with social comparison processes closes a gap in the literature, as past social

comparison research has tended to focus on cognitive processes for self-perception purposes and has only provided limited answers with regard to motivational effects. In Chapter 2, I introduce a new motivational framework paradigm that connects classic similarity assessments and motivational consequences. I generated specific predictions for moderate (vs. extreme) upward (vs. downward) standards' motivational impact in a context of exercising goals. In line with the selective-accessibility model (Mussweiler, 2003), I show evidence of assimilation and contrast in judgmental terms as participant feel similar to moderate compared to extreme standards. In motivational terms, however, I demonstrate that the predictions for moderate vs. extreme standards deviate from the classical social-cognitive model of assimilation and contrast. To give an example, perceived dissimilarity to an extremely downward standard (i.e. contrast effect) does not result in a motivational boost (i.e. increase in goal pursuit) but rather in a clear drop in motivation.

In Chapter 3, the motivational paradigm has been extended and closely aligned with self-regulation research. Furthermore, the framework is tested in an ecologically valid environment by the conduct of an experience-sampling study. The study had a broad focus by including numerous domains of comparison and a heterogeneous sample. Depending on comparison directions (i.e. measuring the valence of target and standard discrepancies on a continuum), the study captured motivational (pushing, coasting, disengagement) as well as emotional correlates (e.g. pride and guilt). I expected pushing motivation to increase with negative discrepancies, but also to decrease when discrepancies become too large. Simultaneously, disengagement was expected to rise with growing negative discrepancies. On the other hand, I predicted coasting effects for positive discrepancies.

In Chapter 4 the motivational framework was tested in a sample of athletes, which may be especially prone towards self-improvement and thus, upward comparison. In addition to the previous chapter, Chapter 4 adds a behavioural effort component consisting of a real-life performance measure. This way, I was able to test the effects of different social comparison and discrepancy assessments on people's actual performance outcomes relating self-improvement. In addition to the main hypotheses concerning pushing, coasting and disengagement as well as emotions, I hypothesized that upward comparison is related to higher behavioural self-improvement scores.

Lastly, in Chapter 5, I deviate into another domain where the classical judgment paradigm does not hold: in the domain of morality where people actually turn towards downward comparisons. Here, people are generally threatened by upward comparison and draw moral affirmation from downward comparison.

Chapters 2 to 5 are based on published manuscripts presenting their own introductions and discussions and thus overlaps will occur. In Chapter 6, I present an overall summery of findings across the main chapters, discuss general implications, research limitations, ideas for future research and finally, a conclusion.

Chapter II: Inspired to Perspire: The Predicted Interplay of Social Comparison Direction and Standard Extremity in the Context of Challenging Exercising Goals

This chapter is based on the following publication:

Diel, K., & Hofmann, W. (2019). Inspired to perspire: The predicted interplay of social comparison direction and standard extremity in the context of challenging exercising goals. *37*(3), 247-265, *Social Cognition*. http://dx.doi.org/10.1521/soco.2019.37.3.247 (Published online: June 2019)

Please note that changes in headings, citation style, and formatting were undertaken to fit the layout of this dissertation. No changes were made to the content of the article.

Abstract

People pursue challenging health goals, often with uncertain success. In this article, we combined a social-cognitive framework of social comparison with motivational principles in self-regulation to investigate how salient social comparison standards impact motivation toward exercising goals. Across three studies, we systematically expose participants to one of four comparison standards varying in comparison direction (upward vs. downward) and standard extremity (moderate vs. extreme). We consistently find that moderately upward standards result in a motivational boost and that extreme downward comparison standards have no motivational potential. However, there was also some evidence, although less consistently, for a motivating effect of moderate downward standards (Studies 1 and 2), suggesting a preventive or competitive demarcation effect, and for extreme upward standards, suggesting overconfidence (Study 3). From a theoretical perspective, our findings suggest that contemporary information-processing models of social comparison (assimilation and contrast), may not be fully adequate to also cover the motivational consequences.

Keywords: social comparison, motivation, self-regulation, fitness behavior

2.1 Introduction

People often want to improve their health and fitness, but the pursuit of reaching health goals is anything but easy. Self-improvement often fails due to a lack of self-control capacity and/or motivation. Whereas these factors point to internal aspects of motivation, a largely open question is how the social environment can shape such processes. In this article, we investigate how social comparison impacts people's self-control and motivation in the pursuit of exercising goals, and how underlying social cognitive processes may explain how different social comparison standards can push or instead hinder the pursuit of those health goals.

Social comparison is ubiquitous in social life. People use information about other people to receive feedback regarding their own abilities and actions. In order to gain an accurate self-view, people turn to similar others as social comparison standards (Festinger, 1954). Moreover, the comparison with superior others (upward comparison) may inspire self-improvement and therefore also positively influence motivation to change behavior (Aspinwall et al., 2002; Collins, 1996; Lockwood & Kunda, 1997, 2000; Markman & McMullen, 2003; Taylor & Lobel, 1989). In the context of health, feeling inferior to others positively influenced dieting intentions (Shakya et al., 2015), motivation and intentions to eat healthy and to exercise (Lockwood et al., 2005), and motivation to increase walking behavior, and even raised the actual number of steps taken (Chapman et al., 2016). Furthermore, in combination with a sun-protection intervention, upward comparison was found to increase sun protection behavior (Mahler et al., 2010). Social comparison processes and upward comparisons are also associated with increased self-regulation (Taylor et al., 1996) and self-control. For instance, present others who engage in prudent behavior increased the likelihood that someone else also displays more controlled behaviors (Polivy et al., 1979; Rosenthal & McSweeney, 1979).

In contrast to upward comparison, downward comparison is less likely to motivate self-improvement (Lockwood et al., 2002), as people were more motivated by healthy than by unhealthy exemplars (Lockwood et al., 2005). Moreover, if a target discovers a positive discrepancy to an inferior target and therefore performs above expectations, a "coasting" pattern may result. Such coasting behavior is characterized by temporarily withholding effort investment toward the goal and

by shifting towards other preferences (Carver & Scheier, 1990). In a health context, downward comparisons paired with a sun-protection intervention undid the positive effects of the intervention by decreasing actual sun protection behavior compared to only the intervention or intervention in combination with upward comparisons (Mahler et al., 2010). However, some research has also found that downward comparison can have a positive influence on health behavior. For instance, comparing to friends who are less fit increased the likelihood to exercise (Shakya et al., 2015) and inferior negative role models motivated individuals to avoid a more negative alternative self (Lockwood et al., 2002). Hence, the specific effects of downward comparison are not particularly well understood.

Moreover, it is not only the direction of comparison may determine whether a comparison standard positively (or negatively) influences health goals. One way to conceptualize the likely motivational outcome in light of past research is to distinguish between moderate and extreme comparison standards and, similarly, whether a given standard is attainable rather than unattainable. For instance, comparing to a regular and persistent runner from your neighborhood can motivate your weekly running attempts. However, comparing to a person who runs a few marathons a year might be rather discouraging. On the other hand, a total couch potato may not arouse your competitive spirit to stay ahead of the curve, but maybe the person who runs just a bit slower than you does. To allow for clearer predictions, we think it may be useful to connect social-cognitive accounts of how people process social comparison information from their environment with motivational principles relevant for the pursuit of challenging health goals, and to jointly manipulate the direction of comparison as well as the extremity of the comparison standard.

Prominent social-cognitive accounts of social comparison spell out how features of the social comparison standard shape information processing and resulting judgments. For instance, according to the selective accessibility model (SAM; Mussweiler, 2003), if a target (i.e., perceiver) is confronted with a salient comparison standard, target knowledge can be activated that is either consistent with the standard (similarity testing), resulting in assimilation, or inconsistent with the standard (dissimilarity testing), resulting in contrast. One of the key decisive factors of whether the perceiver engages in similarity or dissimilarity testing is the extremity and attainability of the standard (Lockwood & Kunda, 1997; Mussweiler, 2003). If a standard is moderate compared to extreme on one dimension,

then similarity-testing (assimilation) is more likely to occur. The same is true if a standard is perceived as attainable. On the other hand, if a standard is extreme or considered unattainable, one will more likely engage in dissimilarity testing and contrasts away from the standard.

As we can see, existing models of social comparison provide guidance regarding what effects may be expected for varying combinations of comparison standard direction and extremity. Their limitation, however, is that such information-processing models are typically geared towards judgmental effects (e.g. "What is the height of the Cathedral of Cologne"; Strack & Mussweiler, 1997). They typically do not spell out how the result of such a similarity assessment would impact actual *motivation* to invest effort towards a personally endorsed goal. Of course, social comparison has not been totally mute with regard to motivational principles; however, the field has mostly focused on broader social comparison motives, such as self-assessment, self-improvement, and self-enhancement (Gerber et al., 2017; Taylor & Lobel, 1989; Wills, 1991), rather than concrete goal pursuit. A closer consideration of the interplay of direction of comparison and standard extremity on goal pursuit may therefore be warranted.

2.1.1 Framework.

In our present research, we systematically distinguished between upward and downward comparison, as represented by physically fit (upward) versus unfit (downward) comparison standards, and moderate versus extreme exemplars in both directions. Assuming that those comparison standards that are evoked by acts of social comparison may serve "as reference values for self-regulation" in a similar way as personal standards (Carver & Scheier, 1982, p. 151), we derived the following hypotheses and predictions regarding their motivational effects, as summarized in Table 1.

Table 1

Interplay of comparison direction and extremity on perceived similarity and motivation

		Standard Extremity	
		Moderate	**Extreme**
Comparison Direction	**Upward (fit)**	Similarity=high:	Similarity=low:
		Standard as inspiration	Standard unattainable
		Motivation: boost (promotion)	Motivation: no boost or drop
	Downward (unfit)	Similarity=high:	Similarity=low
		Standard as threat;	Standard clearly surpassed;
		Opportunity for demarcation	No need for further demarcation
		Motivation: boost (prevention)	Motivation: drop (coasting)

In the case of the *moderately fit standard*, we predicted high perceived similarity (compared to the extremely fit standard) and also a boost in motivation, since we expect the standard to be inspirational and attainable at the same time (Lockwood & Kunda, 1997, 1999, 2000). In contrast, we hypothesized that the *extremely fit standard* is expected to result in lower perceived similarity (compared to the moderately fit standard) and also to lower motivation as it might be inspiring but not attainable because the upward discrepancy is too large (Carver & Scheier, 1990). In case of the *extremely unfit standard*, we predicted that similarity is perceived to be low due to the extreme deviation from the (fitness) goal. Hence, we expected a clear demarcation from this standard, most likely resulting in a coasting pattern (and a corresponding drop in motivation). Lastly, in case of the *moderately unfit standard*, we expected high perceived similarity (compared to the extremely unfit standard). Because the standard is perceived as similar but may also be perceived as a threat (Lockwood et al., 2002), we might expect positive effects on motivation, rather than coasting behavior. For instance, the person may be viewed as a competitor. Therefore, the target may aim to positively distinguish him - or herself from the standard so as to avoid a negative outcome in the future (Crowe & Higgins, 1997). In other

words, whereas moderately fit standards may elicit a promotion focus, moderately unfit standards may motivate people to prevent themselves from ending up in the same category (Higgins, 1998); such a motivational effect would be consistent with other research finding positive effects of downward comparisons on motivation (Shakya et al., 2015).

In sum, we predict that upward comparison may not necessarily benefit goal striving if the standard is too extreme and therefore "out of reach". Likewise, extreme downward comparisons may decrease motivation due to the principle of coasting. The most straightforward motivational boost effect should be expected in cases where people are confronted with a moderately upward standard exemplifying the goal, thus serving as a positive role model. However, we also predicted that a motivational boost may also result from the comparison with moderate downward standards. Note that this pattern of predictions is different from a simple application of the logic of assimilation and contrast for judgmental research with regard to the two downward comparison cells, which would predict assimilation to moderate standards (i.e., reduction in goal pursuit) and contrast away from extreme standards (i.e., an increase in goal pursuit).

2.1.2 The Present Research

In three studies, we implemented this design into the context of exercising goals and paired participants with one of four social comparison standards. We manipulated both the direction of comparison standard (fit/upward vs. unfit/downward), and standard extremity (moderate vs. extreme), as well as measured participants' perceived similarity to a randomly assigned standard and its influences on motivation. This design allowed us to closely trace the motivational effects of a specific standard in connection to a goal.

2.1.3 Statement of Transparency

We conducted four studies in total, from which two were preregistered. One preregistered (Study 3) and two non-preregistered studies (Studies 1 and 2) are reported in this paper. We decided to substitute the former Study 3 with a new Study 3 that was methodologically improved. For instance, we improved the implementation of comparison standards: moderate (vs. extreme) upward (vs. downward) standards were tailored around the own participant's standing and not around a general

average. In addition, dependent measures, such as motivation were worded in a less speculative ("This person would motivate me…"), but more concrete ("This person motivates me…".) manner. Further, in the current studies we included additional variables that were not part of the main analyses, but were assessed for exploratory purposes. All study materials, measurements and preregistrations, as well as results of the former Study 3, are available on Open Science Framework (OSF) (project link: https://osf.io/79cqr/?view_only=95d802e65b2348fe9d0f7fellea7937c.)

2.2 Study 1

The aim of Study 1 was to test how exposure to social comparison standards varying in comparison direction and standard extremity affects perceived similarity towards these standards and the motivation towards exercising. We manipulated the comparison *direction* of the standard (fit/upward vs. unfit/downward) and *extremity* of the standard (moderate vs. extreme), and then measured perceived similarity and motivation in a fitness improvement context.

2.2.1 Method

Participants

The sample consisted of 304 Mechanical Turk (Mturk) users from the United States (M_{age} = 33.83, *SD* = 10.67, 44.4% female). A power analysis with G*Power (Faul et al., 2007) indicated that we needed at least 230 to detect a moderate effect using an alpha of .05 and with 90% power. No participants were excluded. All participants were compensated with $0.80 for participation. Informed consent was obtained from all participants.

Design

The experiment included a 2 (direction: fit vs. unfit) × 2 (extremity: moderate vs. extreme) between-subjects factorial design. As the main outcome variables, we measured perceived similarity to the standard and motivation towards exercising.

Materials and procedure

As a cover story, we introduced an ostensible new fitness app called “Stay fit – together”, which builds on the idea that an online app partner (comparison standard) helps to reach the goal of becoming fit, and participants had to evaluate potential workout partners.

Comparison standards. We used vignettes of ostensible former participants who tell about their fitness level and self-discipline regarding exercising (gender matched). We manipulated direction and extremity of the comparison standard: the standard was either fit (e.g. "I work out on a regular basis") or unfit (e.g. "I do not exercise."). Secondly, their fitness level was either displayed to a moderate extent (e.g. "I work out one or two days a week" (fit) or "I work out every few months" (unfit)) or extreme extent (e.g. "I work out every day" (fit) or "I have never done any exercising" (unfit); see the supplementary materials for the full vignettes). As a manipulation check, we included an item about the perceived fitness level of the assigned partner on a 5-point scale (1 = *very below average*, 5 = *very above average*).[1]

Perceived similarity. Participants indicated how similar they felt to the person and how much they liked the person via five items on a 7-point scale (e.g., "The person and I have things in common", 1 = *Strongly disagree*, 7 = *Strongly agree*). Because of the high internal consistency among all five items (Cronbach‘s $\alpha = .87$) and correlation between the mean value of similarity and the mean value of liking ($r = .63$, $p < .001$), we combined both measures into a single index of perceived similarity (for a list of all items, see the supplementary materials).

Motivation. Participants were instructed to imagine taking part in the app fitness program with their assigned partner and, based on this scenario, rate their own self-control expectancies and motivation for the first weeks of participation. We measured self-control expectations via eight items ($\alpha = .89$), which were adapted from the original brief self-control scale (Tangney et al., 2004) and answered on a 5-point scale (e.g. "I will be self-controlled", 1 = *not at all*, 5 = *very much*) (for a list of all items, see the supplementary materials). In addition, we measured motivation on a single item (“How motivated would you be to participate in regular physical activity over the next four weeks”, 1

[1] Manipulation check results can be found in the supplementary materials at https://osf.io/79cqr/?view_only=95d802e65b2348fe9d0f7fellea7937c

= *extremely unmotivated*, 5 = *extremely motivated*). Because of the high correlation between the measure of self-control expectations and motivation ($r = .65$, $p < .001$) and high internal consistency among all items ($\alpha = .91$), we combined them to a composite motivation score.

In the end, participants indicated their own fitness level (1 = *very below average*, 5 = *very above average*) and filled in demographic information.

2.2.2 Results

Perceived similarity. We hypothesized that participants feel more similar to moderate compared to extreme standards. As predicted, a two-way ANOVA analysis revealed a main effect of extremity on perceived similarity to the standard, $F(1,300) = 50.14$, $p < .001$, $\eta_p 2 = .143$. Participants felt more similar to moderate ($M = 4.81$, $SD = 0.93$) over extreme standards ($M = 3.96$, $SD = 1.18$). The main effect of comparison direction was non-significant, $F(1,300) = 0.77$, $p = .381$, $\eta_p 2 = .003$. Results also demonstrated an interaction effect of direction and extremity on similarity, $F(1,300) = 5.86$, $p = .016$, $\eta_p 2 = .019$. The effect of extremity was larger for fit standards ($M_{moderate} = 5.01$, $SD = 0.78$ vs. $M_{extreme} = 3.86$, $SD = 1.08$, $p < .001$, $d = 1.22$, 95% CI [0.81, 1.49]) compared to unfit standards ($M_{moderate} = 4.61$, $SD = 1.02$ vs. $M_{extreme} = 4.05$, $SD = 1.27$; $p = .001$, $d = 0.49$, 95% CI [0.23, 0.90]). Post-hoc test determined that participants felt more similar to the moderately unfit compared to the extremely unfit standard ($M_{moderate} = 4.61$, $SD = 1.02$ vs. $M_{extreme} = 4.05$, $SD = 1.27$, $p = .001$, $d = 0.49$, 95 % CI [-0.90, -0.23]). Similarly, participants felt more similar to the moderately fit compared to the extremely fit standard ($M_{moderate} = 5.01$, $SD = 0.77$ vs. $M_{extreme} = 3.86$, $SD = 1.08$, $p < .001$, $d = 1.23$, 95 % CI [0.81, 1.49]). Moreover, similarity was higher for the moderately fit compared to the moderately unfit standard ($M_{fit} = 5.01$, $SD = 0.77$ vs. $M_{unfit} = 4.61$, $SD = 1.02$, $p = .021$, $d = 0.44$, 95 % CI [0.06, 0.74]).

Motivation. We conducted a two-way ANCOVA to test the effects of comparison direction and extremity on participants' motivation (composite motivation score). We controlled for participants' self-reported fitness level in order to isolate the contribution of comparison standards to motivation above participants' own fitness level. As predicted, results revealed a main effect of extremity, $F(1,299) = 4.64$, $p = .032$, $\eta_p 2 = .015$. Motivation was higher for moderate ($M = 3.67$, $SD = 0.74$) compared to extreme standards ($M = 3.44$, $SD = 0.83$). The main effect of direction was not

significant, $F(1,299) = 0.05$, $p = .828$, $\eta_p 2 < .001$, nor was the interaction effect of direction and extremity, $F(1,299) = 0.06$, $p = .807$, $\eta_p 2 < .001$.[2] (See Table 2.)

Table 2

Means, Standard Deviations and Cell Sizes for the Effects of Comparison Standard Direction and Extremity on Perceived Similarity and Motivation

		Study 1			Study 2			Study 3		
Variable	**Condition**	***N***	**Mean**	**SD**	***N***	**Mean**	**SD**	***N***	**Mean**	**SD**
Similarity	Extreme Downward	77	4.05_a	1.27	85	3.47_a	1.12	182	4.79_a	1.04
	Moderate Downward	76	4.61_b	1.02	84	4.62_b	0.98	188	5.25_b	0.84
	Moderate Upward	74	5.01_b	0.78	84	4.45_b	0.82	181	5.17_b	0.94
	Extreme Upward	77	3.86_a	1.08	85	3.93_c	1.19	178	4.71_a	0.99
	Control							182	4.79_a	0.92
Motivation	Extreme Downward	77	3.50_a	0.90	85	3.05_a	0.78	182	4.07_a	1.58
	Moderate Downward	76	3.67_a	0.82	84	3.30_a	0.68	188	4.40_a	1.29
	Moderate Upward	74	3.67_a	0.64	84	3.32_a	0.57	181	4.85_b	1.51
	Extreme Upward	77	3.39_a	0.76	85	3.28_a	0.71	178	4.96_b	1.67
	Control							182	4.05_a	1.47

Note: Means with different subscripts differ significantly from each other using Tukey post-hoc tests.

2.2.3 Discussion

As expected, results indicate that people indeed felt more similar to moderate over extreme standards. In addition, we found evidence for a motivational effect of moderate versus extreme comparison standards: Regarding upward and downward comparisons, participants showed higher

[2] Also without the covariate, the main effect of extremity on self-control expectations remains significant, $F(1,300) = 6.36$, $p = .012$, $\eta_p^2 = .021$. Self-reported fitness level was not significantly influenced by the manipulation ($p = .394$)

motivation after comparing to a moderately fit or unfit standard as compared to an extremely fit or unfit standard. In Study 2, we sought to conceptually replicate these findings in the laboratory using an improved manipulation of comparison standards. Instead of vignettes, we gave comparison standards fitness scores that better determined whether they are above expectations (fit/upward standard) or below expectations (unfit/downward comparison).

2.3 Study 2

Study 2 aimed to replicate the findings of Study 1 with a more precise manipulation of comparison standards instead of vignettes. We then again measured perceived similarity and motivation.

2.3.1 Method

Participants

Our sample consisted of 338 German students from the University of Cologne (M_{age} = 23.33, *SD* = 4.98, 79.3% female). A power analysis with G*Power (Faul et al., 2007) indicated that we needed at least 230 to detect a moderate effect using an alpha of .05 and with 90% power. No participants were excluded. Participants were invited into our laboratory to first participate in an unrelated study and then complete this study as a second part that took approximately 10 minutes. Participants received about 2 Euros (approx. $2.15) as compensation. Informed consent was obtained from all participants.

Materials and procedure

The experimental design and the cover story were identical to Study 1. However, instead of vignettes, we manipulated comparison standards by providing information about the standards' past fitness behavior. Comparison standards were introduced as ostensible former app subscribers who participated in an app trail run, where users had to complete a particular workout. This workout was tailored to each of their fitness levels and time schedules. In addition, participants learned that the goal of the former users was to complete a certain workout five times in their first week. In our manipulation, unfit (downward) standards did not achieve the goal of the week by completing less than five workouts. In contrast, fit (upward) standards completed more than five workouts. Beyond

that, each comparison standard was given a self-discipline score (1 = *none at all*, 10 = *very much*) that correlated with the number of completed workouts (see the supplementary materials for all comparison standards description). Again, the standards were matched in gender and the ostensible former users were introduced as students from the University of Cologne. This manipulation of comparison standards was used and successfully pretested in an unrelated study and thus, we did not include a manipulation check for perceived fitness level of the standard. As in Study 1, participants were randomly exposed to one of the four standards and subsequently completed measures of perceived similarity (α = .86) and motivation (α =.87). Lastly, participants reported their own fitness level and demographic information.

2.3.2 Results

Similarity. We expected that participants would feel more similar to moderate compared to extreme standards. As in Study 1, two-way ANOVA results revealed a main effect of extremity on perceived similarity. Participants felt more similar to moderate standards (M = 4.14, SD = 0.90) compared to extreme standards (M = 3.30, SD = 1.18), $F(1,334) = 54.82$, $p < .001$, $\eta_p2 = .141$. The main effect of direction was non-significant, $F(1,334) = 1.68$, $p = .196$, $\eta_p2 = .005$. Results revealed a significant interaction effect of direction and extremity on perceived similarity, $F(1,334) = 7.76$, $p = .006$, $\eta_p2 = .023$. The difference in perceived similarity was less pronounced for fit (upward) comparison standards ($M_{moderate}$ = 4.45, SD = 0.82 vs. $M_{extreme}$ = 3.93, SD = 1.20, $p = .001$, d = 0.51, 95% CI [0.21, 0.84]) compared to unfit (downward) comparison standards ($M_{moderate}$ = 4.62, SD = 0.98 vs. $M_{extreme}$ = 3.47, SD = 1.12, $p < .001$, d = 1.09, 95% CI [0.84, 1.46]).

Motivation. We predicted that participants show more motivation when comparing to a moderate compared to an extreme standard. In line with Study 1, two-way ANCOVA results revealed a main effect of extremity, while controlling for self-reported fitness level. Participants who compared to a moderate standard showed higher motivation (M = 3.31, SD = 0.63) than participants comparing to an extreme standard (M = 3.16, SD = 0.75), $F(1,333) = 5.17$, $p = .024$, $\eta_p2 = .015$. There was no

significant main effect of direction, $F(1,333) = 2.31$, $p = .129$, $\eta_p 2 = .007$, and no significant interaction effect, $F(1,333) = 2.68$, $p = .103$, $\eta_p 2 = .008$.[3] (See Table 2.)

2.3.3 Discussion

In line with Study 1, participants indicated higher perceived similarity for moderate standards compared to extreme standards. Also, motivation was increased when comparing to a moderate standard versus an extreme standard. The motivational advantage of moderate versus extreme standards was more pronounced with regard to downward comparison (i.e., unfit standards), in accordance with the relatively stronger perceived similarity difference between the extremely and moderately unfit standard. Together, both studies so far suggest that when it comes to the motivational implications of moderate and extreme standards, there is neither strong evidence for a classical pattern of assimilation and contrast as typically obtained in the social judgement literature nor strong evidence that these effects would be symmetric for upward and downward comparison directions. We will come back to discussing the larger implications of this insight for motivation and health research in the general discussion.

One limitation of Studies 1 and 2 is that comparison standards are not necessarily different from the participant in the intended direction. In Study 3, we first assess participants' fitness level and then tailor comparison standards around their individual standings. Another limitation of Studies 1 and 2 is that no inferences can be made yet about whether comparison information actually increases (or decreases) motivation compared to no availability of social comparison information. Hence, in Study 3 we added a control group (no social comparison information). Finally, we preregistered our design and analysis plan. Furthermore, in Study 3 we added a behavioral choice measure in order to test whether certain comparison standards can positively influence future health choices.

[3] Without the covariate, the main effect was suggestive, but did not meet our criterion for significance, $F(1,333) = 3.56$, $p = .060$, $\eta_p^2 = .006$. Self-reported fitness level was not significantly influenced by the manipulation ($p > .050$).

2.4 Study 3

In Study 3, we again modified the paradigm of Study 2 by providing participants with a personal fitness score based on a fitness questionnaire. One important innovation in this study was that we then confronted participants with a comparison standard that was customized around their own individual score. This way, we were able to create comparison standards that represented upward or downward standards tailored to each individual fitness score. Furthermore, we added a measure of behavioral choice. In contrast to Studies 1 and 2, the paradigm and adjusted measures of Study 3 are more relevant to participants and, therefore, consist less of hypothetical situations.

2.4.1 Method

Participants

Our sample consisted of 917 Mturk users from the United States (M_{age} = 37.64, *SD* = 11.34, 54.0% female). A prior power analysis using G*Power (Faul et al., 2007) indicated that we needed at least 618 participants to detect a moderate effect using an alpha of 0.05 and with 90% power. Because we would possibly exclude up to 25 % of participants who scored too high or too low on the fitness score as preregistered, we rounded up to 800 participants and in the end received more responses than intended.[4] Due to technical problems with the interface between our survey and the Mturk Hit completion, 90 participants completed the survey twice, so we used only their first response.

Materials and Procedure

At the beginning of the study, we again introduced the fitness app as a cover story. It was our aim to move away from a hypothetical scenario as in Studies 1 and 2 (e.g. "Imagine what it would be like to have this person as your long-term workout partner"), and instead give participants a relevant comparison standard to evaluate in the moment. In order to make the standard relevant, participants had to answer ten questions about their physical fitness (Wood, 2008) (see supplementary materials).

[4] We preregistered that we might exclude participants in an exploratory way who received a fitness score below 100 or above 150 (at both extremes) in order to test whether the effects of comparison direction and extremity on similarity, motivation, and behavior get stronger for participants with a fitness score around the average compared to participants that score at the extremes themselves. We additionally preregistered that we will only exclude those participants if the overall number of to-be-excluded participants does not exceed 25 %. However, the actual number of participants who scored lower than 100 and higher than 150 was 473 (51.6 % of the sample). Therefore, we included the whole sample in the main analyses.

An individual fitness score that could range from 50 to 200 points was calculated for each participant and displayed at the end of the fitness questionnaire. The aim of implementing the fitness questionnaire was to give participants a meaningful (upward or downward) comparison standard by adding or subtracting from their individual fitness score.

Comparison Standards. Next, participants were told to see an ostensible former participant who qualifies as the app partner and who either scored "a little higher/lower" on the fitness test (moderate upward/downward comparison) or "much higher/lower (...) by getting one of the highest/lowest score possible" (extreme upward/downward comparison). On the next page, participants read about the standard who was matched in gender and age. Additionally, the standard's fitness score was displayed, which was (moderately or extremely) higher or lower, depending on the condition. In the moderate (vs. extreme) conditions, 10 points (vs. 50 points) were added or subtracted from their fitness score. As a fifth condition, we added a not-relevant comparison information group that represented the control group. Here, the only information about the standard was their (matched) gender and age. Participants were told that the other person has not yet received a fitness score. Participants were randomly assigned to one of the five groups.

Perceived Similarity and Motivation. After reading the standard's information, participants indicated how similar they feel to the standard on three items ("This person is likeable", "The person and I have things in common", "The person and I would get along", 1 = *not at all*, 7 = *extremely*, α = .83) and how motivated they are by the partner on three items ("The person motivates me to improve my fitness level", "This person motivates me to engage in regular physical activity", "This person encourages me to engage in physical activity", α = .97).[5]

Book choice. Next, participants were informed to have the opportunity to take part in a lottery to win a book of their choice (to be realized after study completion). They could choose their favorite option from four books: two of the books were cookbooks (unhealthy choice, such as a cookbook about desserts) and the other two were fitness books (healthy fitness choice, such as a book about

[5] Compared to Studies 1 and 2, we measured motivation on three items. Instead of using hypothetical items, such as, "This partner would motivate me to engage in regular physical activity" (e.g. Study 1), we changed it to "This person motivates me to engage in physical activity."

exercising; adapted from (Forstmann et al., 2012). We predicted that participants comparing to a moderate (vs. extreme, and vs. control) standard, as well as participants who compare to an upward (vs. downward) standard, would be more likely to choose a fitness book over a cookbook,.

2.4.2 Results

Perceived similarity. In the preregistration, we predicted that participants would feel more similar to moderate than to extreme standards and also compared to the control group. Due to the non-nested nature of our 2 × 2 control group design, we conducted a series of planned contrasts. The first planned contrast analysis, comparing the average of both moderate standards to the average of both extreme standards, revealed that perceived similarity was higher for moderate standards compared to extreme standards, $F(1,912) = 42.82$, $p < .001$, $\eta_p2 = .045$, 95 % CI [0.64, 1.19]. The second contrast compared the average of both moderate fit standards to the control group. Results revealed that perceived similarity was higher for moderate standards compared to the control group, $F(1,912) = 15.55$, $p < .001$, $\eta_p2 = .017$, 95 % CI [0.17, 0.50].[6]

Motivation. We predicted higher motivation for moderate standards compared first to extreme standards and second to the control group. The first planned contrast analysis comparing the average of both moderate standards to the average of both extreme standards did not reveal a significant difference: contrary prediction, motivation was not higher for moderate over extreme standards, $F(1,911) = 0.74$, $p = .391$, $\eta_p2 = .001$, 95 % CI [-0.25, 0.63]. In line with our prediction, motivation was higher for both moderate standards compared to the control group, indicated by the second contrast, $F(1,911) = 16.49$, $p < .001$, $\eta_p2 = .018$, 95 % CI [0.28, 0.81]. Post-hoc analyses indicated that, compared to the control group, motivation was increased for the moderately upward standard ($p < .001$, $d = 0.54$, 95 % CI [0.38, 1.23]) and for the extremely upward standard ($p < .001$, $d = 0.58$, 95 % CI [0.48, 1.34]). Also, the moderately upward standard was significantly more motivating than the moderately downward standard ($p < .036$, $d = 0.32$, 95 % CI [0.02, 0.88]). Motivation did not

[6] When dummy coding the conditions comparison direction (upward vs. downward) and extremity (moderate vs. extreme) and leaving out the control group, a 2× 2 ANOVA reveals a main effect of extremity on similarity ($F(1,725) = 42.13$, $p < .001$, $\eta_p2 = .055$.), but no significant main effect of comparison direction, $F(1,725) = 1.25$, $p = .264$, $\eta_p2 = .002$, or interaction effect of comparison direction and extremity, $F(1,725) < 0.001$, $p = .995$, $\eta_p2 < .001$.

significantly differ between moderately and extremely upward standards (p = .966, d = 0.07, 95 % CI [-0.33, 0.54]). Motivation for the moderately downward standard did not differ from the extremely downward standard (p = .210, d = 0.23, 95 % CI [-0.10, 0.76]) or control group (p = .144, d = 0.26, 95 % CI [-0.07, 0.78]).

In order to control for self-reported fitness level, we added both contrasts as dummy codes together with self-reported fitness level into a regression analysis with motivation as the outcome variable. In line with the former results, Contrast 1 did not significantly predict motivation, b = -0.05, SE =.06, t(913) = -0.81, p = .416, 95 % CI [-0.17, 0.07], but Contrast 2 significantly contributed to motivation, b = -0.40, SE =.10, t(913) = 3.96, p < .001, 95 % CI [0.20, 0.59], while controlling for the effects of self-reported fitness level, b = 1.65, SE =.05, t(913) = 3.09, p = .002, 95 % CI [0.06, 0.27].[7]

Additionally, we conducted a meta-analysis across all four studies (Studies 1 and 3 and the supplementary study), first on the difference between moderately upward and moderately downward standards (Contrast 1) and on the difference between moderately upward and extremely upward standards (Contrast 2) on motivation, for a more comprehensive picture, using the open source statistical system Jamovi with the additional package "MAJOR" (project, 2018). Regarding Contrast 1, the overall relationship from a random-effects model was z = 2.62, p = 0.009, with a 95% confidence interval from 0.049 to 0.338, suggesting a larger motivational effect of the moderately fit compared to the moderately unfit standard. Regarding Contrast 2, the overall relationship from a random-effects model was z = 0.52, p = 0.605, with a 95% confidence interval from 0.049 to 0.338, suggesting no motivational difference between the moderately and extremely upward standard (see supplementary materials for forest plots of both meta-analytic results).

Book choice. As a behavioral measure, we implemented a book choice based on a lottery at the end of the study. Overall, 915 out of 917 participants choose one of the four books as part of the lottery. Across conditions, 686 participants (75%) chose a fitness book and only 229 participants (25%) chose the (unhealthy) cookbook. We predicted that participants comparing to a moderate (vs.

[7] With dummy codes a 2 × 2 ANOVA (without control group) reveals a main effect of comparison direction (F(1,725) = 35.13, p < .001, η_p^2 = .046.), no significant main effect of extremity (F(1,725) = 1.04, p = .309, η_p^2 = .001), and a marginal significant interaction effect of comparison direction and extremity on motivation (F(1,725) = 3.79, p = .052, η_p^2 = .005).

extreme) standard would be more likely to pick a fitness book (vs. unhealthy cookbook), but this was not the case, $\chi 2(1, 728) = 0.02$, $p = .898$, $\varphi_c = .005$. Second, we predicted that participants comparing to an upward standard are more likely to pick a fitness (vs. cookbook) than participants comparing to a downward standard. Again, this prediction was not confirmed, $\chi 2(1, 728) = 0.09$, $p = .764$, $\varphi_c = .011$. Finally, we predicted that participants comparing to a moderate standard would be more likely to pick a fitness (vs. cookbook) compared to the control standard; however, results did not reveal a significant effect, $\chi 2(1, 556) < 0.01$, $p = .986$, $\varphi_c = .001$.

2.4.3 Discussion

In Study 3, we further improved our paradigm by measuring participants' fitness score and calibrating upward and downward comparison standards around their individual fitness scores. Additionally, we compared the effects of all four standards to a control group represented by a standard without relevant information (fitness information). As predicted, we found that participants felt more similar to moderate standards compared to extreme standards and compared to the control group. However, the results for the similarity measure in the control group should be interpreted with caution, because the only information participants received in the control group was regarding gender and age, which both were matched to the participants. Therefore, the standard and participant share all given information.

The moderately upward and extremely upward comparison standard increased motivation compared to the control group and also compared to the downward standards. In line with Studies 1 and 2, we expected the extremely upward standard to lead to lower motivation compared to the moderately upward standard; however, there was no significant motivational difference between the two standards. Both standards equally motivated participants.

Contrary to our predictions, we did not find effects of comparison standards on the behavioral choice measure. Instead, and quite fortunately from a public health perspective, we found that two-thirds of participants chose a fitness book over a cookbook. One possible explanation is that the fitness frame of the study made the fitness goal quite salient to participants and may have washed out additional comparison standard effects regarding book choice.

2.5 General Discussion

How do salient comparison standards in our social environment impact our motivation to make progress with regard to important health goal such as exercising? In a set of three studies, we systematically investigate how being confronted with a particular comparison standard influences participants' motivation and healthy behavioral choices towards an exercising goal. We manipulated the direction of comparison (upward comparison vs. downward comparison) and the extremity of the standard (moderate vs. extreme) in order to predict different (motivational) outcomes. Our findings suggest that people feel more similar to moderate over extreme standards, a finding that is in line with the selective accessibility model (SAM; Mussweiler, 2003). However, the SAM does not make strong predictions regarding how the extremity of an upward or downward standard influences motivation to invest effort towards a challenging health goal. Filling this gap between similarity assessment and motivational consequences, we made specific predictions about each standard's likely motivational impact.

Across all conducted studies, we found that there was always one condition that consistently boosted motivation and in which the comparison standard was perceived at high levels of perceived similarity: the moderately upward comparison standard. Hence, across all studies, it is fair to say that the moderately fit standard clearly emerged as the most reliable "horse" to bet on when the goal is to facilitate the pursuit of health goals via processes of social comparison.

We also found that extreme downward comparison standards were consistently demotivating across studies. We have argued that his may be the case because extremely inferior standards may give rise to an illusion of accomplishment, resulting in coasting behaviors. In support, people did not feel close to the extremely unfit standard. This suggests that participants, on average, may have perceived themselves as clearly surpassing the level of goal progress exemplified by the extremely unfit standard, possibly resulting in a coasting mindset associated with a reduced willingness to invest further effort into the goal (Carver & Scheier, 1990).

However, findings with regard to the moderately unfit standard were less clear. These somewhat inconsistent results may be due to the different manipulations in our studies.

Overall, however, our meta-analytic summary across studies suggests that moderately downward standards are relatively less motivating than moderately upward standards. Nonetheless, additional research is needed to investigate possible motivating effects of the moderately unfit standard we observed in the first two studies. For example, the interplay of downward comparison, goal framing, and dispositional promotion versus prevention focus might play a role here (Crowe & Higgins, 1997; Lockwood et al., 2002), such that, at least under certain conditions, comparing to a similar person who did not complete their fitness goal or has difficulties with regular workout routines and self-discipline may motivate people to prevent a similar fate by investing more self-control and time resources (Shakya et al., 2015).

Finally, our meta-analytic summary across studies suggest that extremely upward social comparison standards were, on average, equally as motivating as moderately upward standards. This finding was surprising, insofar as we had hypothesized that extreme upward comparison should result in lower levels of perceived attainability, and, ultimately, disengagement. One possible explanation for this result is that, *initially*, an extremely upward standard may be judged as motivating, due to motivational biases, such as planning fallacy (Buehler et al., 1994) and overly optimistic self-beliefs. Such positive illusions and false hopes (Polivy, 2001) may indeed facilitate goal setting. However, one possible caveat with extreme upward standards is that people may be prone to set unrealistically high goals (Schwarzer, 1998; Taylor & Brown, 1988). Over time, they may ultimately accrue sufficient disconfirming information regarding the attainability of such high standards, and eventually withdraw effort or fully disengage from their goal (Polivy, 2001; Wrosch et al., 2003). In future research, the different motivational effects of comparison standards should be measured over time to give a better picture of actual self-control and also effort investments towards the health goal. We assume that a long-term comparison to a moderately fit standard is more beneficial than an extremely fit standard, not only for motivational and self-regulatory purposes, but also for positive emotions and self-esteem (Collins, 1996).

2.5.1 Limitations and Future Research

One limitation is that the behavior and extremity of the standards in Studies 1 and 2 were determined globally rather than individually. That is, comparison standards might not always be

different from the target in the intended direction. For instance, a target with a fitness level far below average might perceive a moderately unfit standard as a superior (upward) standard. Note that this is a highly common approach in the social comparison literature to give the direction of comparison as being for instance, athletic versus unathletic (Mussweiler et al., 2004a), or healthy versus unhealthy (Head & Bruchmann, 2018). However, it neglects the participant's own current standing on the given dimension, treating variations in one's own standing as a nuisance factor. In Study 3, we therefore improved the manipulation of the standards by providing participants with a personal fitness score and a comparison standard that was calibrated around their own score. Still, results from Study 3 regarding the extremely upward standard suggest that additional, and even more fine-grained calibration may be fruitful.

2.5.2 Theoretical and practical implications

Our findings connecting social-cognitive approaches to social comparison and motivational aspects in (health) goal pursuit have both theoretical and practical implications. At the theoretical level, our findings suggest that contemporary information-processing models of social comparison (focusing on assimilation and contrast of *judgment*) may not be fully adequate to cover the *motivational* consequences of social comparison for the pursuit of health and other challenging goals. The present research shows that these motivational implications may differ considerably as a function of both the direction of comparison and the standard's extremity. Future research may benefit from a manipulation of standard extremity that is even more fine-grained than what we attempted in the present line of research. This way, we may be able to more precisely map out the exact (nonlinear) function connecting standard-target discrepancies in social comparison with effort investment regarding the focal goal in question. Of particular interest here is the "tipping" point at which the willingness to incest effort turns into a reduction of effort investment or even disengagement, as this tipping point may demarcate the maximally effective social comparison standard discrepancy on the continuum. From an applied perspective, these findings are relevant for health campaigns seeking to boost people's motivation towards health goals by selecting the most appropriate comparison standard. Across all studies conducted, we can conclude that moderately superior standards (i.e., above the average norm in the population) are the primary choice in applied contexts, such as large-scale

campaigns, where the goal is to create an optimal motivational social comparison environment for the largest number of people and where providing more tailored, adaptive feedback is not feasible.

Chapter III: A Motivational Framework of Social Comparison

This chapter is based on following publication:

Diel, K., Grelle, S., Hofmann, W. (2021). A motivational framework of social comparison. *Journal of Personality and Social Psychology*. *120*(6), 1415-1420. http://dx.doi.org/10.1037/pspa0000204 (Published online: January 28 2021)

Please note that changes in headings, citation style, and formatting were undertaken to fit the layout of this dissertation. No changes were made to the content of the article.

Abstract

Whether people's current motivation levels increase or decline also hinge on their social environment. The current research tightly integrates motivational principles from self-regulation research with social comparison processes. In a preregistered experience sampling study including more than 5,400 social comparison situations from people's everyday life, we investigated how discrepancy assessments between the self and a comparison standard influence people's motivation and affect. Results revealed a non-linear relationship between negative discrepancies (upward comparisons) and effort investment ("pushing"): Whereas motivational pushing increases with negative discrepancies, more extreme upward comparisons were associated with less pushing, but increased disengagement ("giving up"). The effect of negative discrepancies on pushing motivation was even more pronounced for people perceiving high control in their domain of comparison or when the domain was considered as important. Positive discrepancies (downward comparisons), on the other hand, were related to a reduction in effort ("coasting"). Similarly, emotional responses, such as an increase or decrease in self-esteem are yet another signal for whether someone needs to invest further effort at a current time. The self-regulatory perspective on social comparison provides a novel framework uniting motivational, emotional, and cognitive processes of social comparison for a better understanding of when social comparison can benefit or hinder people's everyday goal pursuit.

Keywords: self-regulation, motivation, social comparison, everyday life, experience sampling

3.1 Introduction

Healthier, richer, more successful - people harbor various goals in life that steer them towards self-improvement. Consistent and successful goal pursuit is challenging (Hofmann et al., 2011; Kotabe & Hofmann, 2015). If, for instance, one decides to live a more environmentally friendly life, they will frequently face new challenges: Do I take the bicycle or the car on this rainy day? Do I buy the cheap or the much more expensive but sustainable product? In order to keep track of the initial long-term goal in such situations, one must routinely monitor the discrepancy between one's current standing and the aspired standard (Carver & Scheier, 1981, 1982). Am I yet living the eco-friendly life I aimed for? If the answer is no, action can be taken to further adjust goal pursuit and reduce the discrepancy between the current and the actual goal state - even if that means putting on a rain coat to take the bike to work.

Goal pursuit is only partially a private matter though: People do not live in a social vacuum, and a wealth of research has attested that their thoughts, intentions, and actions are influenced by their social environment. Most important, people seek and receive feedback on their abilities, performances, or appearances from their peers and other relevant comparison standards (Festinger, 1954; Taylor et al., 1996), and these standards can be a source of inspiration (Taylor & Lobel, 1989) as well as of self-worth (Morse & Gergen, 1970; Taylor & Lobel, 1989; Wills, 1981). In the present paper, we bring together two literatures whose marriage is long overdue: classic work on self-regulation as a motivated process triggered by discrepancy assessment (Carver & Scheier, 1981, 1982), and the literature on social comparison with reference to the three major social comparison motives identified in this body of research: self-assessment, self-improvement, and self-enhancement. We first lay out the theoretical foundation for this work. In the empirical part, we utilize a large scale experience sampling approach to investigate the interplay of social comparison and self-regulation processes in people's real life settings regarding motivational and emotional consequences of social comparison. We believe that our integration places social comparison in a broader, relevant motivational context that, we believe, sets the stage for a refined understanding of the functionality of social comparison for everyday goal pursuit, helps to recast certain phenomena of interest in a theoretical light (e.g., self-esteem boosts as part of a motivational

"coasting" response), and allows asking novel research question.

3.1.1 The three Social Comparison Motives: Self-evaluation, Self-improvement, Self-enhancement

Leon Festinger's (1954) theory of social comparison already provided the basis of the three main social comparison motives discussed in the literature: First, people engage in comparisons with others who are perceived as similar on an evaluated dimension out of the motivation to accurately assess their own abilities (self-evaluation) (Festinger, 1954; Wheeler, 1966). This way, social comparison is a fast and efficient tool to gain evaluative self-knowledge regarding one's own standing in the absence of an objective standard (Corcoran & Mussweiler, 2009; Mussweiler & Epstude, 2009; Taylor et al., 1996). Second, in order to improve one's abilities (self-improvement), there is a motivation to compare to others who are already more successful (Taylor & Lobel, 1989). Third, people have the need to maintain or enhance their self-esteem (self-enhancement). This need is satisfied by comparing to others who are worse off than oneself (Wills, 1981). Despite their prominence, these three social comparison motives have not been conceptually integrated within a common self-regulatory framework. For instance, when pursuing a long-term goal, a comparison to a superior target (upward comparison) can provide a motivational boost, at least up to a certain point at which discrepancies become too large (e.g. Atkinson, 1964; Heckhausen, 1977; Klinger, 1975). On the flipside, when comparing to an inferior target (downward comparison), people might feel that enough progress has been made already, and hence, withdraw further effort towards their initial goal, a process termed as "coasting" (Carver, 2004; Carver & Scheier, 1990). Perhaps this reluctance of integrating social comparison research more tightly with motivational concepts of self-regulation is due to its strong historical emphasis on social-cognitive mechanisms. Previous influential models on social comparison such as the selective- accessibility model focus on social judgments and spell out when a target selectively assesses standard-consistent or standard-inconsistent knowledge, leading to assimilation or contrast, respectively (Mussweiler, 2001, 2003). For instance, a target is more likely to assimilate with a standard that is perceived as moderate or attainable and contrasts away from a standard that appears to be extreme or unattainable (Lockwood & Kunda, 1997, 2000). However, it is important to realize that similarity and dissimilarity *judgments* do not fully translate

into motivational outcomes (Diel & Hofmann, 2019). Rather, assimilation and contrast merely influence people's subjective perception of self-other discrepancies at the self-monitoring stage; how such discrepancies translate into motivational, such as effort investment or emotional consequences, such as pride are conceptually distinct matters. In the following, we attempt to fill this conceptual space.

3.1.2 Linking Self-regulation and Social Comparison: From Self-assessment to Effort Investment

We aim to combine social comparison principles with motivational principles from self-regulation research by treating social comparison motives as functional and adaptive states that can temporarily support but also hinder goal pursuit. In the process of pursuing goals (i.e. self-improvement motivation), an individual seeks feedback by comparing his or her current state and goal progress to a standard of reference (Carver & Scheier, 1982, 1990). One key proposition for the present purpose is that this salient standard of reference can often stem from the observation of other people in one's environment - especially so, when there is some ambiguity of interpretation and these standards are, in principle, applicable or relevant (Festinger, 1954; Wood, 1989). Carver and Scheier seemed to have foreshadowed this link when distinguishing between two types of comparative standards in a footnote of their seminal book:

> "One important distinction among behavioral standards would seem to be between the relatively private and personal standards that people have (e.g., their attitudes and moral beliefs) and standards that are evoked by such means as acts of social comparison. The latter would seem to be more temporary and situation-specific than the former, though both clearly are used to guide behavior at various times. Our position is that the two are similar in serving as reference values for self-regulation …" (Carver & Scheier, 1982, p. 121, Footnote 5)

Furthermore, self-regulation or goal pursuit operates as a feedback loop, whereby a person repeatedly compares their current standing with a salient standard of reference. The outcome of this comparison process is a discrepancy assessment, which corresponds to the drive to evaluate their abilities and opinions accurately (self-evaluation, Festinger, 1954). Thus, in the process of social

comparison, a target assesses the discrepancy between their own standing and the standing of a comparison standard. As a result, the discrepancy can be positive, meaning the target appears to be more successful on a certain dimension than the standard, or the discrepancy is negative and the target falls behind relatively to the standard.

As shown in Table 1, the social comparison motives can thus be linked to different stages and comparison outcomes of the self-regulation feedback loop. First of all, the self-evaluation motive corresponds to the comparator or the discrepancy-assessment stage of the self-regulation process. In other words, the self-evaluation motive propels people to seek a standard of reference to gather information about their current state on the dimension of interest. For most diagnostic standards in the process of self-assessment, people turn to others who are already similar to them on the evaluated dimension (lateral comparison). The outcome of the self-assessment stage is the detection of possible discrepancies between the person's standing and their salient comparison standard. Importantly, the sign and extent of the ensuing discrepancy can have quite divergent motivational and emotional consequences. If the discrepancy is negative (upward comparison), a target might be motivated to reduce the deviation from the standard, which can translate into increased self-improvement motivation and effort investment (*Pushing*; Fulford et al., 2010). However, not all negative discrepancies result in pushing: If the gap between the target's standing and the standard is too large, the standard seems out of reach, and motivational effects turn into disengagement (e.g. Carver & Scheier, 1982; Gollwitzer, 1996; Jostmann & Koole, 2008; Kuhl, 2000). For instance, a co-worker who commits him- or herself to climate protection can be inspiring for the personally endorsed goal of living a more environmentally friendly life. However, comparing oneself to the environmental activist Greta Thunberg, the discrepancy between one's own and the environment-protecting behavior of the teenager may seem rather discouraging.

It should be noted that several models and traditions converge on this disengagement prediction: First, the idea that people disengage from a current goal if they expect their chances of reducing the discrepancy to be very small is already an essential component of the Carver and Scheier model of self-regulation (e.g. Carver & Scheier, 1981, 1982, 2005) and serves as a basis

for this prediction as long as one accepts the auxiliary assumption above that social comparison standards may act just like personal standards. Second, this prediction dovetails with a larger number of general or domain-specific models in the *expectancy* × *value* tradition positing that low perceived attainability breeds disengagement (e.g. Atkinson, 1964; Klinger, 1975; Wright & Brehm, 1989).[8] In a nutshell, these models emphasize the role of people's expectancies and beliefs regarding focal goal pursuit. And third, the disengagement prediction is also consistent with a complementary but related family of models emphasizing the *functional* role of disengagement (e.g Gollwitzer, 1996; Jostmann & Koole, 2008; Kuhl, 2000). From this more holistic perspective, disengagement is seen as a beneficial cognitive-motivational mechanism that safeguards against futile rumination and the associated distress and frustration resulting from a suboptimal allocation of motivational resources, thus freeing up capacity and time for alternative, more promising courses of action (e.g. Jostmann & Koole, 2008).

[8] For instance, in his classic motivation theory of achievement, Atkinson (1964) posited that, among individuals high in achievement motivation, selection of and effort investment towards tasks of varying difficulty follows an inverse U-shaped function, with a peak at medium levels of difficulty, resulting from the product of success expectation (decreasing with increasing task difficulty) and value of success (increasing with increasing task difficulty).

Table 1

Mapping of key motives from social comparison literature to motivational states in self-regulation research

Motive	Goal of Social Comparison	Direction of Comparison	Motivational Effects of Comparison
Self-Evaluation	Gathering information about oneself (actual states and abilities)	Lateral	Self-assessment: detection of discrepancies between actual standing and salient comparison standard
Self-Improvement	Aiming to achieve something worthwhile	Upward	Investing effort aimed at discrepancy reduction/progress (Pushing); Possibility of disengagement if discrepancy is/remains large
Self-Enhancement	Improving self-esteem	Downward	Coasting on the long-term goal; harvesting short-term rewards more likely

On the other hand, a positive discrepancy (downward comparison) is associated with a boost in self-esteem, but also with a drop in motivation: In comparison to an inferior standard, people perceive themselves as "safe" regarding their goal progress, which leads to temporary effort withdrawal on the motivational dimension. On the emotional dimension, downward comparison has been shown to be linked to a boost in self-esteem and general positive emotional outcomes (Morse & Gergen, 1970; Taylor & Lobel, 1989; Wills, 1981). Due to the traditional focus on the emotional consequences of downward comparison in the literature, this presumably proximal and consequential demotivating effect of downward comparison has been sorely neglected. Hence, we use the term "coasting" to refer to the demotivating potential of downward standards with regard to goal pursuit (Fulford et al., 2010) and to thus interlink self-regulation and social comparison more tightly at this overlooked juncture.

Our predicted relationship between social-comparison-induced discrepancy assessment and motivation (i.e., effort investment) is graphically summarized in Figure 1 as a smooth hypothetical

curve. Note that the function is assumed to be non-linear in nature due to the above-mentioned interleaved effort-enhancing (i.e., pushing) versus effort-reducing (i.e., coasting, disengaging) principles of goal-pursuit set in motion by increasing discrepancies between the own and the target person's standing on the relevant comparison dimension: Up to a certain point, increasing discrepancies between own standing and an upward comparison standard increase the effort people are willing to invest for self-improvement (see Figure 1). This point signifies the optimum of motivational potential, triggered by what we term the *maximally effective standard* (MES) within the possible distribution of available comparison standards in the social environment. Beyond this turning point of the MES, larger discrepancies reduce invested effort, however, leading over to a state of disengagement as people give up the pursuit of the "out of reach" standard (see Figure 1). On the other hand, increasing discrepancies between own standing and worse comparison standards result in a coasting mindset characterized by reduced effort investment and a possible shift towards alternative priorities than the initial goal pursuit as the target has already outperformed the standard (see Figure 1).

The above-mentioned motivational effects are tightly interlinked with affective consequences of social comparisons. In line with the positive effects of downward comparison on people's affect (Morse & Gergen, 1970; Taylor & Lobel, 1989; Wills, 1981), also a positive discrepancy triggers positive emotions, such as happiness and enhanced self-esteem (Carver & Scheier, 1990, 2011). A negative discrepancy, on the other hand, leads to feelings of ambivalence (Carver & Scheier, 1981, 1990, 2011). On the one hand, a person is inspired and hopeful to reach their goal, on the other hand, the same phase is accompanied by feelings of uncertainty if one will eventually reach their goal. However, once the motivational peak has been passed, a negative discrepancy causes that the target no longer believes in the achievement of the goal. Hence, a large discrepancy between target and upward standard results in negative affect and a threat to the self-esteem (Carver & Scheier, 1990, 2011), which is in line with the negative link between upward comparison and negative affect in the traditional social comparison literature (Morse & Gergen, 1970). Overall, those emotional consequences are best thought of as correlates or markers of the underlying motivational states and are hence, of secondary interests to the motivational effects.

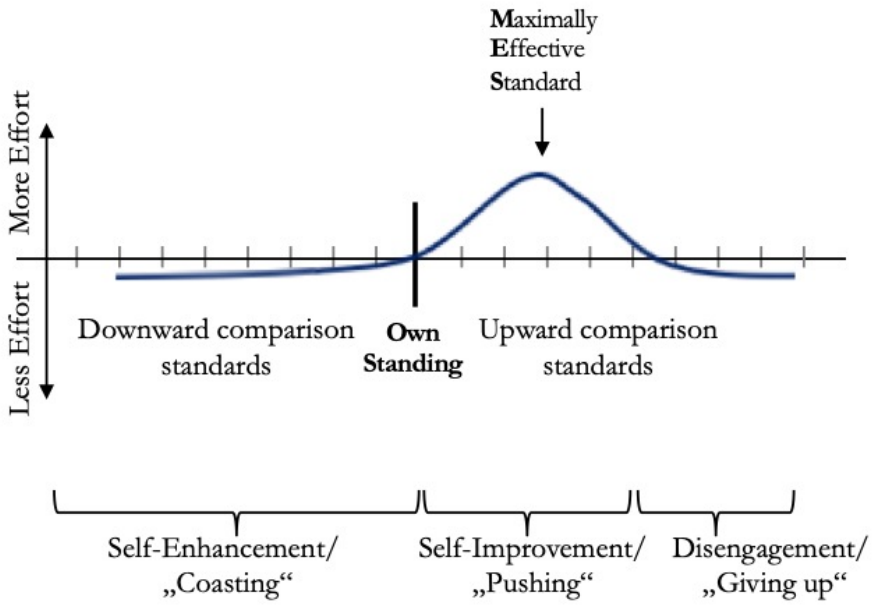


Figure 1. Hypothesized mapping of discrepancy assessment between own standing and relatively better or worse social comparison standards, change in effort investment (curve), and key motivational states of self-improvement ("pushing"), self-enhancement ("coasting"), and disengagement ("giving up").

3.1.3 Boundary Conditions

Furthermore, we investigate a number of potential moderators of the motivational effects. First, we examine how features of the comparison domain, namely domain control and domain importance influence the effects of social comparison on motivation. Past research has shown that people benefit from upward comparison if a given comparison standard is perceived as attainable (Lockwood & Kunda, 1997). Likewise, domains or goals that are seen as important or relevant imply greater goal commitment, which has been shown to increase people's effort investment and persistence in goal pursuit (Fishbach & Ferguson, 2007; Kruglanski et al., 2002; Locke & Latham, 2002). Thus, we predict that a person might derive more self-improvement motivation from a superior standard (a) if perceived domain control is high rather than low, and (b) if the importance ascribed to the comparison domain is high rather than low. Due to the goal pursuit literature separating perceived control and goal commitment (Kruglanski, 2002), we expected domain control and importance to exert additive, separable effects.

Secondly, comparisons can be intentional and strategic or encountered or spontaneous (Wood, 1996). When thinking about specific aspects of the self, one may intentionally select an appropriate

standard that is informative about the own self. A deliberate social comparison search is mostly driven by a particular motivation (e.g. self-improvement; Taylor & Lobel, 1989). On the other hand, comparisons can be made spontaneously without intentions (Alicke, 2007; Mussweiler, 2004; Want, 2009) even if a standard is perceived as logically inappropriate (e.g. stumbling upon famous people in magazines) (Gilbert et al., 1995). We explore how social comparison influences motivation dependent on whether a comparison was actively sought or the target was rather confronted with a standard.

3.1.4 The Present Research

In order to test our predictions and estimate the presumed non-linear relationship between discrepancy assessments in social comparison and their motivational and affective consequences with high resolution and optimal generalizability, we decided against an experimental approach that would only allow to probe the function at a limited number of experimentally induced levels and within a limited number of content domains. Instead, we chose an ecologically valid, big data approach, the experience-sampling method, to sample thousands of real-world social-comparison contexts from the overall "universe" of standard-target discrepancy assessments in an attempt to optimize both resolution, generalizability, and statistical power (Bolger & Laurenceau, 2013; Hektner et al., 2007).

In our study, participants reported their daily social comparisons over a 5-day period. All measures as well as the preregistration can be accessed via https://osf.io/za35c/. Building on our framework, we preregistered our main hypotheses on how social comparison (i.e. social comparison direction) influences motivational outcomes, such as pushing, coasting, disengagement, as well as effort investment and emotional correlates (positive and negative affect, self-esteem). Moreover, we investigated possible moderators of the effect of comparison direction on motivational consequences. Lastly, we preregistered more general predictions concerning base rates of social comparison (e.g. frequency of upward vs. downward comparisons overall). Our main hypotheses (H) were the following:

H1a: We predicted that downward comparison is associated with the motive of self-

enhancement and upward comparison is associated with the motive of self-improvement (preregistered).

H1b: We hypothesized that upward comparison is associated with motivational pushing. However, we predicted that extreme (vs. moderate) upward comparison will lead to disengagement and less pushing. Thus, we predicted a non-linear pattern between upward comparison and motivational pushing – expecting pushing to decrease with more extreme upward comparisons (preregistered).

H1c: We hypothesized that downward comparison is associated with motivational coasting (preregistered).

H2: We predicted that upward comparison is associated with increased effort investment in a given domain. However, we predicted a non-linear pattern between upward comparison and effort investment, expecting effort to decrease with more extreme upward comparisons (preregistered).

Furthermore, we investigated potential moderators of the relationship between comparison direction and motivational effects: perceived control and importance at the level of goal-pursuit, and whether people actively sought the comparison (vs. were confronted with the standard).

H3: We predicted that upward comparison leads to increased motivational pushing when people perceive high (vs. low) control over the comparison domain (preregistered).

H4: We predicted that upward comparison leads to increased motivational pushing when the domain is high (vs. low) on importance (preregistered).

Exploration: We explored whether upward comparison leads to different effects on pushing for people actively seeking the comparison compared to people who were confronted with the comparison.

Exploration: We explored how the interactions of comparison direction with the proposed moderators influence disengagement and coasting.

Moreover, we preregistered following hypothesis on how social comparison influences people's emotional states:

H5: We predicted that upward comparison is associated with less positive and more negative affect, lower self-esteem, and less feelings of pride but more feelings of guilt (preregistered).

As stated in the preregistration we aimed to investigate base rates of social comparison: We investigated the prevalence of upward compared to lateral and downward comparison in people's daily life. A field study of social comparison has revealed that people more often compare downward compared to lateral or upward. However, a recent meta-analysis (Gerber et al., 2018) revealed that people prefer upward comparison over downward comparison. Thus, we make following predictions:

H6: We hypothesized that people more frequently compare upward than downward (preregistered).

Exploration: We explore triggering conditions such as successes and failures providing more insights into antecedents of social comparison (preregistered).

3.2 Method

3.2.1 Participants

We recruited a heterogeneous sample of 454 German participants (M_{age} = 29.32, SD = 8.81, ranging from 18 to 66, 70.1 % female, 55.9 % university students) via panels and research data bases, social media platforms, and flyers at a German university. Statistical power calculation in multilevel models is complicated by the fact that it varies as a function of effect size and intraclass correlations, it differs for fixed effects versus random effects, and it is affected by both the number of Level-2 units (i.e., persons) and the number of Level-1 observations per person

(Kreft & De Leeuw, 1998). However, there is agreement that the number of Level-2 units is generally more crucial (Snijders, 2005). As we did not entertain one cardinal hypotheses, but instead a number of hypotheses (both at Level 1 and Level 2), we therefore sought to optimize power by recruiting as many persons as possible at the top level of the multilevel design through the pooling of available resources ($N = 400$), while following general recommendations to allow for a sufficient number of repeated observations within persons. As stated in our preregistration, we therefore set the targeted sample size to a minimum of 400 participants and, assuming a response rate around 75%, set the number of ESM signals to 25 per person.

Participants were compensated with 5 Euro for completing the intake survey and with additional 15 Euro if they answered at least 70 % of the daily signals. Additionally, we raffled three times 100 Euro and drew the winners after completion of the experiment. Each completed survey counted as a lottery ticket. Thus, the more surveys were completed the higher was the chance to win the lottery. The study was approved by the ethics commission of the German Psychological Society.

3.2.2 Materials and Procedure

Participants completed a screening survey and an intake survey, using the Qualtrics Software. The movisensXS App, Version 0.7.4162 (movisens GmbH, Karlsruhe, Germany) was utilized for the mobile phase. Participants first completed a screening survey introducing the study aim as "investigating social comparison processes in daily life". This survey informed participants about study demands (e.g. duration of the study and required number of completed questionnaires, such as intake survey and mobile phase questionnaires), participant monetary compensation, and requirements that needed to be fulfilled in order to take part in the study. Participants could only participate if they were at least 18 years old, residing in Germany and if their smartphone was equipped with an Android operating system. If all requirements were met, participants were informed about further proceedings of the study. After successful installation of the Movisens App, participants were first asked to complete the intake survey, which took approximately 15 minutes. The intake survey consisted of dispositional measures and demographic information, which included gender, age, and socioeconomic status (for a full list of materials, see OSF). Secondly,

participants answered a questionnaire concerning compensation information for participating in the study. Next, participants began with the mobile phase, which included a practice survey on the day the App was installed and the main mobile phase started on the following day. Participants received five signals over the course of five days (between 10 am and 10 pm). After completing the five- day mobile phase, participants received a last e-mail thanking for participation and were instructed to uninstall the Movisens App.

3.2.3 Mobile Phase

Every time participants received a mobile phone signal, they were asked if they had engaged in a social comparison since the last signal. If participants answered with *Yes*, they indicated how many minutes ago the comparison took place, the domain of comparison (e.g. financial, nutrition, sport/fitness, see OSF for the whole list of domains) and described in a few words what the comparison was about before continuing with the main questionnaire.

If the answer to the first question was *No*, they could report any other situation involving a social comparison in the past. Next, they indicated when the comparison approximately took place (1 = *less than 5 hours ago*, 2 = *more than 5 hours ago*, 3 = *yesterday*, 4 = *last week*, 5 = *last month*, 6 = *more than a month ago*, 7 =*I cannot remember a situation right now*). If the last option was chosen, the questionnaire ended. If any of the other options was chosen, participants continued with the same question about comparison domain and the short comparison description.

Next, participants proceeded with the main questionnaire, which was divided in five blocks that were displayed in a randomized order: Comparison features, comparison direction, motivational states, emotional states, and context.

Descriptive Comparison features. Participants indicated to whom they compared (1 = *romantic partner*, 2 = *close friend*, 3 = *ordinary friend*, 4 = *acquaintance*, 5 = *imaginary person*, 6 = *stranger*, 7 = *family member*, 8 = *famous person*, 9 = *other*), the type of comparison (1 = *direct interaction*, 2 = *short contact/no interaction (e.g. street)*, 3 = *phone/chat*, 4 = *fantasy/daydream*, 5 = *offline*, 6 = *online*, 7 = *other*), and whether the comparison was actively sought (vs. being confronted with) on a 7-point Likert scale (−3 = *was rather confronted with*

comparison, 0 = *both*, 3 = *was rather actively seeking comparison*).

Comparison Direction. In this block, participants indicated the direction of comparison in a fine-grained manner. The question wording was as follows:

"If we compare ourselves with another person in a certain domain, we can perceive ourselves as better, worse or similar to the comparable person. Please indicate how you perceived yourself in the comparison situation", (−5 = *extremely worse*, −4 = *quite a lot worse*, −3 = *a lot worse*, −2 = *moderately worse*, −1 = *a little worse*, 0 = *similar*, 1 = *a little better*, 2 = *moderately better*, 3 = *a lot better*, 4 = *quite a lot better*, 5 = *extremely better*). For the main analyses, we recoded the item so that higher (positive) values represent upward comparison (superior standard) and lower (negative) values downward comparison (inferior standard). For some analyses, we were only interested in the general direction of comparison (upward, downward, or lateral) and coded participant responses accordingly (i.e., above 0, 0, below 0).

Motivational States. Firstly, participants reported on their motivational states on 7-points likert scales: Pushing ("I'm currently motivated to improve myself"), Coasting ("I can currently rest on what I have already achieved"), and Disengagement ("I would like to give up right now", 1 = *not at all*, 7 = *very much*). Secondly, participants indicated their general social comparison motives of self-improvement, self-enhancement, and self-evaluation on a 5-point Likert scale ("I compared to the other person, because..." 1) …I want to improve myself, 2) …to feel better about myself, 3) …to evaluate myself; 1 = *not at all*, 5 = *very much*). Lastly, participants answered an item about effort investment on a 5-point Likert scale ("How much effort would you like to invest in the future to improve in the domain of comparison?", 1 = *none at all*, 6 = *a lot*).

Emotional States. In this block, participants reported on a number of emotional states after the comparison: state self-esteem ("How confident did you feel after the comparison?", −3 = *a lot less confident*, 0 = *same*, 3 = *a lot more confident*), positive and negative affect ("How positive/negative do you feel after the comparison?", 1 = *not at all*, 5 = *a lot*), and pride and guilt ("How proud/guilty did you feel after the comparison?", 1 = *not at all*, 5 = *a lot*).

Context. This block assessed possible moderating variables, such as domain control ("How much control do you have over the domain of comparison?", 1 = *none at all*, 7 = *very much*), and

domain importance ("How important is the domain of comparison to you?", 1 = *not at all important*, 7 = *very important*).

Additionally, we assessed triggering events before the comparison on 7-point Likert scales, including a misfortune (1 = *no misfortune at all*, 7 = *a strong misfortune*) and fortune (1 = *no fortune at all*, 7 = *a strong fortune*).

3.2.4 Analysis Strategy

For the main analyses, we conducted multilevel random intercept analyses using the lme4 package in R (Bates et al., 2014) and ancillary packages (see R script on OSF project page). Out of the data collected from the 454 participants, 425 participants provided data on our independent variable used for our main analyses. All variables at Level 1 were person-mean centered (Enders & Tofighi, 2007). However, we decided against person-mean centered comparison direction scores for interpretation purposes. Comparison direction scores range from -3 (*extreme downward comparison*) to 0 (*similar comparison*) to 3 (*extreme upward comparison*) and thus, contain a meaningful point of zero. Further, we decided to report the sole unstandardized coefficients since in multi-level models the conventional estimates of effect size do not apply. Because the random part of the multi-level regression model is non-invariant for linear transformations, the estimates of the variance components can change, even significantly (Hox, 2010).

3.3 Results

Descriptive and Frequency Data

Overall, our sample of 454 participants provided data for a total of 11,176 measurement occasions including 5,516 social comparison situations. On 32.2% of all occasions, participants reported they had engaged in a social comparison since the last signal. From the remaining occasions the participants reported a social comparison that took place either less than five hours ago (6.1%), more than five hours ago (5.4%), yesterday (9.6%), last week (8.7%), last month (4.8%), or more than a month ago (4.5%). The remaining 67,8% of occasions, participants reported they could not remember a social comparison situation at this moment. Participants compared mostly to acquaintances (16.4%) and close friends (16.0%), followed by strangers (15.4%), ordinary friends

(14.7%), other (11.5%), family members (11.0%), romantic partners (9.5%), famous people (3.6%), and imaginary persons (1.9%).

The most prevalent domains of comparisons were academic and work related (23.7%), sport and fitness related (11.1%) and concerning appearance and looks (8.3%). For an overview of all comparison domains, see Supplementary Figure 1.

3.3.1 Social Comparison and Motivation

General comparison motives. We hypothesized that upward comparison was associated with the social comparison motive of self-improvement and downward comparison with the motive of self-enhancement (H1a). We included the motives of self-improvement, self-enhancement, and also self-evaluation as predictors of comparison direction into a multilevel random intercept model. As predicted, self-enhancement motivation negatively predicted upward comparison (hence, positively predicted downward comparison), $B = -.93$, $p < .001$ and self-improvement motivation positively predicted upward comparison, $B = .55$, $p < .001$. Self-evaluation was not a significant predictor of comparison direction, $B = .03$, $p = .251$.[9] This pattern is consistent with the idea that self-improvement and self-enhancement reflect an increase vs. decrease in motivational aspiration ensuing from social comparison. In contrast, self-evaluation was not associated with any direction, consistent with our assumption that self-evaluation is "directionless" in the sense that it might best reflect a motivational antecedent of the social comparison process (i.e., finding out more about one's own standing ex ante, without preferring an upward or downward social comparison outcome). At a more abstract level, there is a notable parallel between the self-evaluation motive in social comparison and the self-knowledge or self-accuracy motive in broader theories of the self (Sedikides & Strube, 1997), or the deliberative mindset in the mindset theory of action phases (Gollwitzer, 1990).

3.3.2 Mapping Self-Regulatory Discrepancies and Motivational States

We hypothesized that upward comparison is associated with motivational pushing. However,

[9] Also without self-evaluation as a model predictor, self-improvement positively predicted ($B = .56$, $p < .01$) and self-enhancement negatively predicted upward comparison ($B = -.93$, $p < .001$).

we predicted that extreme upward comparison will lead to disengagement and thus, predicted a non-linear relationship between degree of upward comparison and motivational pushing and disengagement (H1b). First, we regressed comparison direction scores on pushing. In order to test the predicted non-linear relationship, we added quadratic, cubic and quartic terms to the regression model. Results revealed a significant linear relationship between upward comparison and pushing B = .12, p < .001, a significant quadratic effect B =.03, p =.004, as well as a significant quartic effect B =−.002, p < .001. The cubic effect was non-significant, B =−.001, p =.332. Hence, pushing increases with less extreme downward comparison and increasing upward comparison. However, when upward comparison becomes too extreme, motivation decreases (Figure 2a). Lastly, we conducted the same multilevel analysis with disengagement as outcome variable. Results revealed a significant linear trend, B =.19, p < .001, as well as a significant quadratic, B =.04, p < .001, and cubic trend, B =.002, p =.029. The quartic effect was non-significant, B < .001, p =.710 (Figure 2b). Hence, disengagement increases with upward comparison, especially towards more extreme upward comparisons.

Moreover, we predicted that downward comparison is associated with motivational coasting (H1c). We conducted a multilevel regression analysis with comparison direction scores as the predictor and coasting as outcome variable to test the predicted linear relationship. As hypothesized, upward comparison negatively predicted motivational coasting, B= -.38, p < .001, indicating that downward comparison has a positive influence on motivational coasting (Figure 2c).

Effort. In line with the motivational pushing effect, we expected a non-linear relationship between upward comparisons and effort investment (H2). We regressed comparison direction scores on effort investment and again, added quadratic, cubic and quartic terms to the regression model. Results revealed a significant linear, B =.10, p <.001, a significant quadratic effect, B =.03, p <.001, as well as a significant quartic effect, B = -.002 p <.001. The cubic effect was non-significant, B < -.001, p =.946. As shown in Supplementary Figure 2, effort initially increases for moderate upward comparisons but then starts to decline with more extreme upward comparisons.

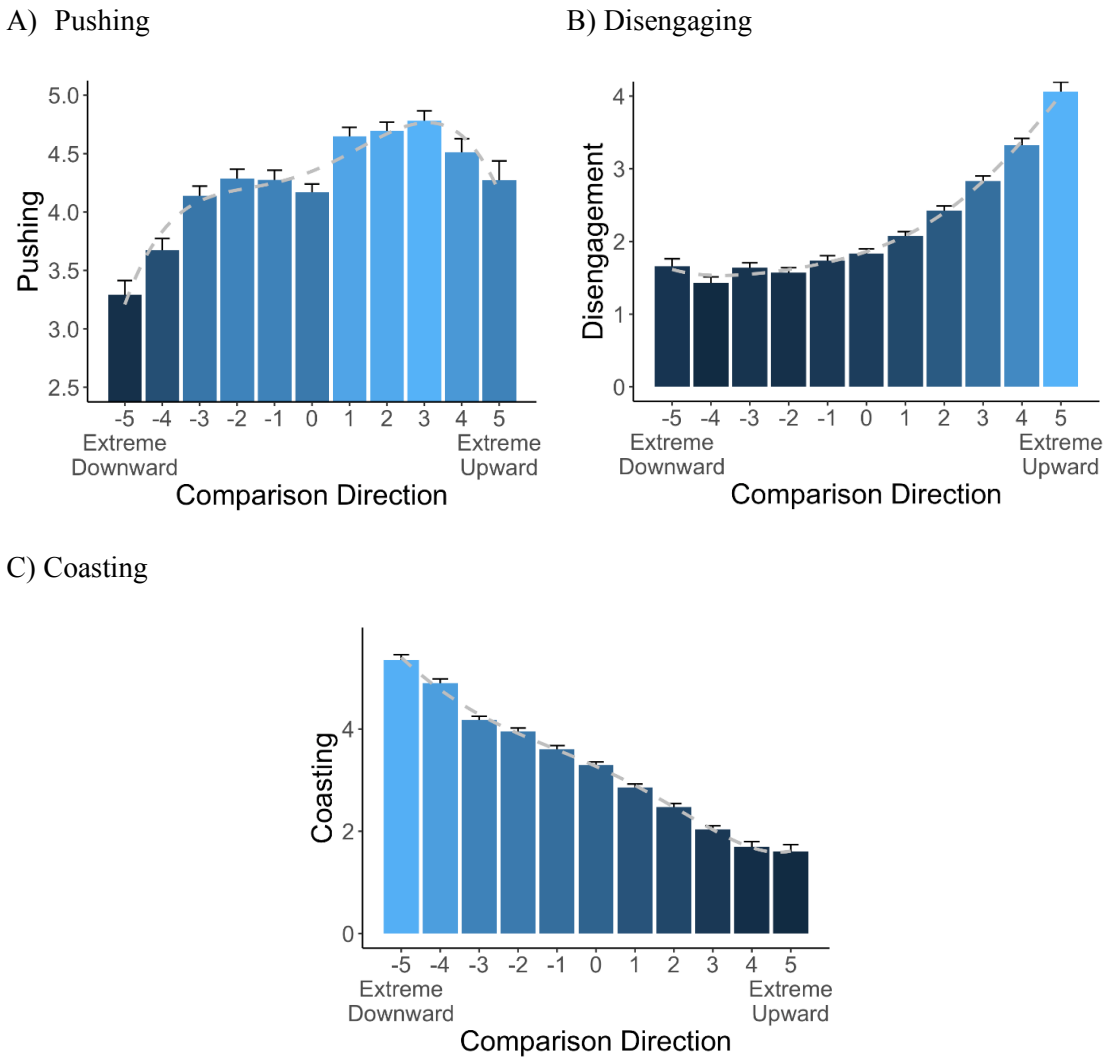


Figure 2. Effects of comparison direction on A) pushing, B) disengaging, and C) coasting. The dotted grey lines show the estimated curves from the multilevel polynomial regression analyses.

3.3.3 Moderating Effects on Motivation

We further aimed to investigate potential moderators that are associated with increases or decreases of the motivational effects of upward and downward comparison. For instance, people may derive more pushing motivation from an upward standard if they believe to have high control over the domain of interest, or if the domain is important to them.

Domain Control. We investigated motivational consequences of comparison direction depending on whether people perceive to have high or low control in their comparison domain. We

predicted that upward comparison is linked to amplified effects on pushing for people high on control compared to people low on control (H3). In order to evaluate the potency of domain control as a moderator in modulating the effect of comparison result on pushing, we compared the deviance of the moderator model with that of a (nested) base model without the interaction terms via the χ^2-distribution. The moderator model includes all previously tested polynomial components and their interaction with domain control (See Supplementary Table 1 for all main and interaction effects of both base and moderator models). This comparison revealed that the interaction model provides a significantly better fit than the base model without interaction terms, $\chi^2(4) = 23.23$, $p < .001$. In addition, we conducted a *region of significance* test to determine the range of comparison direction scores where the moderator elicits statistically significant slopes. Because the region of significance test in a sense "summarizes" the net effects of all polynomial terms, we chose a conservative p of < .001 to highlight those regions with pronounced simple slopes of the respective moderator. As illustrated in Figure 3a, and supported by the regions of significance test, the positive association between comparison direction and motivational pushing is stronger for people with high domain control (compared to low domain control) across most of the social comparison continuum, except for the very strong downward and upward comparisons. The difference was descriptively strongest for moderate upward comparisons.

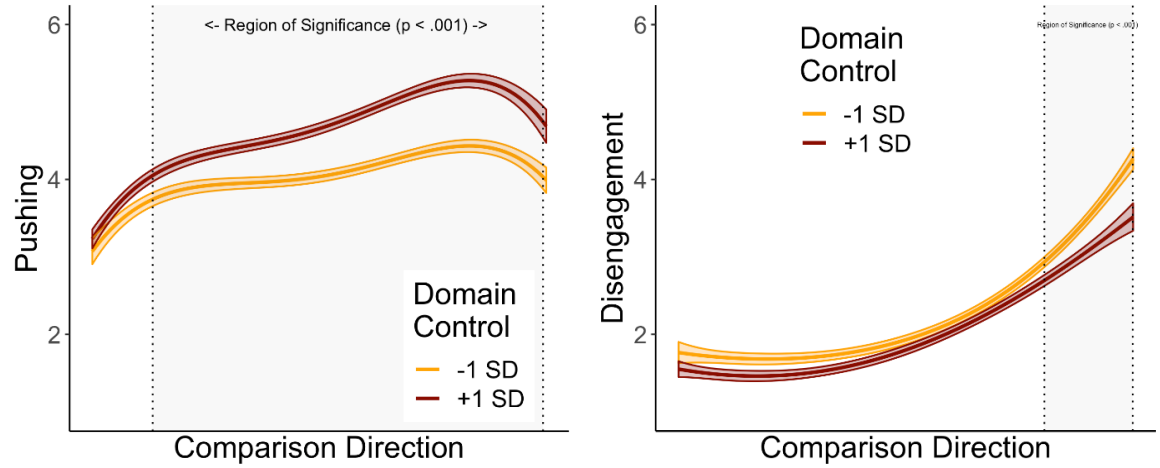
We conducted the same analysis with disengagement as outcome variable. Again, we compared the deviance of the moderator model predicting disengagement with a base model without interaction terms. Results demonstrated that the interaction model provides a significantly better fit than the base model, $\chi^2(4)=10.77$, $p=.029$. According to the region of significance test, it appears that the effect of more extreme upward comparisons on increased disengagement was more pronounced for people low in domain control over people high in domain control (see Figure 3B).

Lastly, we tested the interaction of domain control and comparison direction on motivational coasting. These analyses demonstrate a significant better model fit of the interaction model compared to the base model, $\chi^2(4)=14.01$, $p=.007$. According to the region of significance test, the positive relationship between downward comparison and motivational

coasting was more pronounced for people high in domain control over people low in domain control, especially when engaging in moderate downward comparisons (see Supplementary Figure 3).

In sum, people who perceive they have high control in their domain of comparison, also seem to derive more motivation from upward comparison and to be less likely to disengage motivationally with more extreme upward standards than people who perceive to have low control. On the other hand, people with high control also show a higher tendency of motivational coasting with downward standards than people with low control.

A) High vs. low control (pushing) B) High vs. low control (disengagement)



C) High vs. low importance (pushing) D) High vs. low importance (disengagement)

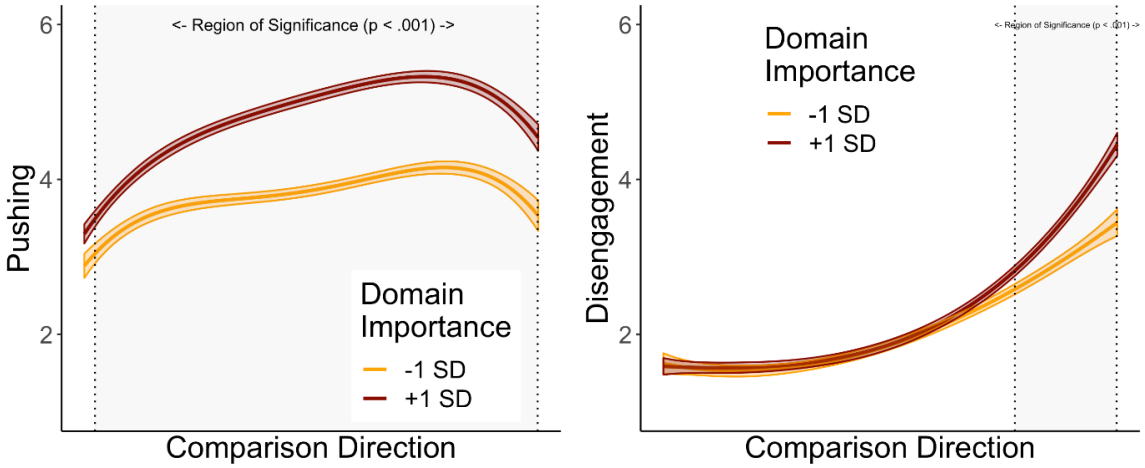


Figure 3. Interaction effects of comparison direction on pushing and disengagement as a function of high and low domain control (Panels A and B) and of high and low importance (Panels C and D). The grey areas determine the regions of significance at $p < .001$.

Domain importance. We investigated whether motivational consequences were differentially associated with comparison direction depending on how important participants perceived their comparison domain. We predicted that the positive association between upward comparison and pushing would be more pronounced if the comparison domain was important for participants (H4). First, we compared the deviance of the moderator model including domain importance with that of a base model without interaction terms using the χ^2 distribution. Results revealed a significant better fit of the interaction model compared to the base model, $\chi^2(4)=34.15, p < .001$. As illustrated by Figure 3c and the range of significance test, the positive relationship between upward comparison and motivational pushing is stronger if the comparison domain is of high importance compared to low importance across most of the comparison direction continuum, except for the very extreme downward comparisons.

Second, the interaction with domain importance predicting disengagement significantly improved the model fit when comparing the moderation model to the base model, $\chi^2(4) = 31.27$, $p < .001$. As demonstrated in Figure 3d, and by the range of significance test, the effect of increased disengagement with upward comparison was more pronounced for people high in domain importance over people low in domain importance especially towards the more extreme upward comparisons.

Lastly, we tested the moderation effects of domain importance on motivational coasting. Results revealed a better fit of the moderation compared to the base model, $\chi^2(4)=10.62, p=.031$. According to the range of significance test, the moderator did not elicit statistically significant slopes at $p < .001$. However, as illustrated by Supplementary Figure 4, high domain importance is linked to lower motivational coasting in association with (extreme) upward comparisons compared to low domain importance and to higher motivational coasting in relation to more extreme downward comparisons (See Supplementary Table 2 for all main and interaction effects of both base and moderator models).

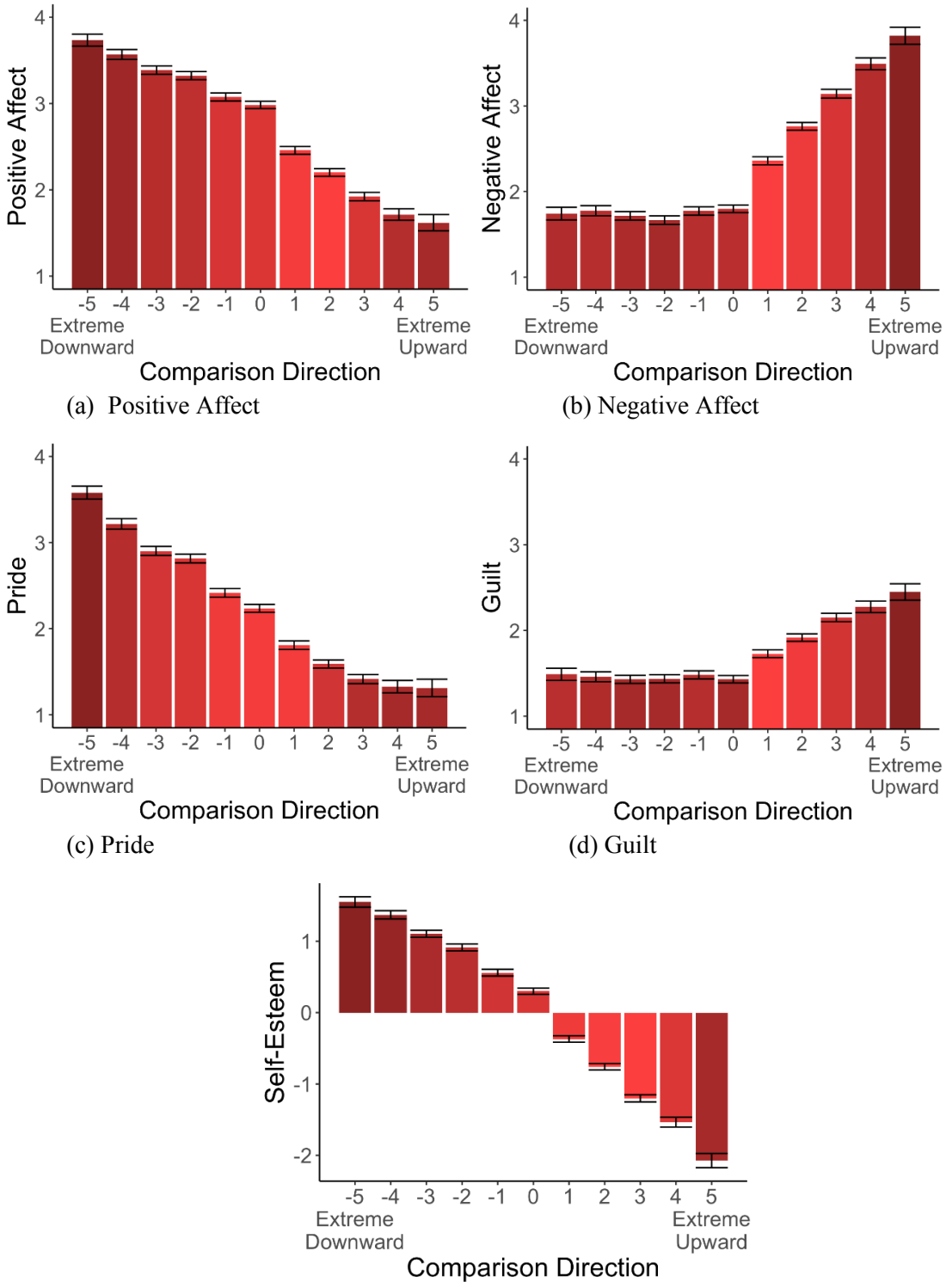
In line with the results on domain control, people with high domain importance also derive more pushing motivation from upward standards than people with low domain importance and seem to be more likely to engage in motivational coasting with more extreme downward standards than people with low domain importance. However, in contrast to high domain control, people with

high domain importance tend to disengage with more extreme upward standards than people with low domain importance.

Actively sought standards. We aimed to explore the effects of upward comparison on motivational pushing for people actively seeking comparison compared to people who were rather confronted with comparisons. Results revealed that the interaction model did not provide a significantly better fit than the base model without interaction terms, $\chi^2(4) = 4.14$, $p = .388$.

We repeated the analysis with disengagement as outcome variable. Comparing the interaction to the base model revealed that the fit difference just missed the cut-off for statistical significance, $\chi^2(4)=8.25$, $p=.083$. However, in the interaction model, the interaction between active comparison seeking and the linear component of comparison direction reached statistical significance, $B = -0.02$, $p = .043$ (see Supplementary Table 3). As illustrated in Supplementary Figure 5, the relationship between more extreme upward comparisons and disengagement seems to be reduced for people who actively sought rather than were confronted with a comparison. However, evidence for a significant moderation of active comparison seeking is only limited as the model comparison did not reveal a significant better fit of the interaction model.

Lastly, results did not reveal a significant difference between the moderation and the base model when testing the influence of active comparison seeking as a moderator between the effect of comparison direction on motivational coasting, $\chi^2(4)=3.55$, $p=.470$.



(a) Positive Affect (b) Negative Affect

(c) Pride (d) Guilt

(e) Self-Esteem

Figure 4. Effects of comparison direction on a) positive affect, b) negative affect, c) pride, d) guilt, and e) state self-esteem.

3.3.4 Social Comparison and Emotions

We hypothesized that people show more positive affect including pride and increased self-esteem, associated with downward comparison and more negative affect including guilt and decreased self-esteem, associated with upward comparison (H5).

Positive and negative affect. First, we regressed comparison direction scores on positive affect following the reported comparison. Results confirmed our prediction: There was a negative relationship of upward comparison and positive affect, $B = -.24$, $p < .001$. Additionally, a multilevel-regression with comparison direction as predictor and negative affect as outcome variable revealed a positive relationship between upward comparison and negative effect, $B = .21$, $p < .001$; Figures 4a and b.

Guilt and Pride. We predicted a positive relationship between downward comparison and guilt and a negative relationship between downward comparison and pride. Multilevel analyses confirmed our predictions: Upward comparison positively predicted feelings of guilt, $B = .10$, $p < .001$ and negatively predicted feelings of pride, $B = -.26$, $p < .001$; Figures 4c and d.

State self-esteem. We predicted that self-esteem increases with downward comparison and decreases with upward comparison. A multilevel regression analysis revealed a negative relationship between upward comparison and state self-esteem, $B = -.37$, $p < .001$. As seen in Figure 4e, self-esteem steadily increases with downward and decreases with upward comparison.

3.3.5 Base Rates of Comparisons

We hypothesized that participants more often compare upward than downward (H6). Results showed that 41.1 % of reported social comparisons were upward and 42.2 % downward comparisons. The remaining 16.4 % were lateral comparisons - meaning, comparing to a standard with a similar standing. The average level of comparison direction was -0.16 ($SD = 2.5$) on a scale from -3 (extreme downward comparison) to 3 (extreme upward comparison); Figure 5. A Pearson's Chi-square test comparing the observed frequencies to the expected frequencies did not reveal a significant

difference between upward and downward comparisons, $\chi^2(1)=1.14, p=.286$, $V=.02$. Hence, contrary to our prediction, we did not find an overall preference for upward over downward comparison and it appears that the proportions of upward and downward comparisons are equally balanced in daily life.

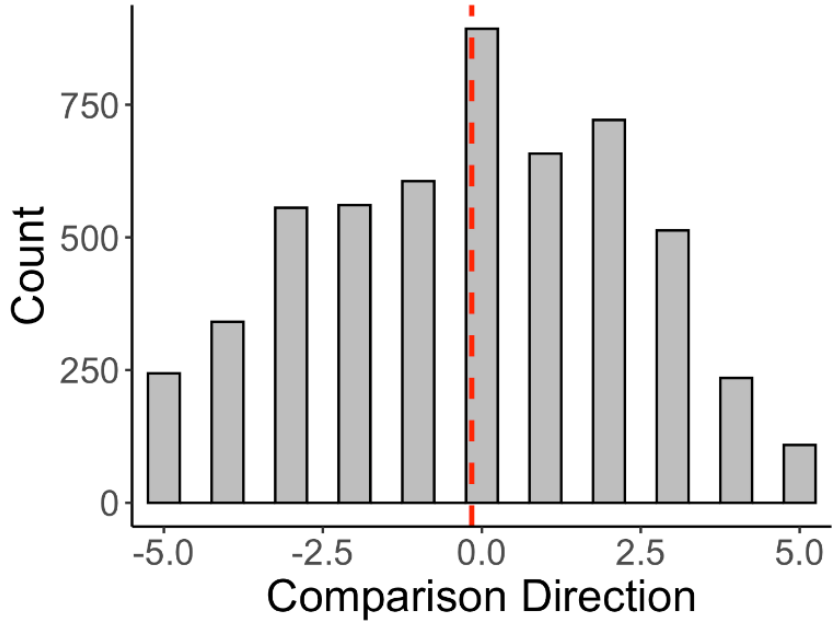


Figure 5. Distribution of comparison direction. Negative values on the x-axis represent downward comparison and positive values represent upward comparison. The red line indicates the grand mean.

Triggering conditions: fortunes and misfortunes. We preregistered that we aimed to explore antecedents of social comparison processes, such as experienced successes or failures prior to the social comparison. We added fortunes and misfortunes before comparison to a multilevel random intercept model: The experience of fortunate events negatively predicted upward comparisons (hence, positively predicted downward comparisons), $B=-.20, p<.001$ and the experience of misfortunes positively predicted upward comparisons, $B=.33, p<.001$.

3.4 General Discussion

People turn to their social environment in order to compare themselves and their current achievements to those around them. The present research provides new insights on the interplay between social comparison and motivation in people's everyday life. Our main contributions are

twofold. From a theoretical perspective, integrating the literature on social comparison more tightly with self-regulation theory provides a novel framework for thinking about the functionality of social comparison as part of people's everyday goal pursuits. As we outline in more detail below, the present set of findings seem to suggest that the time may be ripe to de-emphasize overly specialized and overly cognitive theories of social comparison towards a radically simplified re-conceptualization of social comparison as part of more general principles of self-regulation. From a methodological perspective, we go beyond earlier work in significant ways by capturing a wide range of possible comparison standards within a high number of possible domains, where past research has mostly focused on the effects of social comparison by manipulating social comparison in categorical ways. The high resolution gained by our methodological innovation allowed us to go beyond past work and provides a more comprehensive understanding of how comparison discrepancies relate to motivation and emotion.

Moreover, the sample was deliberately chosen to be heterogeneous in nature, with a relatively large share of non-student participants and a relatively high age range. Together, this endeavor yielded highly generalizable estimates of social comparison effects on key motivational and emotional indicators from our framework that are not only highly representative of people's everyday social comparison environments, but also appear to be generalizable to a typical Western population, across content domains and across other aspects such as whether people actively or passively sought out social comparison contexts.

3.4.1 Needed: A New Motivational Look at Social Comparison

Our findings contribute to cumulative theoretical knowledge with regard to the motivational and emotional effects of social comparison in particular. First, a purely cognitive look at social comparison, as advanced in large parts of the prior social comparison research likely misses a deeper understanding of how social comparison is ultimately "for doing", that is, embedded in broader processes of goal pursuit. As we have argued in our introduction, the "big three" motives in social comparison research, self-evaluation, self-improvement, and self-enhancement may be best understood as representing different phases of a classical self-regulation feedback loop: discrepancy assessment, investment of effort ("pushing"), and reduction of effort ("coasting"),

respectively. In addition, we have clarified how disengagement ("giving up") fits into this larger picture as the probable self-regulatory response to standards that are too far out-of-reach. As our estimated trajectory across the empirical range of social comparison shows (Figure 2), social comparison effects on motivation are strongly consistent with such a broader self-regulatory framework, with increasing invigoration up to a certain point in upward comparison (the maximally effective standard), but also clear evidence for disengagement for increasingly extreme upward comparisons, and with increasing levels of coasting as people compare further and further downwards.

Second, theories of social comparison that treat motivation as largely separate or non-central, may risk studying only parts of the story. We would like to provide two brief illustrations thereof. As a first example, as mentioned in our introduction, the social comparison literature has been dominated by a focus on comparative judgments rather than the behavioral implications of social comparison. This largely experimental literature has revealed evidence for assimilation vs. contrast in judgment as a function of standard extremity (Mussweiler, 2003). Scholars of social comparison of this body literature are likely to expect similar assimilative and contrastive effects of social comparison on behavior resulting from social comparison. However, the present pattern of findings show that what is known from similarity and dissimilarity judgments does not at all translate or generalize onto the motivational dimension. Specifically, nowhere in Figures 2a-c is there any indication of a motivational boost (or reduction in disengagement or coasting) at rather extreme downward comparisons. Rather, assimilation and contrast may be best seen as merely influencing people's subjective perception of self-other discrepancies at the self-evaluation/self-assessment stage. How such discrepancies then translate into tangible motivational effects such as effort investment or emotional consequences is an entirely different matter, one best told, we believe, through the lens of motivational frameworks of self-regulation rather than through principles of assimilation or contrast in judgment.

As a second example, we feel that the neglect of motivation as a primary lens through which to study social comparison has led to a problematic focus on the self-esteem implications of downward social comparison. As reviewed above, this body of literature maintains that downward social

comparison serves self-enhancement (Wills, 1981). In a nutshell, people want to feel good about themselves and selectively seek out downward social comparison standards to satisfy this inner need. Whereas we do not seek to challenge the evidence for self-enhancement as a universal theme in social psychology (Sedikides et al., 2003), we believe our results suggest to view the increase in state self-esteem, which we also found here as embedded in principles of goal pursuit (rather than a self-enhancement need-satisfaction "response" *per se*). Rather, from the present motivational perspective, the boost in self-esteem following downward social comparison can be easily re-conceptualized as positive self-regard ensuing from the perception that one has surpassed a self-relevant standard in goal pursuit. In other words, a boost in state self-esteem may be the emotional readout of a "coasting" response signaling that enough progress has been made (for now), a perception which social comparison often helps to validate. Our interpretation of the present results in terms of "coasting" rather than "boasting" is also supported by the correlated increase in pride, as a marker self-conscious emotion of success in goal pursuit.

3.4.2 The Motivating Effects of Social Comparison (and their Limits)

Third, the present research provides fine-grained and generalizable evidence that upward comparisons are only inspiring up to a certain point, and thus significantly qualifies and extends previous research, which has predominantly focused on positive motivational effects from upward comparison (Lockwood et al., 2005; Lockwood et al., 2005; Lockwood et al., 2004; Taylor & Lobel, 1989). There has been limited research from the field of educational psychology that already found discouraging effect of upward standards, but with a focus on a group of excellent exemplars that clearly stand out from a comparison group of weakly performing exemplars (Cho, 2019; Rogers & Feller, 2016), and thus bias the perception of the norm in the environment (Duckworth & Yeager, 2015; Heine et al., 2002) that is difficult to meet. Our findings show that the discouraging effect of rather extreme standards is not limited to individual domains, such as learning environments and does not require a change of norm in the current social environment, but also occurs with a comparison to a single "out-of-reach" standard.

The general conclusion of the dynamic motivational potential of upward social comparison notwithstanding, our moderator analyses provide an even more precise understanding of the

bounded nature of upward comparisons. Under specific conditions, the motivational potential of upward standards appears to be strengthened whereas it is thwarted under others: Most centrally, we found that participants who perceived higher control in their domain of comparison (e.g. fitness) drew more pushing motivation from upward standards and were less likely to disengage when comparing to upward standards than participants who perceived lower control. In a similar way, people who consider the domain of comparison as more important demonstrated higher motivation from upward comparison than people who regard the domain as less important. That motivational pushing effects were more pronounced for both conditions of higher domain control as well as domain importance strongly supports a self-regulation perspective on social comparison, as advanced here, and is an important step in linking social comparison effects with general principles of goal pursuit, including control perceptions or self-efficacy (Bandura, 1999) as well as goal importance or goal commitment (Fishbach & Ferguson, 2007; Kruglanski et al., 2002). Likewise, the finding that motivational disengagement at more extreme levels of upward comparison becomes considerably more likely with low domain control is consistent with the general idea that perceived control and goal commitment are important inputs into people's decision to keep persisting at a given goal versus invest their efforts in alternative pursuit (Kruglanski et al., 2002; Richter et al., 2016; Wrosch, et al., 2003). Past experimental research into the motivating effects of upward social comparison has assembled important but more or less fragmented pieces of evidence that upward standards have to be perceived as attainable in order to be encouraging (Lockwood & Kunda, 2000, 1997) or that role models only serve their inspirational role when their success is perceived as attainable and controllable (e.g. not due to luck; Morgenroth et al., 2015). We believe that these and other findings can easily be accommodated and integrated in a self-regulation framework of social comparison with the help of just a few key boundary dimensions such as the notion of controllability and commitment. Our work suggests the radical and simple conclusions that, by and large, principles of social comparison may essentially follow the same principles as ordinary goal pursuit—the only difference being that standards are activated by the social context. Rather than devising special, fragmented theories of social comparison and emphasizing distinctness to other areas, such a parsimonious framework builds common ground and emphasizes commonality towards an ultimately

more generalizable science.

Fourth, our research offers new insights into the motivational potential of downward standards. In contrast to (moderate) upward comparisons, inferior standards seem to harm motivation and effort investment. As predicted, downward comparisons were positively related to motivational coasting: With increasing positive discrepancies people are more likely to rest on their superior standing. Moreover, people who perceive to have high control over the comparison domain or perceive the domain as important, are more likely to rest on their current achievement when comparing to an inferior standard. Also here, we close a gap in the social comparison literature, which has mostly focused on affective outcomes of downward standards (Morse & Gergen, 1970; Taylor & Lobel, 1989; Wills, 1981; Wood et al., 2000). Limited past research, which has linked downward comparison and motivation found that prevention-focused individuals or individuals who perceived the standard as a possible future self actually benefit from downward comparisons (Lockwood et al., 2005; Lockwood et al., 2004). Similarly, ego-oriented individuals (i.e. believing that success is an indicator of ability) were more likely to prefer inferior standards compared to task- oriented individuals, which was mediated by decreased self-improvement motivation (Bounoua et al., 2012; Park & Park, 2017). Nevertheless, across many comparison domains and a broad range of comparison standards, we find negative motivational effects of downward comparisons in general: If one has already outperformed a standard, pushing motivation and effort investment drop and individuals tend to rest on what they have already achieved. This does not mean that downward comparisons may not be inspiring under certain circumstances, but suggests that the above findings may rather be seen as the exception from the general rule.

3.4.3 Discrepancy Assessments and Affect

To complement the picture of the functionality of social comparison in everyday life, we additionally provided observations on how discrepancy assessments between the self and a comparison standard influence people's affect. Needless to say, motivation and emotion are intricately linked (Carver & Scheier, 1981, 1990, 2004), and this was also clearly the case in the present results. In line with past research, upward comparison was positively related to negative affect, decreased self-esteem and feelings of guilt (Carver & Scheier, 1990, 2004; Morse & Gergen, 1970).

On the other hand, downward comparison was positively related to positive affect, increased self-esteem and feelings of pride (Carver & Scheier, 1990, 2004; Morse & Gergen, 1970; Taylor & Lobel, 1989; Wills, 1981). At the same time, our exploration of triggering conditions showed that successful events prior to comparison increased the likelihood of downward comparison, whereas failures were more likely to lead to more upward comparisons. As already mentioned above, our results do not support downward comparison theory, which states that people especially compare downwards when feeling threatened to boost or maintain their self-esteem (Hakmiller, 1966; Wills, 1981), but are in line with previous research investigating social comparison in everyday life where threat was associated with a higher frequency of upward, but not downward comparisons (Wheeler & Miyake, 1992; Gerber et al., 2018).

It is important to note that whereas motivational and emotional effects were clearly interlinked, the motivational and emotional effects stemming from discrepancy assessments were not mutually redundant but rather empirically distinguishable. Whereas motivation decreased with more extreme upward standards, negative affect and guilt continued to increase with more extreme upward standards. Hence, people start to disengage in a motivational but not emotional sense. On the other side, positive affect, self-esteem and pride do not drop towards the more extreme ends of downward comparison. For instance, one would expect, that the highest gains in pride stem from outperforming moderately rather than extremely inferior people. Smith's (2000) framework of social comparison-based emotion provides explanations for the linear nature of both positive and negative affect (including pride and guilt) as opposed to the non-linear fashion of motivational effects. First, a negative discrepancy between the self and a standard that is perceived as unchangeable, triggers negative emotions and this effect is even more pronounced if the comparison domain is of high importance or the comparison standard is perceived as similar on a relevant dimension. Thus, with growing negative discrepancies, negative affect steadily intensifies (e.g. guilt). On the other hand, if perceived control to reduce the gap between the self and standard gradually increases, negative affect decreases. Similarly, with increasing superiority (i.e. downward comparison), the perception of control grows which then translates into positive emotions (e.g. pride) (Major et al., 1991; Smith, 2000). Positive emotions signal that one is in a favorable position, which is associated with coasting behavior (Carver, 2003;

Thürmer et al., 2019). In case of a rather small discrepancy to a downward standard, control is perceived as low and negative emotion arise. This emotion is negative in nature (e.g. worry and fear), because one might share the same fate with the comparison person in the future (Major et al., 1991; Smith, 2000). Again, this supports our notion that increased positive emotions triggered by downward comparison are not necessarily the ultimate goal (as represented by the motive of self-enhancement), but also serve important functions as part of goal pursuit, namely, signaling sufficient progress relative to a standard in the environment.

3.4.4 Practical Implications and Future Directions

Upward comparison in daily life can inspire people, but only up to a certain point. If a comparison standard is too far off from the individual's standing, the comparison to more successful others can backfire and motivation turns into disengagement and effort withdrawal. Still, people compare to rather extreme upward standards in their daily life all the time, a trend presumably strengthened through social media (e.g. Appel et al., 2015). The permanent comparison to superior others is also related to people's drive of self-optimization: new year's resolutions are made, most recent App developments help to stay organized and simultaneously offer to share every progress on social media, which in turn facilitates comparison to even more successful others. However, the relentless pursuit of positive emotions can lead to the paradoxical effect of causing in fact unhappiness and disappointment: By setting too high standards for their own happiness and by choosing ineffective strategies to achieve joy (Ford & Mauss, 2014), all of the sudden people achieve the opposite. The same may apply to social comparison and insisting on overambitious goals. According to our research, social comparison should be used strategically to facilitate self-improvement by focusing on those people who are more successful but whose success is attainable at the same time. This way, their motivating potential is fully exploited and negative emotions that also stem from negative discrepancies are kept to a minimum. Future research should examine to what extent social comparisons can actively and strategically be used in everyday life in order to bring out the best in oneself without suffering the negative emotional consequences from upward comparison. Future research could also provide interventions studies inducing selective upward- and downward social comparison orientations, which additionally differ in their levels of discrepancies to the

standards.

3.4.5 Limitations

Although the methodology of experience sampling has the advantages of exploring social comparisons in numerous areas of everyday life, important limitations remain.

Firstly, it probably does not capture all social comparisons, especially those that may be automatic and unconscious (e.g. Mussweiler et al., 2004). Nevertheless, the method still allows one to collect both intentional and unintentional social comparisons (e.g. by browsing through a magazine) in contrast to experimental studies that often only test one or the other (e.g. reaction vs. selection studies, Gerber et al., 2018; Wood, 1996).

Interestingly, and quite remarkably, our moderator analyses regarding active comparison seeking suggest very little differences of intentional versus unintentional social comparisons per se, strengthening the generalizability of findings in this regard, and suggesting that past theoretical distinctions in this direction (e.g. Wood, 1996) may lack a clear empirical confirmation.

Second, we chose to only assess self-other discrepancies in a *relative* manner rather than to separately assess (absolute) self- vs. other-standing on a given dimension of interest as the two components that give rise to the discrepancy. Moreover, in line with the well-known compromise between (enhanced) ecological validly and (reduced) internal validity in field experiments as compared to experimental approaches, the experience-sampling method did not allow us to make use of the experimental technique to hold self-standing on a given dimension of interest *constant*. This limitation makes it difficult to clearly decipher whether a given self-outcome reflects high self-standing or poor other-standing and vice versa. To examine the impact of this limitation, we approximated the statistical control approach in laboratory studies: As shown in Supplement 4, we identified self-perceived domain competence as a proximate of the "self" outcome and re-analyzed all statistical models predicting the relationship between social comparison and motivation controlling for the domain of competence. The results show that above-average comparison outcomes are driven, on average, by the worse outcomes of others and that below-average comparison outcomes are driven, on average, by the better outcomes of others. Thus, adding domain competence as a control did not qualify any of our statistical conclusions regarding

pushing, coasting and disengagement. These supplementary analyses suggest that the effects from tightly controlled past experimental work may generalize well to the intense sampling of comparison outcomes in everyday life settings, in which such tight controls are not possible. However, future research should more directly measure and statistically control for the participants' self-standing than we could accomplish here.

Third, in order to reduce the multiple daily questionnaires to a minimum of questions, constructs are often measured by a single item only. This restriction was necessary to keep the protocol at a manageable completion time, and thus presents an inherent feature of the intense longitudinal design pursued in this work. This design compensates for the reduction in reliability of single item measures by the large degree of statistical power gained from the repeated-measures nature of the data (Bolger & Laurenceau, 2013). Fourth, psychological measures in the current study rely on self-report and may therefore be distorted by social desirability issues (Scollon et al., 2009) or biased self-perceptions. For instance, people see themselves as more favorable (Alicke & Sedikides, 2011) and thus, may also perceive the deviances from successful others as smaller than they actually are. However, the repeated measures nature of the experience-sampling method may buffer some of these issues as the method derives its power primarily from the within-person variation, i.e., relative differences from comparison context to comparison context, whereas personality differences such as social desirability concerns operating at the person level may be less of an issue.

3.4.6 Conclusion

Social comparison is among the most fundamental aspects of our everyday social lives. In this paper, we have argued that, to arrive at a truly comprehensive understanding of social comparison, we need to view the phenomenon through a motivational rather than overly cognitive lens. The present study and framework is one crucial step in such an integrative direction. We believe that a self-regulatory perspective on social comparison allows to see commonalities among seemingly disparate phenomena (e.g., coasting and self-enhancement), to interlink motivation, emotion, and cognition in functional terms, and will ultimately provide us with a better theoretical as well as practical understanding of *when* social comparisons are motivating and when they are not.

Chapter IV: Motivational and Emotional Effects of Social Comparison in Sports

This chapter is based on an earlier version of a manuscript of the following publication:

Diel*, K., Broeker*, L., Raab, M., & Hofmann, W. (2021). Motivational and Emotional Effects of Social Comparison in Sports. *Psychology of Sport and Exercise. 57,* 102048. https://doi.org/10.1016/j.psychsport.2021.102048 (published online: August 2021)

*First shared authorship

Abstract

The comparison to other athletes and the resulting effects on emotion, motivation and effort investment are a natural and integral part of sports performance. The current study tested a bias toward upward comparison in athletes. Further it tested how comparison processes influence motivation (i.e., self-improvement motivation, coasting, disengagement), emotion (e.g. happiness and shame) and performance improvement as a behavioral measure of effort. Freshmen from an elite sport university volunteered to participate in an experience sampling study conducted over one semester (6 months). Contrary to our predictions, athletes did not generally compare more upward than downward, and only few subscales of the two sport-specific dispositional measures (sport-specific achievement motivation and sport orientation) predicted upward comparison. As hypothesized, upward comparison to moderately better standards was associated with greater motivation while extreme upward comparison was related to a drop in motivation and increased disengagement. Still, upward comparison during the semester significantly predicted performance at the end of the semester. Downward comparison was related to coasting motivation and lower levels of performance. Happiness decreased with upward and increased with downward comparison. The opposite was true for feelings of shame. This research emphasizes the role of the social environment in sports and how training partners depending on their inferiority or superiority can boost or hinder motivation and performance in athletes.

4.1 Introduction

In everyday life athletes frequently focus on self-improvement. However, to excel in sport athletes must also focus on outperforming others. It is therefore inevitable that athletes often do just this and adjust their efforts accordingly (e. g. Sheldon, 2003; Williams, 1994). Comparing oneself to better athletes likely motivates greater effort investment and athletic performance, however, will this also result when the athlete compares oneself to one who performs at a *much* higher level? Also, can comparing oneself to underperforming athletes still give a push in the right direction? Previous research investigated how a wide range of possible comparison standards (extremely worse to extremely better) influenced motivation and effort intentions, and thus provides tentative answers to these questions (Pushing, Coasting, and Disengagement; Diel et al., 2021). In the present article we go a step further by investigating the motivational and emotional effects of social comparison in a sample that is likely to be biased towards upward comparisons - freshmen students from a sport university who regularly participate in competitive sport. We also examine whether frequent athletic comparisons promote *actual* performance and thus whether, for instance, upward comparisons are central in this regard.

4.1.1 Social comparison in sports

Comparing to social others occurs naturally in daily life (Festinger, 1954). People tend to compare to similar others (lateral comparison) as a source of self-evaluation (Festinger, 1954; Taylor et al., 1996), superior others (upward comparison) as a source of motivation (Taylor & Lobel, 1989), and inferior others (downward comparison) for a boost in self-worth (Morse & Gergen, 1970; Taylor & Lobel, 1989; Wills, 1981). As we argue in this article, athletes striving for success are especially prone to upward comparison (Gerber et al., 2018; Lockwood et al., 2005). Festinger (1954) posited that when people evaluate their abilities, they have a "unidirectional drive upward" (p. 12). As a motivational force to improve performance, this drive characterizes competitive behavior (Wetherall et al., 2019). In competitive sports "doing better" translates not only as "doing better than last time" but also as "doing better than *others*". Self-reference comparisons, i. e. comparing with one's own past performance, are thus complemented by comparisons to relevant and similar others like competitors and team members. Using others as reference standards is at the core of competition in sports and

physical education (Walton et al., 2020; Xiang et al., 2020). Here, competitive athletes pay regard to better-performing athletes to source salient information on how to improve their own performance (Gotwals & Wayment, 2002). Hence, we assume that athletes are especially prone to compare upwards. Empirical evidence strengthens this viewpoint – intercollegiate athletes were found to frequently make upward comparisons and to infrequently make downward or lateral comparisons (Gotwals & Wayment, 2002). Further, the likelihood of comparing oneself to others who are better at maintaining physical activity increases with greater perceived control over one's own ability to be active (Caltabiano & Ghafari, 2011) and with confidence in one's own abilities (Feltz & Lirgg, 2001; Garcia et al., 2007). This way, upward comparisons pose a challenge by continously "setting a higher bar", a tendency which also conforms with competitive sport culture and athletic attitudes (e.g., the Olympic motto *faster, higher, stronger*; (Committee, 2021). With regards to dispositions, Williams (1994) reported ego-oriented high school athletes, i.e., those athletes oriented towards others' performance, were more likely to use their peers as a source of competence information rather than the self. Also, Harwood and Swain (1998) report ego-orientation and the coherent perception of significant others to predict sport orientation, specifically goal-involvement in competitive contexts. Beyond that, ego-oriented athletes with low perceived ability generally exhibit reduced effort and goal pursuit (Duda & Nicholls, 1992). In addition to unquantifiable self-other discrepancy assessments of physique, skill, etc., objective information like competition outcomes or personal bests can also influence how one compares (e.g., upwards, etc.) and thus to what extent one is motivated. For instance, Berger and Pope (2011) found basketball teams that were behind at half-time and compared upwards to the others teams score were more motivated after half-time and had a higher likelihood of winning the game (relative to teams that did not compare upwards). However, not every losing team or person is sparked with motivation and inspiration. For instance, it is highly probable the FIFA World Cup 2014 host nation Brazil lost hope of turning the tide against Germany when they were 5-0 down at half-time, thus leading them to ultimately lose 7-1. Being *far* outperformed can be devastating and will often tempt people to give up (Wergin et al., 2018). It is thus worthwhile to explicitly examine *how* athletes are fuelled by (upward) social comparisons, which can be inspiring but may also have a discouraging nature.

4.1.2 Pushing, Coasting, and Disengagement

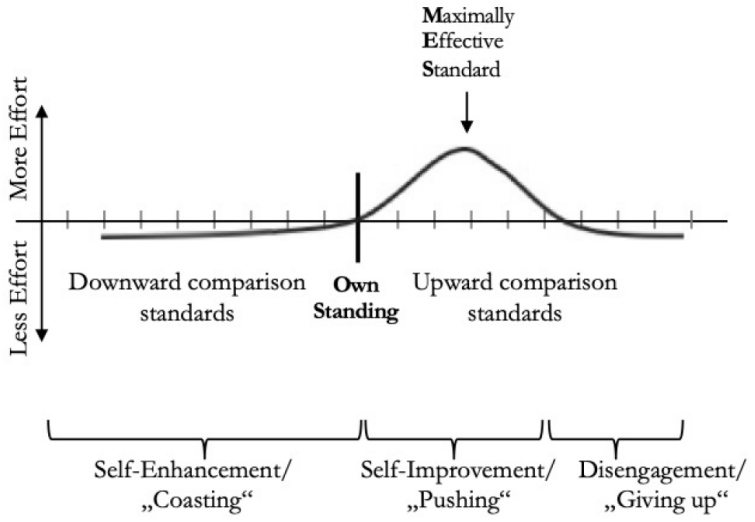
People generate feedback about themselves by repeatedly comparing their current performance to their own self-developed standard (e.g., identifying ones running pace as 16 km/hr, and comparing to one's self-set standard of 17km/hr). Once a discrepancy is detected, people generally need to initiate behavior and invest more effort to reduce it (Carver, 2004). To translate self-improvement motivation into action, people must often resist short-term temptation or impulses in favor of delayed long-term goal rewards (Hofmann et al., 2009, 2011; Kotabe & Hofmann, 2015). Notably, people often also compare their current state with *non*-self based standards (Diel et al., 2021). For instance, one may compare oneself with the lead runner in a race, and ultimately recognize a negative discrepancy. Similarly, comparing oneself with a slower runner is likely to yield a positive discrepancy. Whether a perceived discrepancy is positive or negative is ultimately likely to determine an athletes motivational response (Diel et al., 2021, see Figure 1). According to this theoretical framework, upward comparisons that result in a negative discrepancy are associated with self-improvement motivation and effort investment (*pushing* towards a goal). However, *extreme* upward comparisons to exceptionally superior standards, for instance, comparing to the world's fastest marathoner Eliud Kipchoge, would be expected to often lead to reduced effort and *disengagement* from goal pursuit. Why? Because the discrepancy will often be perceived as too large to minimize soon or without extensive effort. This effect has also been observed in sports - non-elite athletes exercising in a highly selective sports high school who compare themselves to elite athletes scored lower on physical self-concept variables (e. g. endurance, strength) than non-elite athletes comparing to general students (Marsh et al., 1997) – also known as the big fish/little pond effect (see Chanal et al., 2005; Marsh & Redmayne, 2016). The optimal motivational potential would therefore be triggered by an upward reference which is neither out of reach nor already achieved (Diel et al., 2021). This point is called the maximally effective standard (MES, see also Figure 1). Similar evidence with regards to a motivational optimum comes from re-interpretations of the Yerkes-Dodson Law. The original empirical, inverted U-relationship between arousal and performance can be translated to motivation in the context of sports and thus includes an optimal point of motivation for performance that is neither too high nor too low (for an overview see Teigen, 1994). In contrast, a downward comparison to a slower runner results in a

positive discrepancy and is also associated with a drop in motivation: *coasting* - relaxed effort and reduced motivation to improve, as the runner perceives her performance to be sufficient (Diel et al., 2021).

In the present article, we aim to show a bias towards upward comparisons in athletes and also investigate motivational effects and actual performance improvement of comparison processes, such as the motivational potential of upward comparison, but only if the discrepancy between self and the standard is perceived as moderate. We propose that very large perceived discrepancies will be related to feelings of giving up and disengagement. Lastly, we believe downward comparisons will also be associated with a drop in motivation ("coasting") as people rest on their current performance. Additionally, effort investment is assessed by actual performance improvement, which is expected to increase with upward but to decrease with downward comparison. With the experience of motivation, people also respond emotionally to different discrepancy assessment, which we investigate in the last part.

Figure 1

Discrepancy assessment and related investments and motivational states of self-improvement



Note. This is the hypothesized mapping of discrepancy assessment between own standing and relatively better or worse social comparison standards, change in effort investment (curve), and key motivational states of self-improvement ("pushing"), self-enhancement ("coasting"), and disengagement ("giving up") adapted from Diel, Grelle & Hofmann (2021).

4.1.3 Social comparison and emotions in sports

Competitive athletes generally experience a wide range of emotions. Some of the most common negative emotions are sadness, guilt, and shame (Cerin & Barnett, 2006), which can often result after competitions and comparisons. While emotions such as these decrease in intensity and frequency in the days following, they can still hinder performance when perceived as threatening to the self (Jones et al., 2009). Upward comparison or a detected negative discrepancy to a comparison standard evokes negative emotions, such as guilt, shame and a drop in self-esteem (Carver, 2004; Carver & Scheier, 1981, 1990). On the other hand, downward comparison or a positive discrepancy between self and comparison standard is related to positive emotion, such as pride, happiness and a boost in self-esteem (i.e. self-enhancement, Morse & Gergen, 1970; Wills, 1981). Emotional reactions in sports, especially in competitive settings, are slightly more complex and might not only depend on the comparison to a fellow sportsperson (e. g., opponents), but also on a final result. For instance, tennis players' subjective athletic competence decreased when they were defeated by inferior ranked opponents (downward comparison). Thus, the intensity of state self-esteem fluctuations not only depend on the opponent's level but also on the match result (Bardel et al., 2010). Gaudreau et al. (2002), however, reported greater performance-goal discrepancy can predict positive emotions but does not reliably predict negative post-competition emotions. Instead, studies demonstrated that the athlete's mere expectations of future performance influence whether positive and negative emotions result from it (McGraw et al., 2005). In general, having expectations lower than the actual outcome results in happiness and more positive emotion afterwards. However, social comparison outcomes influence emotion irrespective of objective success. For instance, athletes winning a bronze medal were happier than silver medalists who compared (upward) to the gold medal winner ("I just missed the chance to win the gold medal") – while the athletes winning the bronze medal compare to the person on the fourth place ("I reached the podium"), thus comparing downward (see also Allen et al., 2019). We therefore investigate whether the emotional effects of social comparison also apply to our sample of athletes – upward comparison being tied to negative emotions and downward comparison to positive emotions.

4.1.4 The present research

In the current study, we investigate social comparison processes in sports using a sample of sports student from an elite european sports university[10]. First, we examine a bias towards upward comparison by taking into account dispositional measures that typically characterize a group of athletes. For this purpose, we focused on three main measures that aimed to reflect athletes' overall competitiveness and motivation to be successful: their sport-specific achievement motive, sport orientation, and regulatory focus. The sport-specific achievement motive is concerned with the tendency to approach vs. avoid surpassing standards of excellence, striving for skill improvement (e. g., stamina or techniques), and challenges and competition (Elbe et al., 2003; Gröpel et al., 2016). Closely related to this, sport orientation is a multidimensional disposition describing the tendency to "strive for satisfaction when making comparisons with some standard of excellence" (i. e. competitiveness, Gill & Deeter, 1988, p. 191), as well as the desire to win interpersonal competition vs. to reach personal goals (win vs. goal-orientation respectively). Regulatory focus differentiates between focus on accomplishment and aspirations vs. focus on safety and responsibilites.

Research has found significant positive associations between the approach-achievement motive ("hope for success") and athletic performance, and a negative correlation between avoidance-motive ("fear of failure) and athletic success (Elbe et al., 2003; Thomassen & Halvari, 1996). Similarly, sport achievement orientation and its submeasure competitiveness was found to correlate positively with athletic performance (Hellandsig, 1998). Lastly, we focus on athletes' regulatory focus, which consists of two self-regulation strategies: promotion and prevention focus (Higgins, 1997). In sports, a regulatory fit – meaning their chronic regulatory strategy fits their to-be-accomplished goal - was found to increase athletic performance (Klatt & Noël, 2020; Memmert et al., 2009). Thus, we argue that individuals in sports who are dispositionally competitive or achievement-oriented are also more prone to upward comparisons, which leads us to our first set of preregistered hypotheses:

H1: We expect athletes to compare more frequently upward than downward.

[10] In order to study at the Sport University students have to pass a physical aptitude test including 20 individual performances out of which 19 have to be passed. The students tested in this sample therefore deserve to be called athletes. The current failure rate for the aptitude test is 55%.

H2: We expect hope of success as one facet of sport achievement motivation to positively correlate with upward comparison, and the facet fear of failure to positively correlate with downward comparison. Beyond that, we expect a positive correlation between competitiveness, win-orientation and goal-orientation and upward comparison. Finally, we expect a positive relationship between promotion focus and upward comparison, and prevention focus and downward comparison.

In a second part, we test how these comparisons fuel self-improvement motivation, coasting and disengagement as well as effort investment. Effort investment is assessed by actual athletic performance – assessing to what extent students improved their athletic performance by the end of the semester. This leads us to our second set of hypotheses:

H3: We predict that a) upward comparison is positively associated with improvement motivation, but only for moderately (i.e. attainable) compared to extremely (i. e. unrealistic, exceptionally superior) upward comparisons – we thus expect a nonlinear relationship with improvement motivation increasing with moderate upward comparisons and decreasing again with more extreme upward comparisons. We further predict b) a positive association between downward comparison and coasting motivation and c) a positive relationship between upward comparison and disengagement motivation with disengagement increasing especially towards the more extreme ends of upward comparison.

Lastly, we predict that athletic performance improvement is positively associated with upward comparison (H4).

Since comparison processes not only affect people motivationally, but also emotionally, we investigate associations between social comparison and emotions in the third and last part of the paper:

H5: We predict a positive relationship between downward comparison and happiness (5a) and between upward comparison and shame (5b).

For the purpose of the paper, we chose a ecologically-valid environment and applied an experience-sampling method in a sample of students enrolled at the largest sports university in Europe. This sample is special because students, almost all of them being active, competitive athletes, face comparisons with other elite and ambitious athletes day by day in sport-practical courses (Horn & Harris, 2002). The setting allowed us to test long-term effects of social comparison including to what

extent and with what effects people compare themselves with relevant comparison standards in an environment with specific and objectively quantifiable performance goals. Over the period of one semester, freshmen in track-and-field and swimming courses reported social comparisons after each weekly course unit. Simultaneously, students' current standing on their performance goals were frequently assessed over the course of the semester: their measures of running/swimming time were obtained at the beginning and end of the term providing us with an objective measure of their performance improvement.

Furthermore, a more frequent sampling of experienced emotions and motivation, for instance, by experience sampling or diary methods, enables the detection of dynamic fluctuations in comparison type and the ability to examine inter-individual variance, thus yielding a more complete picture of each individual. While measurements post events rely on memory and recall abilities and might produce a biased perception of past events, we measured comparison processes instantaneously after the comparison had taken place by providing students with the opportunity to directly answer weekly questionnaires with an app on their smartphone, thereby recall biases could be minimized. This research aims at better understanding the motivational benefits and drawbacks of social comparison processes in sports as the comparison to the "wrong" people for instance can hinder improvement goals (i. e. comparing to a weaker companion or a to someone who is out of reach). Simultaneously, we aim to show how upward vs. downward processes influence not only planned effort investment (e. g. self-report measure) but actual sport performance. This measure sets our research apart from other social comparison research by objectively quantifying the effects of upward vs. downward comparisons on people's performance improvement.

4.2 Method

The study represents the first and most original separation from a larger study, which included a variety of additional variables and hypotheses. In the present paper, we will only report material and results relevant to our research questions. All study materials, preregistration from the larger project, data and syntax from the present research are publicly available on the Open Science Framework: https://osf.io/bk9xa/?view_only=8b5a4ee12fa6420ebe33da764cd57533.

Because we did not have a focal hypothesis to base a power calculation on, our sampling strategy was to recruit as many students as possible to maximize power: as preregistered, we aimed at a sample size of 200 participants, with a minimum of 100 participants depending on recruitment success. Still, we conducted a post-hoc simulation power analysis with the *simr* package (Green & MacLeod, 2016) for R that runs Monte Carlo simulations designed for mixed models. Results indicated we had around 100% power to find a small effect size (d = .25) based on one of our main hypotheses on motivation (H3a) and around 71.4% power to detect a small effect size of .10 (for interpretation suggestions see Ferguson, 2009). The study was approved by the university's local ethics committee and the ethics commission of the German Psychological Society.

4.2.1 Participants

We recruited participants from an elite Sport University in freshmen classes, of which all were active athletes ranging from state division up to national team level across different sports. One hundred fifty participants were recruited for the study and 114 agreed to take part. After removing 10 participants who had not responded to at least one weekly signal, the final sample consisted of 104 students with an average age of 20.55 years (range 18 – 29, *SD* = 2.46) with 46 % male participants (1% no answer). From these 104 students, 76 completed all weekly signals, 17 completed more than 80% and the remaining 11 completed on average more than 55% of the weekly signals. Overall, participants reported 410 social comparison situations over the course of the semester. All students participated in a mandatory course of either swimming (n = 75, 51.5% female) or track-and-field (n = 29; 62.8% female). Student participants were directly addressed in their introductory unit of their course or were further recruited via flyers and the university's website. Participants who responded to at least one weekly signal during the mobile phase were included in the analysis. Participants received 5 Euros for completing the intake survey and additional 15 Euros if they answered at least 80% of the weekly signals. Each completed signal additionally counted as a lottery ticket with the chance to win one out of two 100 Euro rewards.

4.2.2 Materials and Procedure

We used an experience sampling method (ESM) with weekly questionnaires that have been

validated in former studies (Diel et al., 2021), see section Mobile Phase. The ESM is a research procedure that requires people to provide systematic self-reports about what they do, feel or think at random or scheduled time points during the waking hours of the week (for general validity of ESM see Bolger & Laurenceau, 2013a). ESM has been suggested to "overcome constraints of previous methods by combining the ecological validity of diary approaches with the rigorous measurement techniques of psychometric research" (Larson & Csikszentmihalyi, 2014, p. 23). In this study, ESM allowed us to capture emotional and motivation consequences resulting from comparisons immediately after comparisons had taken place, i.e. after the respective course unit, to ensure the least possible retrospective bias. First, prospective participants filled in a screening survey, testing whether they fulfill the three main criteria for participation: minimum age of 18 years, resident in Germany, and participating in one of the two practical courses of interest (swimming, track-and-field) during the current semester. If all requirements were met, participants received information about the procedure of the study, monetary compensation and how to receive the questionnaires (e. g. how to install the app). The study aim was introduced as investigating "social comparison in practical sport courses". Informed consent was obtained from all participants. If participants were enrolled in the swimming and track-and-field class at the same time, they had to pick one as a basis for the questionnaires. A full protocol can be taken from the supplementary material (S3).

Intake Form. After successfully completing the screening survey, participants proceeded with an intake survey, which consisted of several dispositional measures relevant to our preregistered research questions and exploratory analyses and demographic information.

Further, we assessed performance at the beginning of the semester in the course of interest: If available, students indicated how fast they swim 200 meters or how fast they currently run 100 meters[11]. In order to measure sport-related dispositional variables, we examined the achievement motive with the German version of the Achievement Motive Scale-Sport (AMS-Sport; Wenhold, Elbe, et al., 2008), with the components Hope of Success (HS, 15 items, $\alpha = .95$, e. g. "I want to be successful in what I do in sports"; 1 = *does not apply to me at all*, 4 = *applies completely to me*) and

[11] At the end of the semester, students had to fulfil a certain time requirement in order to pass the class (200 m swimming (men: 240 sec; women: 250 sec) or 100 m running (men: 13.3 sec; women 16.0 sec) depending on the course). Occasionally, the course leader took the time at the beginning of the course to inform students about their current level of performance.

Fear of Failure (FF, 15 items, α = .93, e. g. "I don't like sport situations in which my skills are being tested", 1 = *does not apply to me at all*, 4 = *applies completely to me*), and competitiveness (13 items, α = .94), goal orientation (6 items, α = .81) and win orientation (6 items, α = .82, e. g. "I want to be the best every time I compete", 1 = *strongly disagree*, 5 = *strongly agree*) with the Sport Orientation Questionnaire (SOQ; Elbe et al., 2008). In addition, we assessed regulatory focus (promotion, α = .73, and prevention focus, α = .80, 9 items each) with the Regulatory Focus Scale (e.g. "I often think about the person I would ideally like to become in the future", 1 = *strongly disagree*, 9 = *strongly agree*; Lockwood et al., 2002b ; for validation of German version see Schmalbach et al., 2017). Lastly, participants were asked to indicate their age and gender.

Mobile Phase. To complete the weekly questionnaires, participants chose between an app (movisensXS App, applicable only with Android operating system, movisens GmbH, Karlsruhe, Germany) or questionnaires sent via email (Qualtrics software). Directly after each class, participants filled out the weekly questionnaire via one of these distribution methods. Classes were taught once a week for a total of 14 weeks.

Motivational States. The mobile phase consisted of six blocks of which three were relevant to our research. In the first block, we assessed the motivational states pushing ("I am currently motivated to improve myself"), coasting ("I can currently rest on what I've already achieved"), and disengagement *("I would like to give up right now") on a 7-point Likert scale, 1 = not at all, 7 = very much*; Diel et al., 2021). Next, participants were asked if they had engaged in a social comparison during the recent course unit. If the answer was *Yes*, the questionnaire continued to the *Comparison Direction* block. If the answer was *No*, the questionnaire ended.

Comparison Direction. In this block, we assessed the direction of comparison – meaning how participants perceived themselves in terms of performance compared to the comparison standard ("How did you perceive *your* current performance in comparison to the other person?", -5 = *extremely worse*, -4 = *quite a lot worse*, -3 = *a lot worse*, -2 = *moderately worse*, -1 = *a little worse*, 0 = *similar*, 1 = *a little better*, 2 = *moderately better*, 3 = *a lot better*, 4 = *quite a lot better*, 5 = *extremely better*). For the analyses, the item was recoded for interpretation purposes, so that positive values represent upward comparison (the comparison standard is perceived as superior) and negative values downward

comparison (the comparison standard is perceived as inferior). For the main hypotheses regarding motivational and emotional correlates, the measure was treated as a continuous variable. For H1, comparing the frequencies of upward and downward comparisons, we coded responses categorically (i.e., above 0 = upward comparison, 0 = lateral comparison, below 0 = downward comparison).

Emotional States. Next, we measured additional motivational and emotional states in relation to the comparison standard (How motivated did you feel after the comparison? -3 = *much less motivated*; +3 = *much more motivated*); (How happy did you feel after the comparison? (-3 = *much less happy*; +3 = *much happier*); How proud or ashamed did you feel after the comparison (-3 = *very ashamed*; +3 = *very proud*)[12]. Participants repeated the last two blocks if they had compared to an additional comparison standard. The questionnaire ended with an open textbox for further comments (e. g. problems they had encountered).

Final survey. At the end of the semester, participants filled in a final questionnaire with a focus on their performance. Participants were first asked to again provide their swimming or running times as measured at the beginning of the semester. Second, if they could, they were asked to enter the time taken at the end of the semester, too. For the main analyses, we subtracted end time from start time and took this difference as an *objective* performance improvement score. Moreover, we measured *subjective* performance improvement (i.e. "How much has your performance improved during the semester?"; 1 = *not at all*, 3 = *moderately*, 7 = *quite a lot*). Finally, participants had the chance to provide further comments in an open textbox and to give feedback on the mobile App.

4.2.3 Analysis Strategy

As observations (Level-1, $N = 410$) were nested within persons (Level-2, $N = 104$), we conducted multilevel models with random intercepts and fixed slopes for all main analyses using the lme4 package in R (Bates et al., 2015) and ancillary packages (see R script on OSF project page). Level-2 variables were mean-centered, however, we decided against person-mean centered predictor scores for the Level-1 comparison direction variable (Enders & Tofighi, 2007) for interpretation

[12] In order to reduce the number of items, we decided to measure the two self-conscious emotion, shame and pride, on one continuum based on previous research that demonstrate their symmetrical structure (shame as a negative self-evaluation and pride as a positive self-evaluation; Taylor, 1985; Tangney, 2005).

purposes: The scores range from -5 to 5 ("How do you rate your performance compared to this other person in general?"; -5 = *extremely worse*, -4 = *moderately worse*, (…), 0 = *the same*, (…), 4 = *moderately better*, 5 = *extremely better*) and thus contain a meaningful point of zero.

Furthermore, for H2 we aggregated comparison direction scores to summarize the observation values from Level-1 on Level-2. This way, we were able to conduct a multivariate regression model where a number of dispositional variables predicted average comparison direction scores. Lastly, we conducted linear regression analyses to regress average comparison direction score on subjective as well as objective performance improvement (H4).

4.3 Results

4.3.1 Comparison Direction

We hypothesized that participants would compare more frequently upward than downward (H1). The average level of comparison direction was – 0.05 (*SD* = 2.15) on a scale from – 5 (extreme downward comparison) to + 5 (extreme upward comparison). For H1, we additionally coded comparison direction according to their general direction of comparison (< 0 = downward, 0 = lateral, > 0 = upward) to compare the frequency of upward comparisons to the frequency of downward comparison. Results revealed that 41.2 % of all comparisons were upward comparisons, 38.8 % downward comparisons, and the remaining 20 % lateral comparisons. Comparing the frequencies of upward and downward comparisons resulted in a non-significant difference, $\chi^2(1) = 0.31$, $p = .581$, showing that participants did not compare significantly more upward than downward.

4.3.1.1 Sport-Related, Dispositional Predictors of Comparison Direction

We investigated the effects of several sport-related dispositional variables on overall comparison direction tendencies. In general, athletes could be characterized as moderately achievement-motivated and sport-oriented. Scores regarding hope of success and fear of failure were comparable to the norm sample of the respective original validation study (see Table 1). Further, neither students' win-

orientation was comparable to national level-2[13] athletes, nor were competitiveness and goal-orientation comparable to scores from level-2 athletes and even further afield from international elite athletes (55.47 ± 5.47, and 26.70 ± 2.75, respectively) (Elbe et al., 2008).

Table 1

Descriptive statistics of the sport-specific dispositional predictors in comparison the original studies' norm values

Variable	Study sample		Norm sample		*t*	df	*p*
	M	*SD*	*M*	*SD*			
Hope of Success	34.32	6.46	34.31[a]	6.61	0.02	839	.988
Fear of Failure	12.95	7.75	11.15[a]	7.86	2.19	838	.029
Win-orientation	25.53	3.30	26.41[b]	3.43	2.50	1148	.012
Competitiveness	49.22	9.97	53.36[b]	7.36	5.28	1147	< .001
Goal-orientation	19.33	5.50	22.94[b]	3.43	9.87	1556	< .001

[a] Wenhold, Meier, et al. (2008)
[b] Elbe et al. (2008)

To test H2, we regressed Level-2 dispositional predictors on aggregated comparison direction scores (see Model 1.1, Table 2). Contrary to our predictions, hope of success negatively predicted upward comparison, and fear of failure positively predicted upward comparison. Both promotion and prevention focus did not account for average levels of upward comparison. In line with our predictions, win orientation as a subscale of the sport orientation questionnaire, positively predicted upward comparison, however, contrary to our predictions, the subscales goal orientation and competitiveness were negatively related to upward comparison.

[13] Elbe et al. (2008) use the label level-1 for athletes competing at national/international level (A/B/C squad, (prospective) Olympians, and the label level-2 for athletes competing on regional level (D-squad, state level competitions).

Table 2

Model 1.1. Regression results using comparison direction as the criterion.

Predictor	*b*	*beta*	sr^2	*r*	Fit
(Intercept)	0.01				
AMS					
Hope of Success	-0.03**	-0.10	.00	-.29**	
Fear of Failure	0.02*	0.09	.00	.24**	
Regulatory Focus					
Promotion	0.01	0.07	.00	-.12**	
Prevention	-0.00	-0.11	.01	-.01	
SOQ					
Win Orientation	0.05**	0.17	.02	-.02	
Competitiveness	-0.02**	-0096	.00	-.25**	
Goal Orientation	-0.14**	-0.27	.03	-.32**	
					R^2 = .153**
					95% CI[.11,.19]

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. * indicates $p < .05$. ** indicates $p < .01$.

4.3.2 Comparison Direction and Motivational Effects

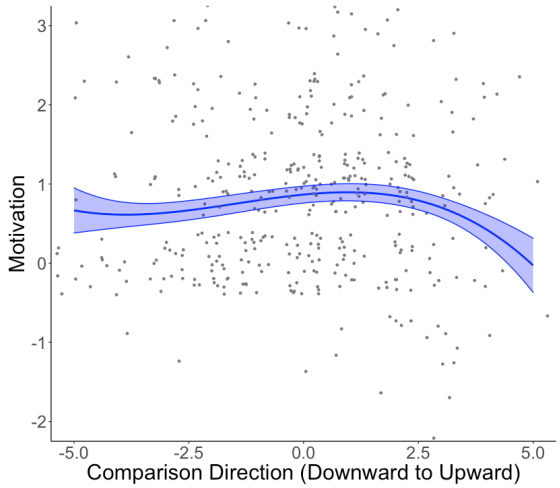
In the second part, we aim to investigate the effects of comparison direction on different motivational outcomes[14]. First, we predicted that upward comparison will be positively associated with motivation after comparison, but only for moderately compared to extremely upward and downward standards (H3a). To begin with, we calculated a random intercept model only (null-model) to decompose the variance in motivation into within- (Level-1) and between-person (Level-2) variance. Results revealed an intraclass correlation coefficient (ICC) of .52 - meaning that 52% of the variance in motivation is explained by interindividual differences and 48% of the variance is potentially explained by situational factors. Next, we tested the predicted non-linear patterns (e. g. motivation goes down with more extreme upward standards) with a random effects, and fixed slopes model where we regressed linear, quadratic and cubic terms of comparison direction scores on motivation. Results revealed a significant quadratic effect on motivation, $B = -0.02$, $p = .013$ ($\beta = -0.09$), but no linear effect, $B = 0.06$, $p = .161$ ($\beta = 0.12$) or cubic effect, $B = -0.01$, $p = .096$ ($\beta = -0.05$). Lastly, comparing our model to the null-model revealed a significant better fit for the polynomial random intercept, fixed slope model, $\chi^2(3) = 8.05$, $p = .045$. As shown in Figure 2, motivation

[14] In order to measure students' motivation, we initially used the variables *pushing* ("I am currently motivated to improve myself") and *motivation after comparison* ("How motivated did you feel after the comparison?"). In this section, we only report the results of the latter variable, because it better fits our research question. Results for pushing can be found in the Supplementary Material.

increases with moderate upward comparisons and decreases with more extreme upward comparisons and also slightly with downward comparisons.

Figure 2

Relationship between Motivation and Comparison Direction



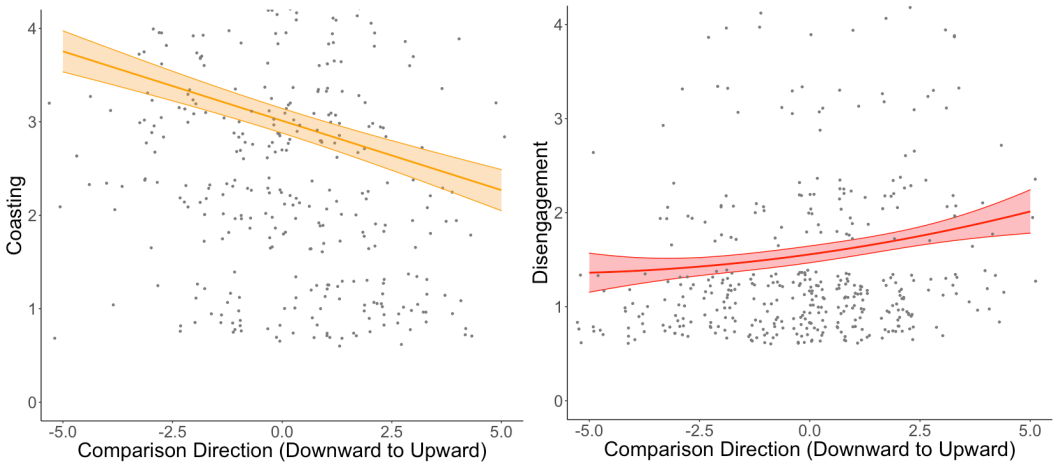
Note. The figure displays the quadratic effect of comparison direction on motivation, showing that motivation increases with moderate upward comparisons and decreases with more extreme upward comparisons and also downward comparisons. Error bands show the 95% confidence interval.

Next, we hypothesized that motivational coasting increases with downward comparison and decreases with upward comparison (H3a). We tested a linear relationship between comparison direction and coasting and in line with our hypotheses, there was a negative association between motivational coasting and upward comparison, $B = -0.15$, $p < .001$ ($\beta = -0.19$), showing that coasting increases with downward and decreases with upward comparison (Figure 3, left half). Likewise, we predicted that disengagement increases with upward comparison, especially towards the more extreme ends of upward comparison. We regressed linear and quadratic terms of comparison direction scores on disengagement. Results showed a significant linear relationship between disengagement and upward comparison, $B = 0.07$, $p = .004$ ($\beta = 0.14$), but no quadratic effect, $B = 0.01$, $p = .541$ ($\beta = 0.02$). Thus, disengagement decreases with downward and increases with upward comparisons,

however, there is no additional boost in disengagement towards the more extreme upward standards (Figure 3, right half).[15]

Figure 3

Relationship between Coasting/Disengagement and Comparison Direction



Note. The figures display the linear relationships between Coasting / Disengagement motive and Comparison Direction. While coasting increases with downward comparison, disengagement increases with upward comparison. Error bands show the 95 % confidence interval.

4.3.3 Performance Improvement

As a measure of effort investment, we assessed performance improvement at the end of the semester consisting of *subjective* improvement (self-report measure) and *objective* improvement (actual running/swimming times).

Descriptive Data on Performance Improvement

While almost all participants rated their subjective improvement on the rating scale, only $n = 38$ students from the swim course, and $n = 5$ from the track-and-field course were able to indicate a baseline plus end result. Thus, with too few running times, only 200 m swim times, more precisely the difference (subtracting baseline result from end result), were used to represent objective improvement for 36.5 % of the sample. At the beginning of the semester, the average 200 m swim time was 227.7 sec ($SD = 33.89$ sec; n = 46) and at the end, it was 212.4 sec ($SD = 33.62$ sec, n = 61). We then calculated difference scores from all those students who could provide a baseline and an end result

[15] For further analyses see Supplementary Material S2 where we tested the relationship of the three social comparison motives self-improvement, self-enhancement, and self-evaluation (Festinger, 1954) and comparison direction.

(positive values representing improved scores). The average improvement over the semester was 18.16 sec (SD = 16.87, n = 38), representing a training gain of 7.2 %[16]. For this subsample, there was a significant relation between subjective ratings and objective measures, $r_\tau = .273$, $p < .001$, so students who were confident they had improved over the course of the semester, also objectively reduced their swim times.

Performance Improvement and Comparison Direction

We hypothesized that performance improvement was positively associated with upward comparison. We conducted two analyses including, first subjective improvement scores predicting comparison direction and secondly, objective scores predicting comparison direction.

Subjective Improvement. First, we regressed scores aggregated Level-1 predictor comparison direction on subjective improvement. As expected, upward comparison significantly predicted subjective performance improvement, $B = 0.23$, $p < .001$ ($\beta = 0.29$). The higher students scored on average (upward) comparison direction, the higher they reported performance improvement.

Objective Improvement. Next, we regressed aggregated comparison direction scores on objective improvement. Again, upward comparison significantly predicted objective performance improvement, $B = 2.68$, $p < .001$ ($\beta = 0.30$). The higher students scored on average (upward) comparison direction, the more they improved their performance. However, it is important to keep in mind that the number of observations ($N = 38$) and thus the statistical power of the test was rather small (47% based on a R^2 of .09 and an alpha level of .05).[17]

4.3.4 Comparison Direction and Emotional Effects

To complement the picture, we investigated the relationship between social comparison and emotions in athletes. We predicted a positive relationship between downward comparison and happiness (H4a) and a positive relationship between upward comparison and the pride to shame measure (H4b). First, we regressed comparison direction scores on happiness after comparison. As

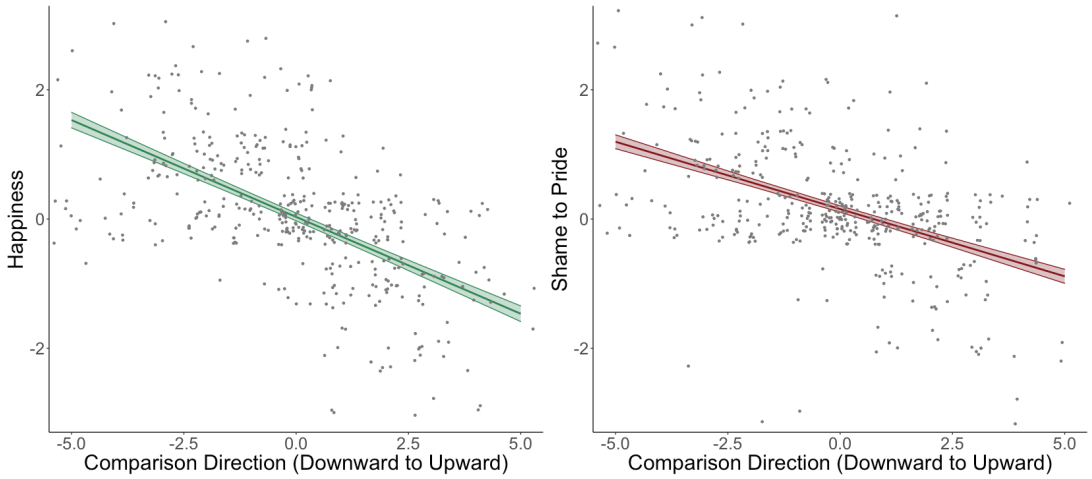
[16] To give a reference frame, the current world record for 200 m breaststroke is 126.12 sec, and the maximum time to qualify for a swimming badge bronze (adults) would be 420 sec.

[17] Since we had more observations of swimming times from the end of the semester (n = 61), we regressed comparison direction scores on swimming performance at the end of the semester as an exploratory analysis. Again, upward comparison positively predicted performance, $B = 9.11$, $p < .001$ ($\beta = 0.43$). Here, we had power of 96 % based on R^2 of .19 and an alpha level of .05.

predicted, upward comparison negatively predicted happiness, $B = -0.30$, $p < .001$ ($\beta = -0.58$). As seen in Figure 4A, happiness decreases with upward but increases with downward comparison. Second, results revealed a negative association between upward comparison and pride, $B = -0.21$, $p < .001$ ($\beta = -0.48$). As illustrated in Figure 4B, downward comparison is related to increasing values of pride and upward comparison to increasing values of shame.

Figure 4

Relationship between Happiness/Shame to Pride and Comparison Direction



Note. The figure shows that happiness and pride were associated with downward comparison and shame was related to upward comparison.

4.4 General Discussion

Athletes compare their abilities and performance to competitors, fellow athletes, or idols of their own discipline. Taking these comparisons as a natural and integral part of sports, the present research provides insights about the interplay between social comparison processes and motivational and emotional affect as well as sport-specific dispositions and comparison tendencies in a real-life sports context. Our testbed allowed us to go beyond past work as it provided the unique opportunity of testing how upward and downward comparison influence actual performance improvement. Beyond that, experience sampling over one semester yielded a more reliable picture of emotional effects than a measurement after single occasions like big tournaments.

Results showed that an increase in motivation is related to the comparison to a superior standard, but only if the discrepancy between self and comparison standard is perceived as moderate rather than

extreme. This is in line with previous research (Diel et al., 2021) and again extends past research, which has predominantly focused on positive motivational effects resulting from upward comparison (Lockwood & Kunda, 1997, 2000). On the other hand, the comparison to inferior standards can also hamper motivation: here, the discrepancy between self and standard is positive, resulting in a coasting mindset ("Compared to the other person, I am doing just fine"; Diel et al., 2021). This has direct practical consequences for goal attainment in sports. Athletes seeking to fully exploit their motivational potential should orient themselves at attainable upward comparison standards. For instance, a marathon runner would be advised not to orient herself to slower runners but instead to focus on the few runners in front of her instead to striving towards the very top, which may be out of reach. Hence, when aiming at strategically implementing comparisons into training routines, it is not advisable to pose low or extremely high aims and comparisons standards (e. g. a non-athlete or Usain Bolt), as corroborated by the findings that downward comparison does not benefit motivation and that increasing upward comparisons make disengagement more likely. The comparison to out-of-reach standards may also trigger choking – a sudden decline in athletic performance due to pressuring circumstances – which not only impacts athletic performance negatively in the short-term, but also in the long run (Hill & Shaw, 2013). Similarly, strong feelings of pressure to meet coaches and parents expectations result in negative emotions (Hellstedt, 1990). Adhering to such a separation and rather orienting at achievable upward standards can also prevent athletes from developing a low self-concept as shown by the big fish/little pond effect in sports (Marsh et al., 1997b). In this context, another overlap between the academic and sport domain becomes apparent: As part of general achievement orientation, students that are performance or ego-oriented (vs. task and learning oriented; Dweck, 1986) rather focus on others as a reference in order to determine their own success and to show superiority over others (Ames, 1995; Garcia et al., 2007). However, even athletic and academic role models that contribute to (sport) students' motivation (Eccles et al., 1998) or help to overcome motivational constraints (Bandura, 1986) need to be perceived as similar in order to boost enhance self-efficacy and motivation (McCullagh & Weiss, 2001; Weiss, 1995).

The risk of lower self-concept and discouraging effects of dissimilar or "out-of-reach" standards brings us to the general aspect of less positive emotions relating to upward comparison beyond

possible motivational boosts. In line with past research, upward comparison was associated with less positive emotions (Carver, 2004), so less happiness and increased shame. It is important to note that motivational and emotional affect associated with discrepancy assessments can be empirically distinguishable. Whereas motivation decreased with more extreme upward standards, negative emotions continued to increase towards more extreme ends. Therefore, people seem to motivationally disengage but do not emotionally resign with extremely upward standards. On the other side, happiness and pride do not drop towards the more extreme ends of downward comparison – meaning, even if one is clearly superior to another person, positive emotion still increase. Stoeber and colleagues (2007) argued that overall perfectionism is typically associated with higher cognitive and somatic competitive anxiety in athletes, however they also demonstrated that once the influence of negative reactions to imperfection was partialled out, striving for perfection was associated with lower anxiety and higher self-confidence. Practitioners and coaches could thus promote moderate upward comparisons for motivational reasons, but work on techniques that allow to successfully identify and control negative reactions to extreme upward comparison standards. The comparison to inferior performances on the other hand can have emotional benefits but comes with the risk of impeding motivation. Those techniques would as well protect damaging comparison with others (Walton et al., 2020) and provide a healthy performance climate. At the same time, excellent athletes who are currently at the top of their game might not benefit from social comparison as their only comparison direction is thus, downward. Instead, they might be more focused on maintaining their current state as the ultimate goal (Ecker & Gilead, 2018) and simultaneously, focus on temporal self-comparisons – thus, beating their personal record. But even top athletes may then compare their performance to performance records in the past, thereby focusing on setting new records in the future.

In general, our student athlete sample reported almost as many upward as downward comparisons (approximately 20% were lateral comparisons), so there was no evidence that athletes have a general bias towards upward comparison standards. However, the impact of a lateral comparison should not go unnoticed. According to the social rank model, it is not primarily important to be superior over others, but rather to be included (Gilbert, 2014). Considering that the athletic condition of our sample

was superior to normal reference populations, and students are aware of being part of an exclusive environment, a lateral comparison could be interpreted as keeping up with an already high standard.

Furthermore, many students received an objective run or swim score at the end of the semester but were missing feedback at the beginning. If upward comparisons predominantly relied on objective criteria (Gotwals & Wayment, 2002), then a lot of the comparisons may have resulted from a subjective perception of co-students which may have eventually been influenced by the other person's outward appearance, charisma, overall fitness level, muscularity, etc. but not necessarily by their real swimming/running abilities. Still, past research revealed that body-related envy triggered by self-other comparisons in young adults influenced their motivation to exercise positively but also negatively depending on whether the facet of envy was benign and the outcome was achievable (e. g. feelings of admiration led to inspiration) or malicious and the outcome unachievable (e.g. feelings of injustice leading to maladaptive strategies, such as avoidance; Pila et al., 2014). Hence, the objective performance score is only one criteria of someone's athletic abilities. We can assume, however, that other characteristics that speak for athletic success (e. g. a muscular body) are also used for self-other discrepancy assessments and thus, influence motivation and emotion similarly.

4.4.1 Sport-specific dispositional predictors

Overall, the influence of sport-specific dispositional predictors on comparison direction did not show the expected pattern. We suspect that one possible explanation for this outcome lies in the differentiation of performance goals. In sports, outcome-, performance- and process-goals can be differentiated (Filby et al., 1999). An outcome-goal is well relatable to social comparison because it measures success by contrasting own and competitors' performances, for instance, to be on top of the podium after the marathon. However, performance- and process-goals are often self-reference standards and can therefore be achieved relatively independently of others. While a typical performance-goal could be to run 10 secs faster than last time, a process-goal could be to focus more on breathing technique. It seems likely that student athletes of our sample pursued the performance goal of lowering their swim time from the beginning of the semester in order to pass the exam, but not to outperform the other students of the class. Achievement-motivated athletes, especially those who are hoping for success, are motivated to surpass themselves rather than others (Wenhold et al., 2008),

which could explain why we found a mild negative relationship to upward comparisons. The same explanation holds for sports-orientated individuals. High goal-orientated individuals can be characterized as constantly seeking to enhance their own performance (Gill & Deeter, 1988). Thus, athletes considering self-reference norms may eventually not be dependent on external comparison standards, because they are able to evaluate a negative outcome of a competition as a success when for instance a personal best was achieved. Therefore, a positive relation between goal-orientation and downward comparisons, does not necessarily contradict our assumptions on motivational effects. Secondly, the finding that win-orientation positively predicted upward comparison fosters the self- vs. other-reference explanation, since win-orientation is the only disposition that is directly related to comparison as victory vs. defeat is the central characteristic to it (Elbe et al., 2008). Finally, we did not find the predicted patterns for social comparison direction in relation to prevention versus promotion focus as part of the regulatory focus theory (Higgins, 2012). According to past research, a prevention or promotion focus only impacts performance if there is a fit between the dispositional trait (e. g. a chronic promotion focus) and the present goal (e. g. improving personal swimming record). However, a regulatory fit does not guarantee improvement: Other research revealed that a regulatory fit only positively influenced performance when performing easy over difficult task (Vogel & Genschow, 2013) or when it involves experts rather than amateurs (Memmert et al., 2009). Thus, whether a chronic promotion focus is actually associated with a tendency of upward comparison might be dependent on other factors, such as the salient regulatory focus goal of the present course (e. g. passing the class by either improving performance or by preventing a drop in the performance) and whether this goal is perceived as easy or difficult.

4.4.2 Performance Improvement

Results on performance improvement were unique to this study. This study, to our best knowledge, is the first to jointly include subjective and objective performance measures to examine the distance between the two and its relation to comparison direction. In contrast to often reported overconfidence biases in athletes (e.g. Fogarty & Else, 2005), we were able to show that students' subjective improvement ratings were related to actual objective improvement. Moreover, upward comparison was associated with both subjective and actual performance improvement. However, we

have to keep in mind that the sample of students that provided a performance score was smaller than anticipated and thus, must be interpreted cautiously. Moreover, one explanation for the relationship between upward comparison and performance improvement is that students who initially performed on average or below average had the greatest scope to improve throughout the semester and simultaneously, had more opportunities to compare upward to students performing above average.

4.4.3 Limitations and Future Research

Some limitations of the present research should be noted. As outlined in the introduction, Gotwald and Wayment (2002) argued that athletes need objective performance information of others to properly compare to them. Considering the lecturer in our swim classes communicated the swim result immediately after students left the pool, it is likely they heard their own results but not the results of the other swimmer(s). Thus, students may have lacked the objective criterion (i.e., the results of other swimmers) in which to make comparisons

Moreover, the effects of social comparison on motivation (pushing, coasting and disengagement) may have been smaller in this sample of athletes compared to other domains or populations. Students of a sport university – especially in their first semester – might have been very motivated from the start, which is shown by relatively high scores on motivation and low scores on disenagement. Future research could investigate social comparison effects in athletes (e.g. university students) who go through a rather difficult phase in a motivational sense, e.g. during a later stage in their studies, where giving up becomes more likely.

ESM is a useful method to capture responses in ecologically valid settings. However, ESM can also have limitations. First, the weekly survey limited social comparison observations to one specific class and time of the week. Hence, the questionnaire could not capture all sport-related comparisons, especially those triggered by media (e. g. watching big sport events on TV, or social media). This ESM approach also prevented the measurement of performance changes, for example, the change in swimming performance depending on whether someone swims along faster or slower swimmers. Future research should therefore ask whether people have moved between comparison standards while performing. Also, the reported comparisons were course-specific – this may be problematic as comparisons may differ *across* courses and disciplines. For instance, a student water polo player might

compare towards the top swimmers in their swimming class to a higher extent than a student soccer player, who is satisfied with comparing to the average swimmers but would certainly focus on superior students in ball sports classes. Beyond that, the blocks of our weekly questionnaire repeated for every comparison object mentioned. We saw a trend that in the first few questionnaires more comparisons were mentioned than at the end, but this trend does not necessarily mirror less comparisons but a lower motivation to report all comparisons towards the end of the semester. Related to this, one limitation to be noted is the use of single-item questions. Thereby, we tried to reduce the weekly load of questions in advance by using single-item questions, which is recommended in experience sampling research to decrease burden and increase data quantity and quality (Eisele et al., 2020). Lastly, our results are of correlational nature and thus, causality cannot be inferred. In this vein, results should be complemented by experimental research investigating causal paths between comparison processes and motivation. On the one hand, people can be exposed to comparison standards that then fuel or hinder motivation (e. g. Diel & Hofmann, 2019) and thereupon actual performance. On the other hand, people who already aim at improvement may then show a bias of upward comparison (see also S4 on the relationship between main social comparison motives and social comparison direction). Moreover, future research may focus on how discrepancy-reducing goals can be achieved. For instance, one could not only benefit from an upward standard as representing an ultimate goal but also by following a superior standard's training strategies that led to such achievements in the first place. Future research should also look at how excellent athletes who are already on top may benefit from social comparison while focusing on their goal to maintain their current state or set new records. And finally, it might be worthwhile to investigate gender differences. Since young male and female athletes differ in their physical abilities in sports (e. g. Bessem et al., 2017), the other sex might become irrelevant as a comparison standard.

4.4.4 Conclusion

Social comparison is a fundamental aspects in sports. On the path to success, athletes do not only constantly beat themselves but also others. The present study emphasizes once again that athletes can benefit from upward comparisons, but only if the upward comparison standard performs moderately better. Extremely upward comparisons on the other hand lead to declining motivation and increasing tendencies of disengagement. Still, actual sport performance appears to be positively influenced by

upward comparison. On the other hand, comparison to downward standards is related to lower performance and increased coasting motivation – the feeling that someone deserves to rest on what has been already achieved. In terms of emotions, athletes just like other populations benefit from downward comparison and suffer from upward comparison. Coaches and their training sessions should benefit from this knowledge by exposing athletes to attainable upward standards that maximize motivation and effort and minimize negative emotions. However, athletes do not appear to have a more pronounced tendency towards upward comparison within their athletic environment where they turn to superior others as much as they turn to inferior others.

Chapter V: More Threatening and More Diagnostic: How Moral Comparisons Differ From Social Comparisons

This chapter is based on following publication:

Fleischmann, A., Lammers, J., Diel, K., Hofmann, W., & Galinsky, A.D. (2021). More threatening and more diagnostic: How moral comparisons differ from social comparisons. *Journal of Personality and Social Psychology, 121*(5), 1057-1078. http://dx.doi.org/10.1037/pspi0000361 (Published online: March 1 2021)

Please note that changes in headings, citation style, and formatting were undertaken to fit the layout of this dissertation. No changes were made to the content of the article.

Abstract

The current research tests how comparisons in the moral domain differ from other social comparisons in three ways. First, an initial experience-sampling study shows that people compare downward more strongly in the moral domain than in most other domains (Study 1, $N = 454$), because people like to feel moral and present themselves as moral. Second, the classic threat principle of social comparison holds that people choose downward comparisons to improve their well-being after a threat to their self-esteem. We propose that in the moral domain the threat principle is intensified because morality is a uniquely important and central comparison domain. Across seven experiments (2ab, 3abc, 4ab), we find that people search for downward comparisons much more than in other domains. This effect is so strong that people are willing to forgo money and incur time costs to avoid upward moral comparisons when threatened. Third, another classic principle of social comparison holds that people only consider comparisons that are diagnostic (i.e., close or similar) and therefore self-relevant, while dismissing extreme or dissimilar comparisons as irrelevant. We propose that this diagnosticity principle is attenuated because morality is a binding code that applies equally to all humans. Across four experiments (Experiments 5ab, 6ab), we find that even the most extreme and dissimilar moral (but not other) comparisons are deemed relevant and potentially threatening. Together, these twelve studies (total $N = 5,543$) demonstrate how moral comparisons are a ubiquitous but fundamentally distinct form of social comparison with altered basic principles.

Keywords: morality; social comparisons; moral identity; downward comparisons; motivated comparisons

5.1 Introduction

Moral exemplars encourage and inspire people to be virtuous and moral. Even decades after their deaths, millions find motivation in Martin Luther King's struggle for civil rights and Mohandas Gandhi's pursuit of Indian independence from the British Empire. People from different faiths call forth their own saints, prophets, and martyrs to conjure up inspiration in their elevated deeds and moral behavior. In more recent times, people look up to the young Pakistani Malala Yousafzai or Greta Thunberg for their advocacy for women's rights and environmental protection. People also find inspiration in less extreme moral comparisons. For example, they admire colleagues who do their work in an exemplary manner or family members who have lived particularly virtuous lives. Scholars have proposed that morally elevated people motivate others to become a better person (Haidt, 2000; Keltner & Haidt, 2003). Although exalted exemplars are predicted to encourage people to meet these lofty standards, a largely unanswered question is whether these exemplars actually produce more virtuous behavior.

To understand how moral exemplars might impact behavior we consider a foundational process of human psychology: social comparisons. Psychological research has distinguished two basic principles of social comparisons. The first principle of social comparison, which we refer to as the *threat principle*, holds that people tend to engage in more downward than upward comparisons after a relevant aspect of the self is threatened (Wills, 1981). To illustrate this with a sports example, a young soccer player who experiences a threat to their self-worth after missing an important penalty kick may choose to compare themselves with an untalented teammate to protect their well-being and feel better about the self. A second classic principle of social comparison, which we refer to as the *diagnosticity principle*, is that upward comparisons are only threatening when the comparison standard is close or similar on the comparison dimension (or other relevant dimensions) and therefore diagnostic and relevant (Tesser, 1988). Thus, extreme upward comparisons can even be inspiring (Lockwood & Kunda, 1997). To come back to the sports example, the unsuccessful soccer player would avoid comparing themselves with a more successful teammate, but would not avoid thinking about Lionel Messi because the extreme performance of FC Barcelona's forward is not diagnostic or relevant for their own performance.

Given the central role of morality as a standard guiding our judgments and decisions, it stands to reason that people also compare frequently in the moral domain and that such comparisons strongly affect our behavior. Nonetheless, to our knowledge no research has systematically tested how people make comparisons in the moral domain and whether such comparisons follow the general principles of comparison. Our aim in the current research is to test whether the general principles of comparison apply in the moral domain. First, we aim to demonstrate that in their daily lives people frequently make downward comparisons in the moral domain. Because morality is central to identity and social standing, we predict that people frequently make downward social comparisons to both view and present themselves as moral.

Next, we test whether moral comparisons follow or violate the threat and diagnosticity principles of social comparisons. Our core proposition is that comparisons in the moral domain differ from comparisons in other domains due to the unique features of morality. Specifically, we propose that morality intensifies the threat principle of social comparisons, but that morality reduces the diagnosticity principle. We theoretically explain these and empirically explore these predictions after first providing a background summary of the literature on social comparisons.

5.1.1 Social Comparisons

Social comparisons are a basic element of human cognition (Festinger, 1954; Mussweiler, 2003). They are so fundamental that even children as young as five months already possess the ability to make comparisons (Baillargeon, 1991; Gentner & Medina, 1998; Gentner & Rattermann, 1991). And they are so essential that they facilitate fast and effortless cognitive processing (Corcoran & Mussweiler, 2009). By comparing oneself with others, people gain a sense of their standing when absolute judgments are not possible, and gain a relative standing compared with others—whether they do better or worse, and whether their opinions are similar to those of others (Festinger, 1954). However, people can also use comparisons in strategic ways. Although people use lateral comparisons to gain diagnostic information about the self, they engage in upward comparisons to improve their abilities, and in downward comparisons to maintain and enhance self-esteem (Corcoran et al., 2011; Taylor & Lobel, 1989).

The importance of social comparisons has been shown in a large variety of domains. For example, social comparisons are related to lower levels of body satisfaction, especially for women (Myers & Crowther, 2009). With regard to academic performance, students in classes with high-performing peers lower their academic self-concept due to upward comparisons, whereas the academic self-concept of students in classes with low-performing peers is boosted due to downward comparisons (Huguet et al., 2009). In sports, tennis players' perceptions of their athletic competence after a lost match depends on the comparison direction, with especially negative consequences if the comparison partner was supposed to be a downward comparison, but won the match (Bardel et al., 2010). People even use upward and downward comparisons strategically: Cancer patients rely on downward comparisons to patients who received a grim prognosis to protect the self, while using upward comparisons to patients who do better than expected to give them hope (Taylor & Lobel, 1989). Across all these domains—irrespective of whether it is academia, sports, or health—the processes of comparison are largely similar.

5.1.2 Moral Comparisons

We propose that comparisons in the moral domain differ from comparisons in other domains, and test this general hypothesis in three predictions outlined below. Morality is such a central comparison domain that it is experienced as a general, panhuman code that applies to all people equally and in the same manner (Haidt & Graham, 2007; Shweder et al., 1987). Even more than academic performance or sports, morality is central to our identity, as it is perceived to be uniquely human (Goodwin et al., 2014; Monin, 2007; Strohminger & Nichols, 2014). Moral judgments are one of the earliest judgments that people make; children younger than one year already form stable impressions about others' morality (Goodwin et al., 2014; Hamlin et al., 2007; Kuhlmeier et al., 2003; Pizarro & Tannenbaum, 2012). Moral judgments are also fast, automatic, chronically accessible, and can be based even on minimal information (Bar et al., 2006; Todorov et al., 2008; Willis & Todorov, 2006; Wojciszke et al., 1998).

One reason why we consider morality to be such an important part of our identity is that moral judgments dominate person perception: In general, people's perception of others are first and foremost influenced by their moral inferences about them. If people want to form an overall impression of

someone, they focus their efforts on discerning that person's moral traits (Brambilla & Leach, 2014; Brambilla et al., 2011; Goodwin et al., 2014; Hartley et al., 2016; Landy et al., 2016; Pizarro & Tannenbaum, 2012; Wojciszke et al., 1998). Furthermore, perceptions of morality are dominant because they can strongly guide the meaning of other traits. For example, competence is a desirable trait, but becomes undesirable if a target is immoral (Landy et al., 2016). In summary, the general impression people have of others crucially depends on morality.

Our first prediction is that people compare downward more strongly in morality than in most other domains. As morality is both central to our identity and to how others see us, people have a strong tendency to want to see and present themselves as moral (Hauke & Abele, 2019; Monin, 2007; Ybarra et al., 2012) and will avoid information suggesting otherwise. That is, people are likely to have stronger self-enhancement motives in morality than in other domains. In turn, self-enhancement motives will result in stronger downward comparisons in morality as a way to maintain a positive self-image (Corcoran et al., 2011; Wood & Taylor, 1991). In contrast to our prediction for moral social comparisons, a recent meta-analysis found that people generally prefer to compare upward (Gerber et al., 2018). Therefore, we predict that because morality is so central to the self and the basis for person perception, people will be more likely to compare downward in morality than in other domains.

5.1.3 Morality Intensifies the Threat Principle of Social Comparisons

Our second prediction also builds off the fact that morality is so important for an individual's understanding of the self (Monin, 2007; Strohminger & Nichols, 2014), to propose that morality intensifies the threat principle of social comparisons. The threat principle states that people are motivated to avoid upward comparisons and therefore focus on downward comparisons when the self is threatened (Hakmiller, 1966; Pyszczynski et al., 1985; Taylor & Lobel, 1989; Wills, 1981). For example, an early study found that students who did badly on a test avoided comparing with others who were more successful to avoid harming their academic self-concept even more (Pyszczynski et al., 1985). This study also revealed a tendency for these students to compare themselves to other students who were even less successful than they were to restore their academic self-concept. Similarly, people with learning disabilities, who experience constant social comparison threats, used downward comparisons in both the learning domain and other domains to create a positive sense of

self (Finlay & Lyons, 2000). A recent meta-analysis established that, although people normally prefer to choose upward comparisons, they opt more for downward comparisons after a threat (Gerber et al., 2018).

We propose that morality increases the tendency to make downward social comparisons after threats because such threats are particularly threatening. Therefore, when the moral self is threated, people will choose downward comparisons in morality even more often and more intensely than they choose downward comparisons in other domains. This prediction is supported by two ideas in the research literature: First, people self-enhance more for morality than for other positive dimensions such as intelligence; that is, they describe themselves to be more moral, but not necessarily more intelligent than other people (Allison et al., 1989; Van Lange & Sedikides, 1998; Ybarra et al., 2012). Second, people often react negatively to moral exemplars, i.e., others who demonstrate a higher morality than they do, particularly if their own moral self is threatened (Cramwinckel et al., 2015; Cramwinckel et al., 2016; Cramwinckel et al., 2013; Minson & Monin, 2012; Monin, 2007; Monin et al., 2008). For example, after inadvertently displaying racist behavior, people react negatively to others who successfully avoided racist expressions (Monin et al., 2008). Similarly, after having eaten meat, people dislike vegetarians who frame their refusal to eat meat in moral terms (Cramwinckel et al., 2013).

Therefore, we predict that moral comparisons not only follow the threat principle, but that this principle is intensified when comparisons are moral in nature. Specifically, we predict that upward moral comparisons will be so threatening in the moral domain that people are especially likely to avoid them and instead show more interest in pleasant downward comparisons when the moral self is threatened.

5.1.4 Morality Reduces the Diagnosticity Principle of Social Comparisons

Our third prediction is that moral comparisons reduce the diagnosticity principle of social comparisons. Specifically, we propose that the tendency to engage in downward moral comparisons is less moderated by the diagnosticity of that comparison than for other domains. In other words, we predict that even upward moral comparisons that are particularly extreme, different, or distant in time

and place from the person will still be threatening and therefore avoided by those who experience a threat to the self.

This prediction violates one of the basic premises of social comparisons: Comparisons depend on the closeness and relevance of the comparison target (Corcoran et al., 2011; Festinger, 1954; Lockwood & Kunda, 1997; Major et al., 1991; Tesser, 1988, 1991). To return to the example we provided earlier, a young soccer player of modest skills may feel better about their skills if they compare their own skills with the worst player on their team, but is unlikely to be affected by comparisons to an elderly individual who can barely walk, or to the world's best player, because those comparisons are too distant to be relevant. In contrast to this sports example, however, we argue that even very extreme upward *moral* comparisons—for example with Martin Luther King, Malala Yousafzai, or Greta Thunberg—will be seen as relevant and therefore threatening to those who feel insecure with their own moral standing. Similarly, we predict that even very extreme downward immoral comparisons—for example with Al Capone or Charles Manson—can be used to feel better about the self.

We propose that even extreme moral exemplars will be considered diagnostic and relevant because morality typically applies equally to all people, regardless of time and place. Although people with different political backgrounds have different conceptualizations of what constitutes morality, most people believe their own standards to be objectively correct (Graham et al., 2013; Graham et al., 2009; Haidt & Graham, 2007; Haidt et al., 2009; Haidt & Joseph, 2006; Koleva et al., 2012). Despite the sizeable cross-cultural differences in moral norms and values, a common finding in anthropology and sociology is that people treat their own moral norms and values as universal, and expect people from other cultures to act according to them (Shweder et al., 1987; Shweder et al., 1997; Tetlock, 2003). In fact, one of the most well-established schools of thought in philosophy—deontology—is based on this idea (Kant, 1974/1929). In other words, people believe that morality provides absolute standards that apply to all humans, across time and space. People recognize that different standards make it unfair to compare the soccer performance of a soccer star and their grandmother. But when it comes to morality, we propose that it is appropriate to compare to even extremely different targets.

Based on this idea that people perceive morality as an absolute standard that applies to all other people, regardless of time and space, we argue that moral comparisons reduce the diagnosticity principle. Support for this prediction again comes from two findings in the research literature: First of all, moral standards are often considered sacred, absolute values that are protected from trade-offs with other values, especially economic ones. Reflecting this sacredness, people react with strong negative reactions to even the thought of making such trade-offs (Atran et al., 2007; Tetlock, 2003; Tetlock et al., 2000). Second, although people realize that sometimes others are forced to make immoral choices due to exceptional circumstances, they still find it difficult to overcome their feelings of revulsion or anger, and fail to sufficiently take into account the strength of the situation that can drive unethical acts (Haidt, 2001). Therefore, we expect that even upward moral comparisons to other people who are distant, extreme, or otherwise not diagnostic or relevant for the self (Corcoran et al., 2011; Festinger, 1954; Lockwood & Kunda, 1997; Major et al., 1991; Tesser, 1988, 1991) are nonetheless threatening, and that people avoid them.

5.1.5 Summary and Overview of Experiments

We conducted twelve studies to test these three predictions and explore how classic social comparisons principles function in the domain of morality. To test our first prediction—that people compare downward more strongly in morality than in other domains—Study 1 uses an experience sampling methodology to test whether people make more downward comparisons in morality compared to a range of other domains in their everyday life.

Our first set of experiments then moves on to the second prediction and tests whether the threat principle of social comparisons is intensified in the moral domain. Experiments 2a and 2b test the basic assumption that comparisons in morality follow the threat principle of social comparison, namely that people choose more downward comparisons when their moral self is threatened. Experiments 3a – 3c establish the robustness of this effect by showing that people are even willing to incur substantial costs to avoid upward comparisons in the moral domain, such as increased time spent on a task (Experiment 3a), a reduced likelihood of winning bonus money (Experiment 3b), or even a lower amount of money actually paid out (Experiment 3c). Experiments 4a and 4b provide evidence

that morality intensifies the threat principle, by testing moral comparisons against athletic (Experiment 4a) and economic (Experiment 4b) comparisons.

With our next four experiments we move to the third prediction and test whether the diagnosticity principle of social comparison—which states that only close and similar comparisons are diagnostic and therefore relevant and threatening—is reduced for morality. We first show that theoretically irrelevant comparisons standards (extreme standards in Experiment 5a, extreme, distant, and different standards in Experiment 5b) are still considered relevant and elicit downward comparisons in morality. Finally, Experiments 6a and 6b demonstrate that comparisons in the moral domain follow the diagnosticity principle less strongly than in other domains.

In all experiments (i.e. from Experiment 2a on), our main method of testing our hypotheses is to measure participants' self-reported interest in reading upward and downward comparisons using Likert scales. Across studies, we also measure participants' choice of selecting either an upward or downward comparison. Given that the latter is a single dichotomous item while the former is an index of multiple Likert scale items, we focus primarily on the former to test our hypotheses, but we also shortly mention any effects on choice as a secondary test of the hypothesis.

It is important to note that across studies we did not define the content of morality for our participants. Instead, we took a bottom-up approach of letting people decide for themselves what comparisons (Study 1) or which memories (Experiments 2a – 6b) they considered moral. We did so by only giving the category name of morality (Study 1; citation blinded for peer review), or by using prototypical moral characteristics and several examples of everyday morality (Experiments 2a – 6b; Conway & Peetz, 2012). Although this stands in contrast to research on moral content, such as research on moral foundations (Graham et al., 2013; Graham et al., 2009), we believe this approach offers two advantages: First, because the content of morality might vary by culture, political orientation, or other interindividual characteristics, letting participants define morality ensured that each participant considered those stimuli both relevant and moral. Second, using artificial stimuli can lower external validity. By letting participants categorize (Study 1) or generate (Experiments 2a – 6b) their own stimuli we ensure a high level of external validity.

Throughout these experiments, we determined our sample size a priori and report how we did so.[18,19] We did not look at any of the results prior to collecting all data, we did not exclude any data, we report all manipulations, and we report all measures. Materials and raw data are available at osf.io/fd6at/?view_only=9298bd746f97465f80583d2643ceb594. Across the experiments, participation was restricted to MTurkers from the US with an approval rate of at least 95. Participants could only take part in one of these experiments.

5.2 Study 1: Downward Comparison in Morality and Other Domains in Everyday Life

Study 1 used an experience-sampling methodology to examine the social comparisons people engage in during their everyday lives. This method allowed us to compare the direction of comparisons in the morality domain with all other comparison domains. We predicted that people would make more downward comparisons in morality than in other domains. To address this question, we analyzed data from [citation blinded for blind review][20]. The study was designed to investigate social comparison processes and their motivational and emotional correlates and consequences in everyday life. Over a 5-day period, participants reported their daily social comparisons and indicated, for instance, in which domain the comparison took place (e.g. academic/work, sport/fitness, financial) and the comparison outcome (e.g., downward or upward comparison). We not only explored whether moral social comparisons increased the use of downward comparisons, but also tested whether the importance of the comparison domain to the person can explain the stronger downward comparisons in morality.

5.2.1 Method

Participants

[18] Power analyses in this paper were based on the common assumption of zero heterogeneity in effect sizes (Kenny & Judd, 2019).

[19] We present another experiment in the supplemental material. While this experiment does not speak directly to the research questions addressed here, we report it to avoid a file drawer.

[20] The analyses reported in this study are based on data that were collected in the *A Motivational Framework of Social Comparison* Study, a large experience-sampling project on social comparison and motivation in everyday life. The addressed preregistered hypotheses and analyses in this article do not overlap with the main report from this project (Diel, Grelle, & Hofmann, 2021).

Participants were 454 Germans (70% female; average age 29.32 years, *SD* = 8.81) who were recruited via panels, social media platforms, and flyers. Participants received five Euro for completing an intake survey and an additional 15 Euro if they responded to at least 70% of their daily signals. Additionally, participants received one lottery ticket for each completed survey, with which they could win 100 Euro. Due to the design, we conducted a simulation power analysis with half-standardized effect sizes. Half-standardized effect sizes indicate how many standardized units on the dependent variable the mean changes from one condition, that is, baseline, to the other condition. A simulation with the simr package (Green et al., 2016) for R (R Core Team, 2017) indicated that we had around 100% power to find a big (β = .80) or medium (β = .50) effect, and 64% power to find a small effect (β = .20).

Materials and Procedure

After completing an intake survey that consisted of dispositional measures and demographic information, participants started the mobile phase with the experience-sampling application *Movisens (Version 0.7.4162, movisens GmbH, 2016)*. Over a 5-day period, they received a total of five daily signals at random time points between 10 am and 10 pm, that led them to a survey asking whether they had made a social comparison since the last signal or, if not, whether they could report any other social comparison from the recent past. If they answered *yes* to either of the questions, participants continued with the main questionnaire. If they were unable to think of any situation, the questionnaire ended. Overall, the main questionnaire consisted of five blocks, which appeared in a randomized order: comparison features (e.g., who was the comparison target?), comparison direction (e.g., was the target perceived as superior or inferior), motivational states, emotional states, and context features (e.g., importance of the comparison domain). A detailed overview of all measures is available at osf.io/za35c/?view_only=9a45d5b7d5bb44da8733d052e9fff56c. The preregistration is available at osf.io/bx9p3?view_only=9de74a1c764e41a6b3f39f3631e1d663.

As part of the comparison feature block, participants indicated the domain of comparison and were provided with a multiple-choice list of 18 common comparison domains (e.g., academic/work, sport/fitness, financial) including the domain moral/ethical behavior. In the comparison direction block, participants indicated the direction or outcome of the comparison ranging from -5 = (compared

to the other person I perceived myself as) extremely worse, to 0 = similar, to 5 = extremely better. For the main analyses, we recoded the item so that higher values represent superior (hence, upward) targets and lower values inferior (downward) targets. Lastly, the context feature block included a measure of domain importance ("How important is the domain of comparison to you?", 0 = *not at all important*, 6 = *very important*).

5.2.2 Results

To answer our main question whether people made more downward comparisons more in the morality domain than in other domains (as preregistered), we conducted a multilevel regression model in which we included domain of comparison as a categorical variable (1 = morality, 0 = other, non-morality) predicting direction of comparison. The analysis of variance with Satterthwaite's method revealed a significant main effect for domain, $F(1, 5260.4) = 51.32$, $p < .001$, $\beta = -0.605$, $CI_{95}[-0.765, -0.458]$[21]. The domain morality was associated with higher levels of downward comparisons than non-morality domains. We also tested whether the stronger downward comparisons in morality were based on the importance of the comparison domain. However, the main effect for domain remained when adding domain importance (person-centered) as a covariate to the model, $F(1, 5248.4) = 49.84$, $p < .001$, $\beta = -0.596$, $CI_{95}[-0.774, -0.419]$. That is, the stronger tendency to compare downward in morality could not be explained by morality being a more important comparison domain than other domains.

We also examined what other domains people also showed stronger downward comparisons in other than morality (see Figure 1). Descriptively, people made more downward comparisons relative to morality in two domains, both related to addiction, namely alcohol and smoking. However, only smoking was significantly different from morality, $F(1, 1719.8) = 7.70$, $p = .006$, $\beta = 0.448$, $CI_{95}[0.159, 0.777]$, while alcohol was not, $F(1, 218.9) = 2.95$, $p = .087$, $\beta = 0.213$, $CI_{95}[-0.063, 0.510]$.

5.2.3 Discussion

Consistent with our predictions, Study 1 found that people make more downward comparisons in morality than in other domains. The prominence of downward comparisons for morality occurred

[21] Confidence intervals reported in Study 1 are bootstrapped percentile confidence intervals (Efron, 1982) following Lai (2020).

relative to a range of domains covering most comparison opportunities. Importantly, these were comparisons made in the course of everyday life situations. The only cases that produced more downward comparisons were related to addiction domains such as smoking and alcohol, while moral behavior even scored higher than other addiction domains such as drugs, media consumption, or caffeine. There may be two reasons why downward comparisons are most frequent for addiction-related behaviors. First, there are constant threats in this domain; for example, cigarette packs come with strong warning labels and graphic pictures of the side effects of smoking. Therefore, stronger downward comparisons in these domains may be caused by the constant presence of threat. Second, many people frame addiction in moral terms as a failing of character. Given its link to morality, it is not surprising that addictions produce so many downward comparisons.

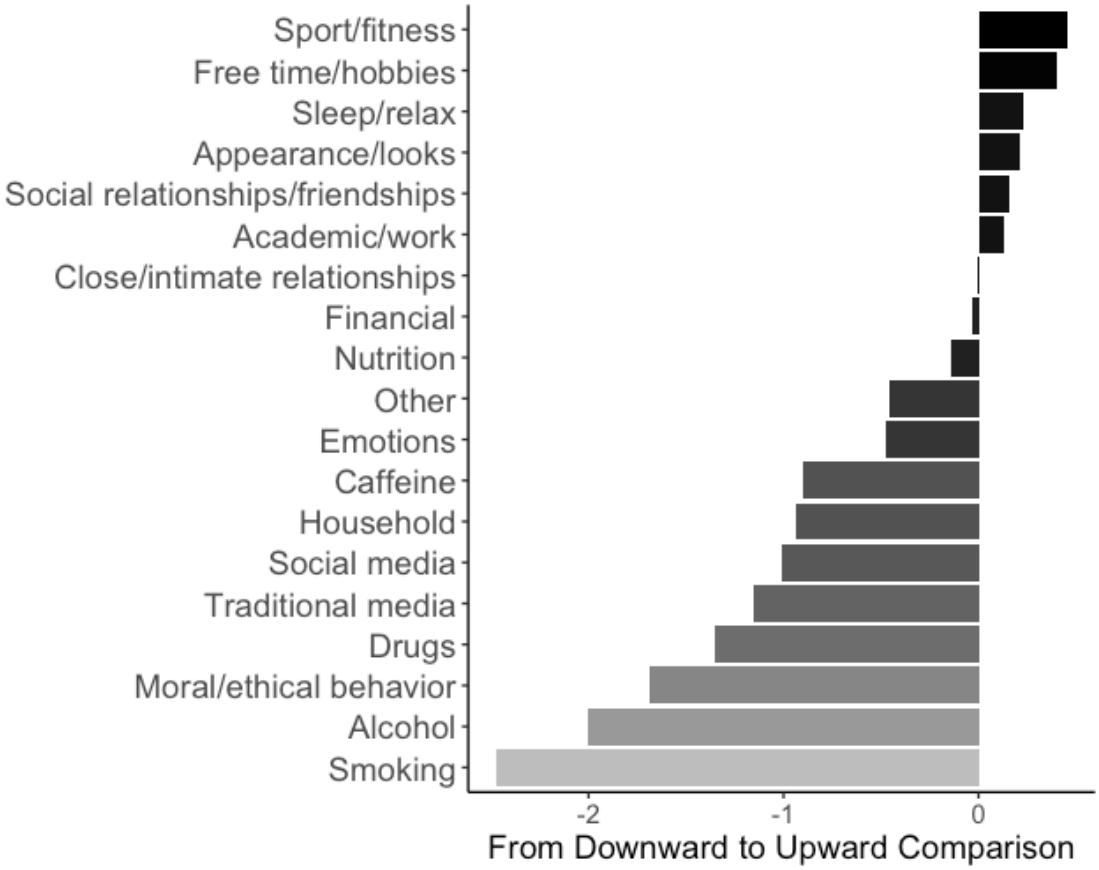


Figure 1. People's comparison direction in all comparison domains in their daily lives. People compare more downward in morality than in (most) other domains. People only compare significantly more downward in smoking, a domain with constant threat.

Interestingly, previous results of a meta-analysis indicate that people normally compare upward, and this trend even occurs under threat (Gerber et al., 2018). In contrast to this meta-analytic finding, we found that people compare downward in morality, even without threat. We also tested whether the stronger downward comparisons in morality can be directly explained by the importance of the domain to the person comparing, and find that this is not the case. These findings mirror those of Skitka et al. (2005), who find that moral mandates have effects over and above their attitude strength. Study 1 offers initial evidence that moral comparisons differ from other social comparisons.

5.3 Experiments 2a – b: The Threat Principle in Morality

Having found that people make more frequent downward comparisons in morality than in (most) other domains, we now turn to the question whether moral comparisons differ with regard to the threat principle of social comparison. Experiment 2a tests the degree to which comparisons in morality follow the threat principle, i.e., people choose downward comparisons more often after a threat. We examined whether threats to participants' sense of their own morality lead people to show more interest in downward than upward moral comparisons. Experiment 2b is a preregistered, exact replication of Experiment 2a, with the addition of a control condition. Given their similarity, we discuss both experiments together. Materials, data and analyses are available at osf.io/fd6at/?view_only=9298bd746f97465f80583d2643ceb594.

5.3.1 Method

Participants and Design

In Experiment 2a, 134 MTurkers (68 female, 66 male; M_{age} = 33; 81% White, 5% African American, 9% Hispanic, 5% other) participated for a compensation of $0.50 and were randomly assigned to a 2(Threat: low vs. high; between) × 2(Comparison Standard: upward vs. downward; within) mixed design. Sample size was set a priori to 134 to achieve 90% power based on a calculation with GPower (Faul et al., 2007) and an effect of η_p^2 = .046 obtained in a pretest (N = 107). This sample size gave us 80% power to find at least an effect of η_p^2 = .035 and 95% for η_p^2 = .057. In Experiment 2b, 461 MTurkers (216 female, 244 male, 1 other; M_{age} = 38; 81% White, 8% African American, 4% Hispanic, 7% other) participated for a compensation of $0.80 and were randomly

assigned to a 3(Threat: low. vs. control vs. high, between) × 2(Comparison Standard: upward vs. downward, within) mixed design. Sample size was set a priori to 460 (rounded up from 455) to obtain 90% power based on an effect of φ = .152 in a pretest for choice of story, as we wanted to set power for the test with higher sample size requirements in the replication (see preregistration, http://aspredicted.org/blind.php?x=99jx8k). For the interest in upward or downward comparisons variable, this gave us 80% power to find an effect of at least $\eta_p^2 = .021$, 90% power for $\eta_p^2 = .028$, and 95% power for $\eta_p^2 = .033$.

Threat Manipulation

To manipulate threat in the domain of morality, we used a behavior recall paradigm (Conway & Peetz, 2012; Sachdeva et al., 2009). In the *low-threat condition*, participants recalled a time when they acted in a way that they felt honorable and righteous. In the *high-threat condition*, participants recalled a time when they acted in a way that they felt guilty or ashamed. In the control condition (only Experiment 2b), participants recalled a day of their last week. Participants were asked to describe the situation, and to write about when it was, who was involved, what they did, and how they felt. To establish whether participants engaged similarly with the recall manipulation, we counted the words participants wrote for Experiment 2a. Overall, participants wrote 98 words on average in each condition.

Comparison Standard

Next, participants were presented with moral (upward) or immoral (downward) comparison standards. As a cover story, participants read that we needed their help in rating stories of other MTurkers and that—to make this more attractive—they could pick one story from a short selection. They were presented with short descriptions of six different stories, each describing mundane everyday events (see osf.io/fd6at/?view_only=9298bd746f97465f80583d2643ceb594). The moral (upward) stories were entitled "Surprised my grandma", "Helped homeless youth", and "Volunteered in a hospital", while the immoral (downward) stories were "Cheated on an exam", "Cheated on my wife", and "Lied to a friend". Order of presentation was randomized. To increase external validity, the

stories were adapted from actual stories written by other MTurkers who took part in a previous study and gave their consent that their stories could be used for future research.

Main Measures

Participants indicated for all stories how interested they would be in reading them, on 7-point scales from *not at all* (1) to *very much* (7). Internal reliability for the interest measure was satisfactory, for the interest in moral stories, Cronbach's $\alpha = .83$ for Experiment 2a, Cronbach's $\alpha = .81$ for Experiment 2b, for the interest in immoral stories, Cronbach's $\alpha = .66$ for Experiment 2a, Cronbach's $\alpha = .73$ for Experiment 2b. Then, participants chose which story to read, and actually read the story.

Further Measures

In Experiment 2a, participants also rated how interesting the story was and how much they enjoyed reading it and—for exploratory purposes—indicated how similar they felt to the standard, indicated how moral the standard was, indicated how moral they themselves felt, and completed the moral identity scale (Aquino & Reed, 2002). These variables are not discussed further. In the end, participants answered demographic questions.

In Experiment 2b, we also included two manipulation checks. As a first manipulation check, participants indicated how *moral*, *good*, *bad*, *righteous*, *honorable*, *guilty*, and *ashamed* they felt when recalling the experience, on 7-point scales from *not at all* (1) to *very much* (7), Cronbach's $\alpha = .94$. As a second manipulation check, participants also indicated how *threatened, concerned* and *alarmed* they felt when recalling the experience, on 7-point scales from *not at all* (1) to *very much* (7), Cronbach's $\alpha = .79$.

5.3.2 Results

Experiment 2a

In support of our predictions, a 2(Threat: low vs. high; between) × 2(Comparison Standard: upward vs. downward, within) mixed ANOVA on reported interest in the stories showed the predicted interaction between threat and comparison standard, although it was marginally significant, $F(1, 132) = 3.47$, $p = .065$, $\eta_p^2 = .026$, CI_{90} [<.001, .084]. Simple effect analyses found that participants in the low-threat condition ($n = 65$) were more interested in reading the moral than the immoral stories (see

Table 1), $F(1, 132) = 6.96$, $p = .009$, $\eta_p^2 = .050$, CI_{90} [.070, .121], presumably out of a desire to read the inspiring examples provided by the moral stories. In the high-threat condition ($n = 69$), interest in reading these inspiring examples was reduced; there was an equally strong interest in reading moral and immoral stories (see Table 1 and Figure 2), $F(1, 132) = 0.002$, $p = .965$, $\eta_p^2 < .001$, CI_{90} [<.001, .001].

Table 1

Means and Standard Deviations for Interest in Upward and Downward Comparisons in Experiments 2a to 4b

	Low Threat		High Threat	
	M	***SD***	***M***	***SD***
Experiment 2a				
Upward	4.18	1.93	3.63	1.86
Downward	3.28	1.89	3.62	1.44
Experiment 2b[a]				
Upward	4.33	1.74	3.73	1.89
Downward	3.13	1.79	3.66	1.68
Experiment 3a				
Upward	4.50	1.87	4.10	1.67
Downward	3.28	1.82	4.20	1.56
Experiment 3b				
Upward	5.65	1.41	5.42	1.45
Downward	3.52	2.00	4.36	1.73
Experiment 3c				
Upward	5.46	1.66	4.96	1.71
Downward	4.17	1.91	4.25	1.64
Experiment 4a				
Morality				
Upward	5.32	1.72	5.13	1.73
Downward	3.62	2.07	4.63	1.70
Athletics				
Upward	5.01	1.51	5.01	1.70
Downward	4.26	1.77	4.67	1.72
Experiment 4b				
Morality				
Upward	4.28	1.68	4.07	1.67
Downward	3.31	1.67	3.50	1.68
Economics				
Upward	3.57	1.79	3.61	1.77
Downward	3.66	1.66	3.70	1.65

[a]Experiment 2b also included a control condition: interest in upward comparison, $M = 4.04$, $SD = 1.74$, interest in downward comparison: $M = 3.50$, $SD = 1.71$

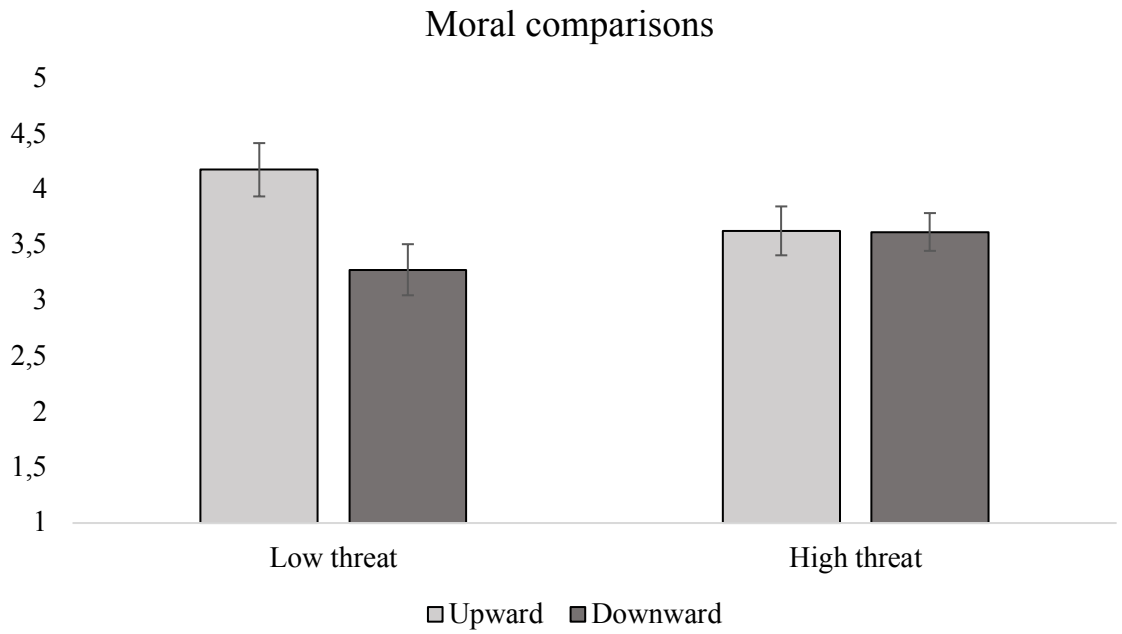


Figure 2. Experiment 2a: Threat increases interest in downward and decreases interest in upward moral comparisons. Bars show mean interest (error bars show the standard error) in making upward and downward moral comparisons.

These differences were also reflected in participants' choices. Participants in the low-threat condition chose immoral and moral stories equally, while participants in the high-threat condition tended to avoid upward comparison and chose to read the immoral stories over the moral stories (see Table 2), $\chi^2(1) = 6.68$, $p = .01$, $\phi = .22$, CI_{95} [.06, .39].

Experiment 2b

First, we examined the manipulation checks. We averaged the items measuring feelings of morality and threat, and used the same two orthogonal contrasts as those preregistered for our main analyses. Participants felt less moral in the high-threat ($n = 148$) condition compared to the low-threat ($n = 159$) and control ($n = 154$) condition, $F(1, 458) = 1021.72$, $p < .001$, $\eta_p^2 = .690$, CI_{90} [.655, .719]. Participants also felt less moral in the control condition than in the low-threat condition, $F(1, 458) = 57.83$, $p < .001$, $\eta_p^2 = .112$, CI_{90} [.070, .158], but this effect was descriptively 84% smaller. The overall effect of threat on feelings of morality was also significant, $F(1, 458) = 542.02$, $p < .001$, $\eta_p^2 = .703$, CI_{90} [.494, .582].

Table 2

Percentages of Chosen Upward and Downward Comparisons in Experiments 2a to 4b

	Low Threat		High Threat	
	Upward	**Downward**	**Upward**	**Downward**
Experiment 2a	49%	51%	28%	72%
Experiment 2b[a]	52%	48%	33%	67%
Experiment 3a	50%	50%	36%	64%
Experiment 3b	81%	19%	67%	33%
Experiment 3c	69%	31%	68%	33%
Experiment 4a				
Morality	65%	35%	57%	43%
Athletics	74%	26%	67%	33%
Experiment 4b[b]				
Morality	33%	37%	31%	32%
Economics	11%	19%	15%	21%
Experiment 5a				
Diagnostic	76%	24%	52%	48%
Non-Diagnostic	77%	23%	56%	44%
Experiment 5b				
Diagnostic	47%	53%	40%	60%
Non-Diagnostic	73%	27%	64%	36%
Experiment 6a				
Morality	65%	35%	49%	51%
Economy	87%	13%	84%	16%
Experiment 6b				
Morality	60%	40%	50%	50%
Athletics	69%	31%	61%	39%

[a]Experiment 2b also included a control condition: 57% chose downward comparisons, and 43% chose upward comparisons.
[b]Participants in Experiment 4a experienced high or low threat for both the moral and economics domain and could then choose between moral and economic upward and downward comparisons.

Participants also felt more threatened in the high-threat condition compared to the low-threat and control condition, $F(1, 458) = 51.22$, $p < .001$, $\eta_p^2 = .101$, CI_{90} [.061, .145]. However, participants felt marginally more threatened in the control than the low-threat condition, $F(1, 458) = 3.80$, $p = .052$, $\eta_p^2 = .008$, CI_{90} [.000, .027], and this effect was descriptively 92% smaller. The overall effect of threat on feeling threatened was also significant, $F(1, 458) = 27.39$, $p < .001$, $\eta_p^2 = .107$, CI_{90} [.027, .094]. We therefore assessed our manipulation as successful. Based on the effect sizes, we also assumed that the

effect was mostly due to the effect of moral threat in the high-threat condition, and not as much due to moral affirmation in the low-threat condition.

Next, we examined interest in stories, again with the two preregistered orthogonal contrasts. In comparison with participants in the low-threat or control condition, participants in the high-threat condition showed more interest in downward comparison standards and less in upward comparison standards, $F(1, 458) = 9.09$, $p = .003$, $\eta_p^2 = .019$, CI_{90} [.004, .045], in line with our predictions. Compared to the low-threat condition, participants in the control condition also showed more interest in downward comparison standards and less in upward comparison standards, $F(1, 458) = 4.71$, $p = .030$, $\eta_p^2 = .010$, CI_{90} [<.001, .031], but the effect was descriptively 48% smaller. We also repeated these analyses with a 3(Threat: low vs. control vs. high; between) × 2(Comparison Standard: upward vs. downward, within) mixed ANOVA as in Experiment 2a to make both experiments more comparable. There was a significant effect of comparison standard, $F(1, 458) = 23.47$, $p < .001$, $\eta_p^2 = .049$, CI_{90} [.021, .084], but no effect of threat, $F(2, 458) = 0.16$, $p = .855$, $\eta_p^2 = .001$, CI_{90} [.000, .008], and, more importantly, a significant interaction of threat and comparison standard, $F(2, 458) = 6.96$, $p = .001$, $\eta_p^2 = .029$, CI_{90} [.002, .038], as in Experiment 2a.

Finally, we tested whether participants with different levels of threat chose different comparison standards. As preregistered, we first compared participants' choices in the high-threat condition to the low-threat and control condition. Compared to the other two conditions, participants in the high-threat condition were more likely to choose downward relative to upward comparisons, $\chi^2(1) = 8.25$, $p = .004$, $\phi = .134$, CI_{95} [.042, .221], in line with our predictions. While participants in the low-threat and control condition chose downward and upward comparison nearly equally, participants in the high-threat condition chose downward over upward comparisons (see Table 2). As preregistered, we also examined whether participants in the control condition choose downward comparisons relative to upward comparisons more often than participants in the low-threat condition, but this was not the case, $\chi^2(1) = 2.38$, $p = .123$, $\phi = .087$, CI_{95} [-.024, .197]. Descriptively, participants in the low-threat condition chose downward and upward comparisons equally, while participants in the control condition seemed to slightly choose downward over upward comparisons (see Table 2). Again, to make this comparable to Experiment 2a, we also analyzed whether choice of stories differed between

all three threat-conditions. Participants chose differently depending on threat, $\chi^2(2) = 10.68$, $p = .005$, Cramer's V = .152, CI_{95} [.070, .247].

5.3.3 Discussion

Experiment 2a provides evidence that comparisons in morality follow the threat principle of social comparisons: People demonstrated a relative tendency to choose downward comparisons when their own moral self was threatened. Participants who felt moral expressed more interest in moral than immoral stories and chose them equally often; in contrast, participants with a threatened morality expressed equal interest, and then actually chose immoral comparisons in the majority (72%) of cases. Experiment 2b replicated Experiment 2a. It showed that participants in the high-threat condition felt less moral and more threatened than participants in the control and participants in low-threat condition. It also replicated that threatened participants expressed more interest in downward and less interest in upward comparison standards. Threatened participants also chose more downward than upward comparison standards compared to non-threatened participants. This effect was mainly (interest in comparison standards) or completely (choice of comparison standards) due to a difference between the high-threat and control conditions, indicating that moral threat was driving the process and not moral affirmation in the low-threat condition.

5.4 Experiments 3a – c: Robustness Tests by Increasing the Attractiveness of Upward Moral Comparison

In the next set of experiments, we test the robustness of motivated downward comparisons in morality by increasing the attractiveness of making an upward moral comparison, compared to choosing a downward comparison. This allows us to test whether participants would still choose downward moral comparisons when choosing them was otherwise less attractive. In designing these experiments, we sought to connect as best as possible to the dominant motivations and desires of our participants. Specifically, MTurkers prefer to finish studies quickly (so that they can take part in more studies) and prefer to be paid as much as possible (Brawley & Pury, 2016; Buhrmester et al., 2011; Litman et al., 2015). Tapping into to these preferences, Experiments 3a – c increased the attractiveness of upward comparisons by making them shorter to complete (Experiment 3a), by increasing the

likelihood that completing them would confer bonus money (Experiment 3b), and by increasing the direct compensation for making such comparisons (Experiment 3c). Our manipulation was non-deceptive, meaning that participants' choices actually affected their time spent on the experiment or their actual monetary outcomes. Given their similarity, we discuss the three experiments together.

5.4.1 Method

Participants and Design

In all three experiments, we used a 2(Threat: low vs. high; between) x 2(Comparison Standard: upward vs. downward, within) mixed design. In all experiments, sample size was set a priori to 240, based on the effect size of Experiment 2a and a power of 90%, calculated with GPower (Faul et al., 2007). In Experiment 3a, 242 MTurkers (135 female, 107 male; M_{age} = 34; 72% White, 11% African American, 7% Hispanic, 10% other) participated in exchange for $0.40. This gave us 80% power to find an effect of at least η_p^2 = .02 and 95% power to find an effect of at least η_p^2 = .032. In Experiment 3b, 239 Mturkers (141 female, 98 male; M_{age} = 35; 80% White, 8% African American, 5% Hispanic, 7% other) participated in exchange for $0.40 and tickets for a lottery. This gave us 80% power to find an effect of at least η_p^2 = .02 and 95% power to find an effect of at least η_p^2 = .033. In Experiment 3c, 242 MTurkers (124 female, 116 male, 2 other; M_{age} = 33; 71% White, 8% African American, 6% Hispanic, 15% other) participated in exchange for $0.40 and additional payment based on the story chosen. Again, this gave us 80% power to find an effect of at least η_p^2 = .02 and 95% power to find an effect of at least η_p^2 = .032.

Materials

All experiments used the stories from Experiment 2a. The only difference between experiments was that we added additional incentives to make selecting upward comparisons about 30% more attractive than downward comparison. Specifically, in Experiment 3a, we showed participants the length of the stories (in words). The three moral (upward comparison) stories were between 150 and 180 words, while the three immoral stories (downward comparison) were between 190 and 230 words, making them less attractive. Indeed, participants who chose to read the longer, immoral stories spent 34% more time doing so than those who chose to reading the shorter, moral

stories, $t(240) = 4.48$, $p < .001$, $d = .58$. In Experiment 3b, participants received a different amount of lottery tickets (for winning a $10 prize) depending on their choice of story. Selecting one of the three moral stories resulted in 12 to 14 tickets, while choosing one of the three immoral stories resulted only in 9 to 11 tickets, making the immoral stories less attractive. In Experiment 3c, participants received a bonus payment depending on their choice of story. Each of the three moral stories yielded higher bonuses ($0.12 to $0.14) than each of the three immoral stories ($0.09 to $0.11), making the immoral stories less attractive

Measures

We used the same measures as in Experiment 2a, including the 7-point measure of interest and the actual choice. Internal reliability for the interest measure was good, for the interest in moral stories, Cronbach's $\alpha = .81$ for Experiment 3a, Cronbach's $\alpha = .83$ for Experiment 3b, and Cronbach's $\alpha = .84$ for Experiment 3c. For the interest in immoral stories, Cronbach's $\alpha = .70$ for Experiment 3a, Cronbach's $\alpha = .85$ for Experiment 3b, and Cronbach's $\alpha = .78$ for Experiment 3c. We dropped the exploratory items that we administered in Experiment 2a, except for in Experiment 3c, where we did include the moral identity scale (Aquino & Reed, 2002), which is not discussed further.

5.4.2 Results

Experiment 3a

In line with our predictions, a 2(Threat: low vs. high; between) x 2(Comparison Standard: upward vs. downward, within) mixed ANOVA on reported interest in the stories showed the predicted interaction of threat and comparison standard, $F(1, 240) = 13.59$, $p < .001$, $\eta^2_p = .054$, CI_{90} [.017, .106]. Simple effects analyses showed that participants in the low-threat condition ($n = 123$) showed more interest in reading the shorter, moral stories that featured upward comparisons than reading the longer, immoral stories, $F(1, 240) = 23.76$, $p < .001$, $\eta^2_p = .090$, CI_{90} [.040, .151]. In contrast, in the high-threat condition ($n = 119$) this difference was eliminated; participants showed an equally strong interest in reading the shorter moral stories and the longer immoral stories, $F(1, 240) = 0.14$, $p = .707$, $\eta^2_p = .001$, CI_{90} [<.001, .015]. Despite their longer length, the psychological benefit of downward

comparisons made these immoral stories more interesting for participants who felt threatened in their morality (see Figure 3).

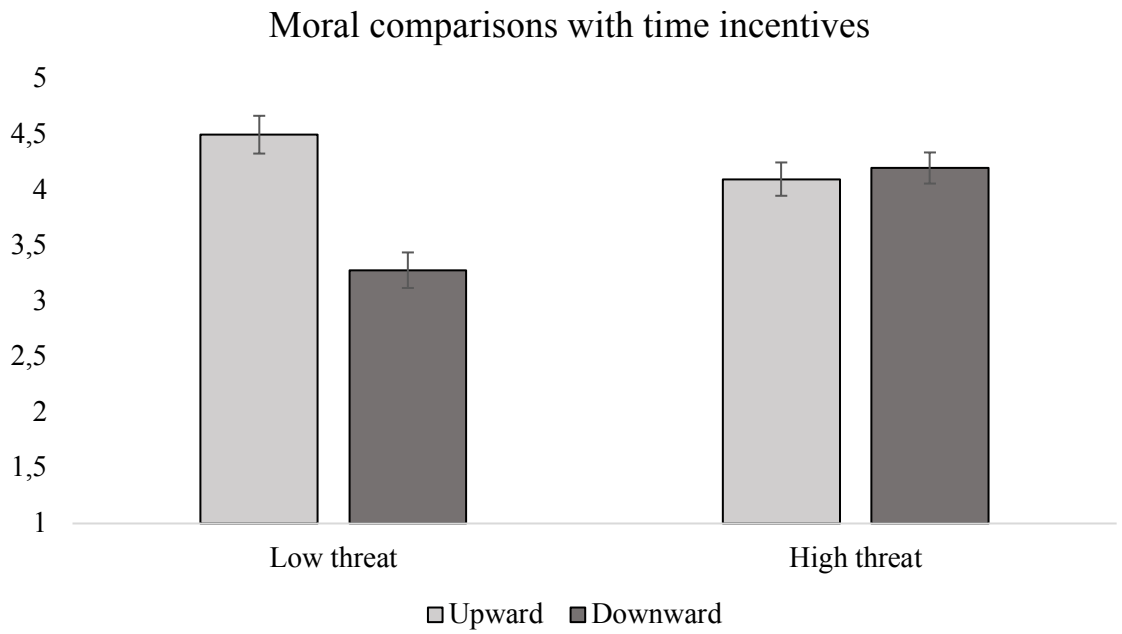


Figure 3. Experiment 3a: Threat increases the interest in downward and decreases the interest in upward moral comparisons, despite time incentives for upward comparisons. Bars show mean interest (error bars show the standard error) in making upward and downward moral comparisons depending on threat.

These effects were also reflected in participants' selection of stories. In the low-threat condition, participants were equally likely to pick the immoral or the moral stories, while in the high-threat condition, participants chose the immoral stories over the moral stories (see Table 2), even though this required them to spend more time reading, $\chi^2(1) = 4.47$, $p = .034$, $\varphi = .14$, CI_{95} [.01, .26].

Experiment 3b

Supporting our predictions, a 2(Threat: low vs. high; between) x 2(Comparison Standard: upward vs. downward, within) mixed ANOVA on reported interest in the stories showed the predicted interaction of threat with comparison standard, $F(1, 237) = 13.76$, $p < .001$, $\eta^2_p = .055$, CI_{90} [.017, .108]. Simple effects analyses showed that when participants did not feel threatened ($n = 117$), they were more interested in reading the moral stories, with the associated higher likelihood of gaining money, than reading the immoral stories, with their lower likelihood of gaining money, $F(1, 237) =$

107.62, $p < .001$, $\eta^2_p = .312$, CI_{90} [.234, .383]. In the high-threat condition ($n = 122$), however, this difference was greatly reduced. Although participants still were more interested in the moral stories with more lottery tickets than the immoral stories with less lottery tickets, the effect size of this difference was descriptively 66% smaller, $F(1, 237) = 27.99$, $p < .001$, $\eta^2_p = .106$, CI_{90} [.051, .169]. These effects were also reflected in participants' selection of stories (see Table 2). In the low-threat condition, only a minority of participants selected immoral stories, but in the high-threat condition, this amount rose to a third, $\chi^2(1) = 6.08$, $p = .014$, $\varphi = .16$, CI_{95} [.04, .28].

Experiment 3c

In line with our predictions, a 2(Threat: low vs. high; between) x 2(Comparison Standard: upward vs. downward, within) mixed ANOVA on reported interest in the stories showed the predicted interaction of threat with comparison standard, $F(1, 240) = 4.25$, $p = .040$, $\eta^2_p = .017$, CI_{90} [.001, .054]. Simple effects analyses again showed that participants in the low-threat condition ($n = 122$) were more interested in the well-paying moral stories than the less-paying immoral stories, $F(1, 240) = 41.55$, $p < .001$, $\eta^2_p = .148$, CI_{90} [.085, .216]. In the high-threat condition ($n = 120$), this difference was greatly reduced. Although participants still showed more interest in the well-paying moral stories than the less-paying immoral stories, the effect size of this difference was descriptively 68% smaller, $F(1, 240) = 12.18$, $p = .001$, $\eta^2_p = .048$, CI_{90} [.014, .099]. In contrast to Experiments 3a and 3b, this effect was not reflected in a significant effect of conditions on participants' choices (see Table 2), χ^2 (1) = 0.05, p = .821, $\varphi = .01$, CI95 [-.11, .14].

5.4.3 Discussion

Replicating Experiment 2a – b, Experiments 3a – c found that people are more interested in downward comparisons and less interested in upward comparisons when their morality is threatened—even when doing so is costly in terms of time or money. Specifically, participants who chose to make an upward comparison spend 34% less time (Experiment 3a), had a 27% higher chance of winning $10 (Experiment 3b), and received 24% more compensation (Experiment 3c) than participants who chose to make a downward comparison instead. And indeed, participants in the low-threat condition consistently were more interested in the better-paying upward comparison. Yet if participants felt

threatened in the moral self, this difference was reduced by between 45% (Experiment 3c) and 92% (Experiment 3a). Given the strong motivation of MTurkers to maximize compensation and minimize participation time (Buhrmester et al., 2011), we consider this robust evidence that the threat principle—choosing downward comparisons after threat—operates in morality.

5.5 Experiments 4a – b: The Threat Principle in Moral vs. Other Social Comparison Domains

Our prior experiments found that comparisons in the moral domain comply with the threat principle: People showed a tendency to be more interested in and to choose more downward and less upward comparisons when their moral self was threatened (Experiments 2a – b), even when doing so was costly in terms of time or money (Experiments 3a – c). However, because these experiments did not compare moral comparisons to other social comparisons domains, they do not yet provide evidence that moral comparisons intensify the threat principle. In Experiments 4a – b, we directly tested comparisons in the moral domains against other commonly studied domains (Gerber et al., 2018): athletic performance (Experiment 4a) and economic success (Experiment 4b, preregistered). In Experiment 4a we used a between-participants design, meaning that participants were exposed either to moral or athletic threats and selected either moral or athletic standards, while in Experiment 4b we used a within-participants design, meaning that participants were exposed to both types of threats and standards.

5.5.1 Method

Participants and Design

In Experiment 4a, 759 MTurkers (426 female, 331 male, 2 other; M_{age} = 35; 77% White, 9% African American, 6% Hispanic, 8% other) participated in exchange for $0.50, and were randomly assigned to one cell of a 2(Threat: low vs. high; between) × 2(Domain: morality vs. athletics, between) × 2(Comparison Standard: upward vs. downward, within) mixed design. Sample size was set to 750 a priori, based on the effect size of Experiment 3b, to achieve a power of 90% (calculated with GPower, Faul et al., 2007), and under the assumption that there would be no effect in the athletics condition

(Simonsohn, 2014). This gave us 80% power to find an effect of at least $\eta_p^2 = .041$ and 95% power to find an effect of at least $\eta_p^2 = .066$.

In Experiment 4b, 401 MTurkers (170 female, 231 male; $M_{age} = 34$; 71% White, 13% African American, 7% Hispanic, 9% other) participated in exchange for $0.60, and were randomly assigned to one cell of a 2(Threat: low vs. high, between) × 2(Domain: morality vs. economics, within) × 2(Comparison Standard: upward vs. downward, within) mixed design. Based on the effect size of Experiment 3b, we needed 368 participants to achieve 90% under the assumption that there would be no effect in the economy domain (Simonsohn, 2014), which we rounded up to a target sample size of 400 (see preregistration here: http://aspredicted.org/blind.php?x=x94uh3). This gave us 80% power to find an effect of at least $\eta_p^2 = .039$, 90% power for $\eta_p^2 = .051$, and 95% power for $\eta_p^2 = .063$.

Threat Manipulation

Experiment 4a used a between-participants design. Participants in the moral conditions received the same threat manipulation (low vs. high) as in previous experiments, while participants in the athletics condition recalled a sports experience that made them feel good (low threat) or bad (high threat). Experiment 4b used a within-participants design. Participants read both the morality and the economics threat manipulation (in randomized order). The moral threat manipulation was the same as before. The economic threat manipulation asked participants to recall a time when they made an economic decision that made them feel good (low threat) or bad (high threat).

Comparison Standard

In Experiment 4a, participants read either morality or athleticism standards. In Experiment 4b, participants read both morality and economic standards. To manipulate the morality-related standards, we used the same stimuli as in Experiment 2a. Similarly, we manipulated the athleticism and economic standards using stories about every day events (see osf.io/fd6at/?view_only=9298bd746f97465f80583d2643ceb594). The descriptions of the athletic standards included “Went jogging for three days a week for a month” and “Finished last in a race”. The descriptions of the economic standards included "Got a bonus for being employee of the month" and "Fell for a fraud company on the internet".

Experiment 4a used the same lottery ticket manipulation from Experiment 3b to test the robustness of the effects. Upward comparisons provided more (12 to 14) lottery tickets than downward comparisons (9 to 11 tickets). No lottery tickets were used in Experiment 4b.

Measures

We used the same measure of interest in the stories and story choice used in previous experiments. Internal reliability for the interest measure was good, for the interest in moral stories, Cronbach's $\alpha = .85$ for Experiment 4a, Cronbach's $\alpha = .83$ for Experiment 4b. For the interest in immoral stories, Cronbach's $\alpha = .83$ for Experiment 4a, Cronbach's $\alpha = .78$ for Experiment 4b. For the interest in athletic stories, Cronbach's $\alpha = .76$, for the interest in the non-athletic stories, Cronbach's $\alpha = .84$, in Experiment 4a. For the interest in stories of economic success, Cronbach's $\alpha = .86$, for the interest in stories of economic failure Cronbach's $\alpha = .81$, in Experiment 4b.

However, participants in Experiment 4b read both economics and morality standards, as stated above. In contrast to Experiment 2a, participants in both experiments were not required to actually read the story (as we were only interested in their choice).

5.5.2 Results

Experiment 4a

Supporting our predictions, a 2(Threat: low vs. high; between) × 2(Domain: morality vs. athleticism, between) × 2(Comparison Standard: upward vs. downward, within) mixed ANOVA on expressed interest in the stories showed the expected three-way interaction between domain, threat, and comparison standard, $F(1, 755) = 6.91$, $p = .009$, $\eta_p^2 = .009$, CI_{90} [.001, .024], see Figure 4. To inspect this interaction, we looked at the moral and athletic conditions separately. In the morality condition, we found a significant interaction of threat and comparison standard, similar to previous experiments, $F(1, 376) = 26.75$, $p < .001$, $\eta_p^2 = .066$, CI_{90} [.031, .110]. Although that same interaction was also found in the athletic condition, $F(1, 379) = 4.64$, $p = .032$, $\eta_p^2 = .012$, CI_{90} [.001, .037], its effect size was descriptively 82% smaller than in the morality condition.

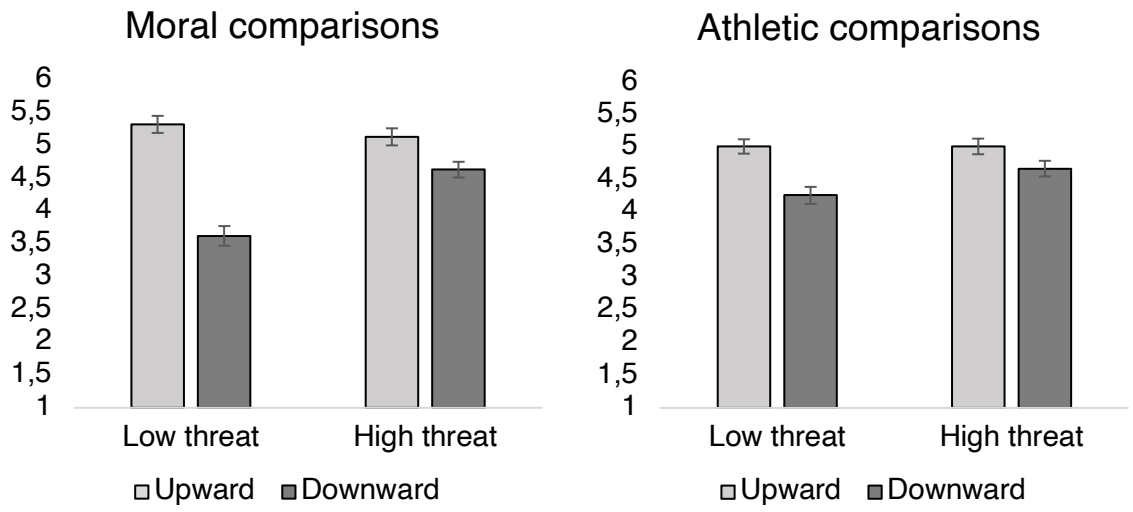


Figure 4. Experiment 4a: Threat increases the interest in downward and decreases the interest in upward moral comparisons more in morality than in athletics. Bars show mean interest (error bars show the standard error) in making upward and downward moral comparisons, depending on threat (high vs. low) and domain of those comparisons (morality vs. athleticism).

As expected, for the morality condition, participants in the low-threat condition (n = 188) were more interested in reading the moral (upward) stories that provided more lottery tickets than the immoral (downward) stories that provided less lottery tickets (see Table 1), $F(1, 376) = 107.09$, $p < .001$, $\eta_p^2 = .222$, CI_{90} [.163, .279]. In contrast, although participants in the high-threat condition (n = 190) were also more interested in reading the better-paying moral than reading the immoral stories (see Table 1), the effect size was descriptively 89% smaller than in the low-threat condition, $F(1, 376) = 9.42$, $p = .002$, $\eta_p^2 = .024$, CI_{90} [.005, .056].

For the athletics condition, participants in the low-threat condition (n = 190) were more interested in reading the upward athletic stories that paid more lottery tickets than reading the downward stories that paid less lottery tickets (see Table 1), $F(1, 379) = 30.82$, $p < .001$, $\eta_p^2 = .075$, CI_{90} [.038, .121]. Participants in the high-threat condition (n = 190) also showed more interest in reading upward than downward stories (see Table 1), and the effect size was descriptively only 79% smaller, $F(1, 379) = 6.33$, $p = .012$, $\eta_p^2 = .016$, CI_{90} [.002, .044].

Turning to our secondary dependent measure, the dichotomous choice responses, the above-described effects observed on the items of primary interest were not mirrored in a similar effect on choice. Although participants chose differently in the high and low threat condition, $\chi^2(1) = 4.91$, $p =$

.027, ϕ = .08, CI_{95} [.01, .15], this did not differ for domain of comparison (see Table 2), as the interaction of threat and domain was not significant, b = .009, p = .997, CI_{95} [-0.606, .0597].

Experiment 4b

In line with our predictions, a 2(Threat: low vs. high; between) × 2(Domain: morality vs. economics, within) × 2(Comparison Standard: upward vs. downward, within) mixed ANOVA on reported interest in the stories showed a marginally significant three-way interaction between domain, threat, and comparison standard (n = 213 for low threat, n = 188 for high threat), $F(1, 399) = 3.36$, p = .068, η_p^2 = .008, CI_{90} [<.001, .029]. Looking at moral comparisons, we (marginally) replicate the previously found interaction of threat and comparison standard, $F(1, 399) = 3.59$, p = .059, η_p^2 = .009, CI_{90} [<.001, .030]. Participants in the low-threat condition showed more interest in reading the moral stories than the immoral stories (see Table 1), $F(1, 399) = 44.32$, $p < .001$, η_p^2 = .100, CI_{90} [.058, .148]. In the high-threat condition, participants also showed more interest in the moral than the immoral stories (see Table 1), but the effect was descriptively 68% smaller, $F(1, 399) = 13.33$, $p < .001$, η_p^2 = .032, CI_{90} [.010, .066].

In contrast, for economic comparisons threat and comparison standard did not interact, $F(1, 399)$ = 0.001, p = .976, $\eta_p^2 < .001$, CI_{90} [<.0001, .0001]. People did not show differential interest in the stories of economic success or economic failure in either condition. In the low-threat condition, they were interested in stories of economic success as well as stories of economic failure (see Table 1), $F(1, 399)$ = 0.41, p = .523, η_p^2 = .001, CI_{90} [<.001, .013]. The same was true for the high-threat condition (see Table 1), $F(1, 399) = 0.41$, p = .522, η_p^2 = .001, CI_{90} [<.001, .013]. As in Experiment 4b, these effects of condition on our measure of primary interest did not affect the measure of secondary interest—participants' choice of stories (see Table 2), $\chi^2(3) = 1.94$, p = .584, Cramer's V = 0.07, CI_{90} [0.00, 0.12].

5.5.3 Discussion

Replicating the previous experiments, participants showed more interest in downward comparisons and less interest in upward comparisons after a threat to their morality. More importantly, we found that this tendency was stronger for moral than for other, commonly studied (Gerber et al.,

2018) types of social comparisons: athleticism and economic success. In other words, the domain of morality intensified the threat principle: People tend to be more interested in downward and less interested in upward moral comparisons when their morality is threatened, and this threat-based tendency is stronger than it is in other commonly studied domains of social comparisons (athletics or economics).

5.6 Experiments 5a – b: The Diagnosticity Principle is Reduced in Moral Comparisons

Having demonstrated that morality intensifies the threat principle, Experiments 5a and 5b test our third prediction, that the diagnosticity principle of social comparisons is reduced in moral comparisons. That is, we expect that the choice of downward comparisons in morality depends less on the diagnosticity of the comparison than in other comparison domains. To test this prediction, we rely on Festinger's (1954) famous suggestion that only diagnostic comparisons impact the self, an idea that has been robustly supported (Lockwood & Kunda, 1997; Major et al., 1991; Tesser, 1988, 1991). Critically, because we argue that morality is an absolute standard, we predicted that people would choose even non-diagnostic comparisons, including extreme (Experiment 5a) and extreme, distant, and different (Experiment 5b) comparison partners, after a self-threat. Given the similarity of these experiments, we discuss them together.

5.6.1 Method

Participants and Design

In both experiments, participants were assigned to one cell of a 2(Threat: low vs. high; between) × 2(Comparison Standard: upward vs. downward, within) × 2(Standard Diagnosticity: diagnostic vs. non-diagnostic, between) mixed design. Sample size for both experiments was set to 560 a priori, based on the effect size of Experiment 3b, a power of 80% (calculated with GPower, Faul et al., 2007), and the assumption that there would be no effect in the non-diagnostic comparison conditions (Simonsohn, 2014). In Experiment 5a, 570 MTurkers (347 female, 218 male, 5 other; M_{age} = 33; 77% White, 8% African American, 5% Hispanic, 10% other) participated for a compensation of $0.35. This gave us 90% power to detect an effect of at least η_p^2 = .071 and 95% for η_p^2 = .087. In Experiment 5b,

561 Mturkers (297 female, 263 male, 1 other; M_{age} = 36; 73% White, 9% African American, 7% Hispanic, 11% other) participated for a compensation of $0.25. This gave us 90% power to detect an effect of at least η_p^2 = .072 and 95% for η_p^2 = .088.

Material

In the diagnostic condition of Experiment 5a, participants were presented with six moderate stories (the four used in Experiments 1 and two new ones that featured donating blood and living off your parents' money). In the non-diagnostic condition, participants were presented with six more extreme versions of the existing moderate stories. For example, the moderate moral story "Helped homeless youth" was matched by the extreme story "Gave away my house and all my belongings to the homeless". As another example, the moderate immoral story "Cheated on an exam" was matched by the extreme story "Got a medical degree by cheating" (see osf.io/fd6at/?view_only=9298bd746f97465f80583d2643ceb594 for all materials). In this way, we sought to manipulate only extremity, but not type of (im)morality. To further increase similarity to previous experiments and provide a strong test of the manipulation, we again used the same lottery manipulation as used in Experiment 3b and 4a in this experiment.

In Experiment 5b, we used the same moderate stories as in previous experiments in the diagnostic condition. For the non-diagnostic condition, we used stories of Polish war heroes and war criminals—which are a very extreme, distant, and different comparison standard. For example, a moral story was "Ekaterina Vasilieva helped many lost children find their families back after WWII", and an immoral story was "Mikhail Semenov used poison gas against civilians, including children" (see osf.io/fd6at/?view_only=9298bd746f97465f80583d2643ceb594).

Measures

We used our previous measure of interest in the stories and choice of stories. The internal reliability of the interest measure was good, in Experiment 5a, Cronbach's α = .75 for the interest in moderate moral stories, Cronbach's α = .82 for the interest in moderate immoral stories. Cronbach's α = .84 for the interest in extreme moral stories, Cronbach's α = .86 for the interest in extreme immoral stories. In Experiment 5b, Cronbach's α = .75 for the interest in moderate moral stories, Cronbach's α

= .77 for the interest in moderate immoral stories, Cronbach's α = .70 for the interest in extreme moral stories, Cronbach's α = .90 for the interest in extreme immoral stories. In Experiment 5a, participants actually read the stories. In Experiment 5b, after rating their interest and choosing a story, participants read that we were only interested in which stories they would choose.

5.6.2 Results

Experiment 5a

As predicted, a 2(Threat: low vs. high; between) × 2(Comparison Standard: upward vs. downward, within) × 2(Standard Diagnosticity: diagnostic vs. non-diagnostic, between) mixed ANOVA on interest in stories revealed the same interaction between threat and comparison standard as in previous experiments, $F(1, 566) = 60.61$, $p < .001$, $\eta_p^2 = .097$, CI_{90} [.061, .136], that—importantly—was not qualified by any three-way interaction with standard diagnosticity, $F(1, 566) = 1.89$, $p = .170$, $\eta_p^2 = .003$, CI_{90} [<.001, .016]. Instead, in line with our predictions, we found the same threat × comparison standard interaction as in previous experiments, in both the diagnostic ($n = 137$ for low threat, $n = 147$ for high threat), $F(1, 282) = 22.27$, $p < .001$, $\eta_p^2 = .073$, CI_{90} [.032, .126], and non-diagnostic condition ($n = 144$ for low threat, $n = 142$ for high threat), $F(1, 284) = 38.99$, $p < .001$, $\eta_p^2 = .121$, CI_{90} [.067, .181] (see Table 3 and Figure 5). In fact, the interaction was descriptively even larger in the non-diagnostic than in the diagnostic condition.

Table 3

Means and Standard Deviations for Interest in Upward and Downward Comparisons in Experiments 5a and 5b

	Low Threat		High Threat	
	M	***SD***	***M***	***SD***
Experiment 5a				
Diagnostic				
Upward	5.39	1.47	5.05	1.56
Downward	3.68	2.12	4.54	1.62
Non-Diagnostic				
Upward	5.31	1.36	5.04	1.74
Downward	3.37	2.13	4.81	1.72
Experiment 5b				
Diagnostic				
Upward	4.34	1.78	4.11	1.94
Downward	3.39	1.82	3.72	1.79
Non-Diagnostic				
Upward	4.77	1.46	4.67	1.58
Downward	2.76	1.86	3.38	1.97

Similar results were found for choice of story. Here, we found that participants were more likely to choose the moral stories (upward comparison) in the low-threat than in the high-threat condition, $\chi^2(1) = 38.45$, $p < .001$, $\phi = .26$, CI_{95} [.18, .34], but importantly this did not depend on the diagnosticity (or lack thereof) of those upward comparisons (see Table 2), as the interaction between threat and diagnosticity was not significant, $b = 0.256$, $p = .485$, CI_{95} [-0.474, 1.061]. That is, this effect was found both for the diagnostic (moderate), $\chi^2(1) = 23.92$, $p < .001$, $\phi = .29$, CI_{95} [.18, .40], and the non-diagnostic (extreme) condition, $\chi^2(1) = 14.75$, $p < .001$, $\phi = .23$, CI_{95} [.11, .34].

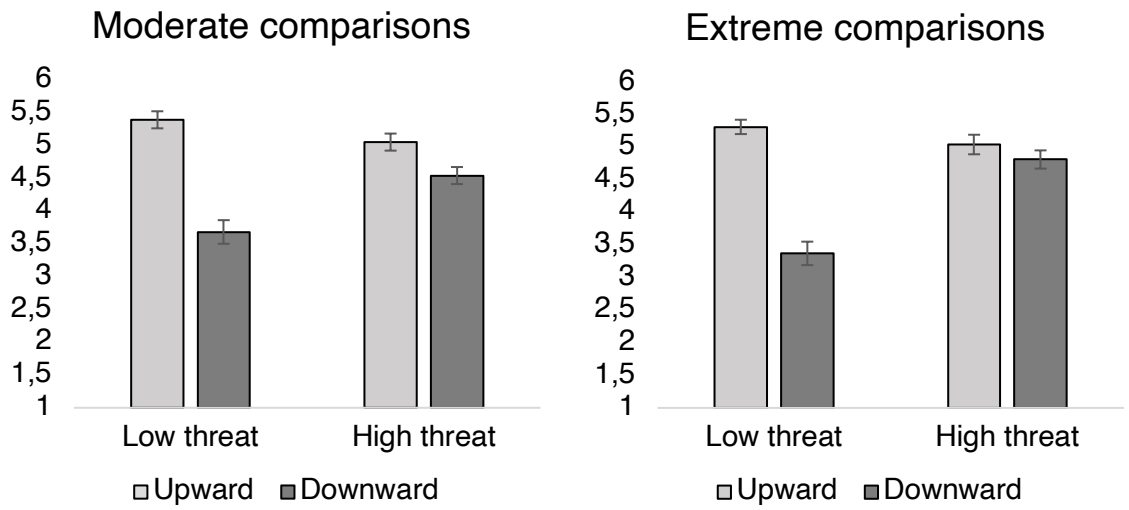


Figure 5. Experiment 5a: Threat increased interest in downward moral comparisons and decreased interest in upward moral comparisons, even when the comparisons were so extreme to be rendered non-diagnostic. This violates one of the basic principles of social comparison. Bars show mean interest (error bars show the standard error) in making upward and downward moral comparisons, depending on threat (high vs. low) and diagnosticity of those comparisons (diagnostic moderate vs. non-diagnostic extreme comparisons).

Experiment 5b

Replicating Experiment 5a, the 2(Threat: low vs. high; between) × 2(Comparison Standard: upward vs. downward; within) × 2(Standard Diagnosticity: diagnostic vs. non-diagnostic, between) mixed ANOVA showed the predicted threat × comparison standard interaction, $F(1, 557) = 8.15$, $p = .004$, $\eta_p^2 = .014$, CI_{90} [.003, .035], and this interaction not qualified by the three-way interaction with standard diagnosticity, $F(1, 557) = 0.11$, $p = .741$, $\eta_p^2 < .001$, CI_{90} [<.001, .006]. Following our predictions, there was again the same comparison standard × threat interaction in both conditions, marginally for the diagnostic (MTurker) condition ($n = 145$ for low threat, $n = 139$ for high threat), $F(1, 282) = 2.93$, $p = .088$, $\eta_p^2 = .010$, CI_{90} [<.001, .038], and significantly for the non-diagnostic (Polish war heroes and criminals) condition ($n = 140$ for low threat, $n = 137$ for high threat), $F(1, 275) = 5.58$, $p = .019$, $\eta_p^2 = .020$, CI_{90} [.002, .055] (see Table 3). Again, descriptively the effect size was even bigger in the non-diagnostic conditions.

Turning to our measure of secondary interest—participants' choice—we found that consistent with our predictions the effect of threat did not differ between the diagnostic and non-diagnostic conditions, $b = -0.103$, $p = .772$, CI_{95} [-0.803, 0.580]. Overall, participants chose marginally different

stories to read in the low vs. high-threat condition, $\chi^2(1) = 3.49$, $p = .062$, $\phi = .08$, CI_{95} [-.003, .16]. Participants in the low-threat condition chose moral stories more and immoral stories less than participants in the high-threat condition (see Table 2).

5.6.3 Discussion

Experiments 5a – b found initial evidence that the diagnosticity principle is reduced for morality. Specifically, we tested whether the avoidance of upward comparisons depended on the diagnosticity of the comparisons. Because social comparison theory predicts that only close and similar comparison standards are diagnostic and thus relevant (Festinger, 1954; Goethals & Darley, 1977), threat should only affect diagnostic comparisons. However, based on our theorizing that morality represents an absolute standard, we predicted and found in Experiment 5a that when threatened, people show less interest in upward and more in downward comparisons, independent of the extremity of that comparison. Someone who donates all their belongings to the homeless appears to functions as an upward threat in the same way as someone who makes a small donation. Similarly, someone who leaves his chronically ill wife for an affair with her best friend functions as a downward comparison in the same way as someone who cheats on his wife. In fact, Experiment 5b showed that even when comparisons were taken from stories of Polish war heroes and war criminals—meaning totally disconnected from participants' daily experiences—it produced the same motivated tendency to show more interest in downward and less in upward comparisons when threatened, as when these comparisons were of a mundane nature.

5.7 Experiments 6a – b: The Diagnosticity Principle in Morality vs. Other Domains

To provide further support for our reasoning that the diagnosticity principle is reduced in morality, Experiments 6a – b tested whether the effect of diagnosticity is smaller in morality than it is for other forms of social comparisons. Mirroring our approach in Experiments 4a and 4b, we test this by comparing morality to the domains of economic success and of athleticism.

5.7.1 Method

Participants and Design

Participants were randomly assigned to one of the four cells of a 2(Threat: low vs. high, between) × 2(Domain: morality vs. economic/athletics, between) × 2(Comparison Standard: upward vs. downward, within) mixed design. Sample size was set to 750 a priori, based on the effect size of Experiment 3b and 90% power (calculated with GPower, Faul et al., 2007), and the same assumption about the interaction as in previous experiments (Simonsohn, 2014). In Experiment 6a, 757 MTurkers (451 female, 305 male, 1 other; M_{age} = 35; 77% White, 8% African American, 6% Hispanic, 9% other) participated for a compensation of $0.35. In Experiment 6b, 752 MTurkers (387 female, 357 male, 8 other; M_{age} = 36; 72% White, 9% African American, 8% Hispanic, 11% other) participated for a compensation of $0.50. This gave us 80% power to detect an effect of at least $\eta_p^2 = .041$ and 95% for $\eta_p^2 = .066$ for both experiments.

Threat Manipulation

The moral threat manipulation was the same as in prior experiments. In the economic condition in Experiment 6a, participants were asked to recall an economic decision that either made them feel good (low threat) or that made them feel bad (high threat). In the athletics condition in Experiment 6b, participants were asked to recall an experience with sports that either made them feel good (low threat) or that made them feel bad (high threat).

Comparison Standard

We used the same extreme moral standard materials as in Experiment 5a. For the economic condition, we created three stories of extreme economic success and three stories of extreme economic failure. To ensure that these stories were clearly extreme and therefore non-diagnostic compared to the experiences typically recalled by our participants, we included stories about people making or loosing hundreds of thousands of dollars, and establishing or ruining big companies (see osf.io/fd6at/?view_only=9298bd746f97465f80583d2643ceb594). Similarly, for the athletics condition, we created five stories of extreme success and five stories of extreme failure in athleticism. Of these stories, we later chose three each in a pretest. Again, these stories were clearly extreme and non-diagnostic, as they included becoming the best athlete in differing sports, or not being able to do the simplest sports at all (see osf.io/fd6at/?view_only=9298bd746f97465f80583d2643ceb594).

Pretest Experiment 6a

To establish that the stories from the different domains did not differ in their extremity, we conducted a pretest for each experiment. In the pretest for Experiment 6a ($N = 201$), participants rated the moral ($M = 6.05$, $SD = 1.01$) and immoral ($M = 1.54$, $SD = 0.81$) stories, on a 7-point scale from 1 (*extremely immoral*) to 7 (*extremely moral*), and the economic successful ($M = 6.28$, $SD = 0.93$) and unsuccessful ($M = 1.69$, $SD = 0.90$) stories, on a 7-point scale from 1 (*extremely unsuccessful*) to 7 (*extremely successful*). A 2(Comparison Standard: upward vs. downward, within) × 2(Domain: morality vs. economy, within) repeated measures ANOVA showed a main effect of comparison standard, $F(1, 200) = 2093.02$, $p < .001$, $\eta_p^2 = .913$, CI_{90} [.900, .925], confirming that the upward comparisons were rated higher than the downward comparisons. There was also a non-interesting main effect of domain, $F(1, 200) = 19.06$, $p < .001$, $\eta_p^2 = .087$, CI_{90} [.034, .153], meaning that judgments of stories in the economic domain were overall higher than judgments of stories in the moral domain. Critically, we found no interaction, $F(1, 200) = 0.71$, $p = .401$, $\eta_p^2 = .004$, CI_{90} [<.001, .030], indicating that the moral and economic standards were similar in extremity.

Pretest Experiment 6b

In the pretest for Experiment 6b, participants again rated the moral and immoral stories, on a 7-point scale from 1 (*extremely immoral*) to 7 (*extremely moral*), and the athletic and unathletic stories, on a 7-point scale from 1 (*extremely unathletic*) to 7 (*extremely athletic*). We chose three athletic ($M = 6.39$, $SD = 0.91$) and unathletic stories ($M = 2.02$, $SD = 1.46$) that were similar in interest to the immoral ($M = 1.77$, $SD = 1.38$) and moral ($M = 6.00$, $SD = 1.14$) stories. A 2(Comparison Standard: upward vs. downward, within) × 2(Domain: morality vs. athletics, within) repeated measures ANOVA showed the expected effect of comparison standard, $F(1, 199) = 959.76$, $p < .001$, $\eta_p^2 = .828$, CI_{90} [.800, .852], as the upward comparisons were rated as clearly more athletic than the downward comparisons. There was also an uninteresting effect of domain, $F(1, 199) = 44.28$, $p < .001$, $\eta_p^2 = .182$, CI_{90} [.107, .259], meaning that the stories in the athletic domain were overall rated higher than the stories in the moral domain. Most importantly, we found no interaction, $F(1, 199) = 1.72$, $p = .191$, $\eta_p^2 = .009$, CI_{90} [<.001, .042], indicating that the moral and athletic standards were similar in extremity.

Measures

We used the same measure of interest in the stories and choice of stories as in prior experiments. Internal reliability for the interest measure was good, for the moral stories, Cronbach's $\alpha = .73$ for Experiment 6a and Cronbach's $\alpha = .72$ for Experiment 6b. For the immoral stories, Cronbach's $\alpha = .79$ for Experiment 6a and Cronbach's $\alpha = .78$ for Experiment 6b. For the interest in stories of extreme economic success, Cronbach's $\alpha = .62$, for the interest in stories of extreme economic failure, Cronbach's $\alpha = .82$ in Experiment 6a. For the interest in stories of extreme athletic success, Cronbach's $\alpha = .70$, for the interest in stories of extreme athletic failure, Cronbach's $\alpha = .80$, in Experiment 6b.

5.7.2 Results

Experiment 6a

As predicted, a 2(Threat: low vs. high; between) × 2(Domain: morality vs. economics) × 2(Comparison Standard: upward vs. downward, within) mixed ANOVA on interest in stories showed the expected three-way interaction of threat, comparison standard and domain, $F(1, 753) = 9.31$, $p = .002$, $\eta_p^2 = .012$, CI_{90} [.003, .028], see Figure 6. To explain this interaction, we tested both domains separately. In line with our predictions, the interaction of threat and comparison standard was significant in the moral domain, $F(1, 368) = 25.68$, $p < .001$, $\eta_p^2 = .065$, CI_{90} [.030, .109]. Simple main effects showed that as in previous experiments, participants who did not feel threatened ($n = 191$) clearly showed more interest in the moral than the immoral stories (see Table 4 and Figure 6), $F(1, 368) = 88.61$, $p < .001$, $\eta_p^2 = .194$, CI_{90} [.137, .251]. For participants who felt threatened ($n = 179$), this difference was much weaker (see Table 4), $F(1, 368) = 4.24$, $p = .040$, $\eta_p^2 = .011$, CI_{90} [<.001, .036].

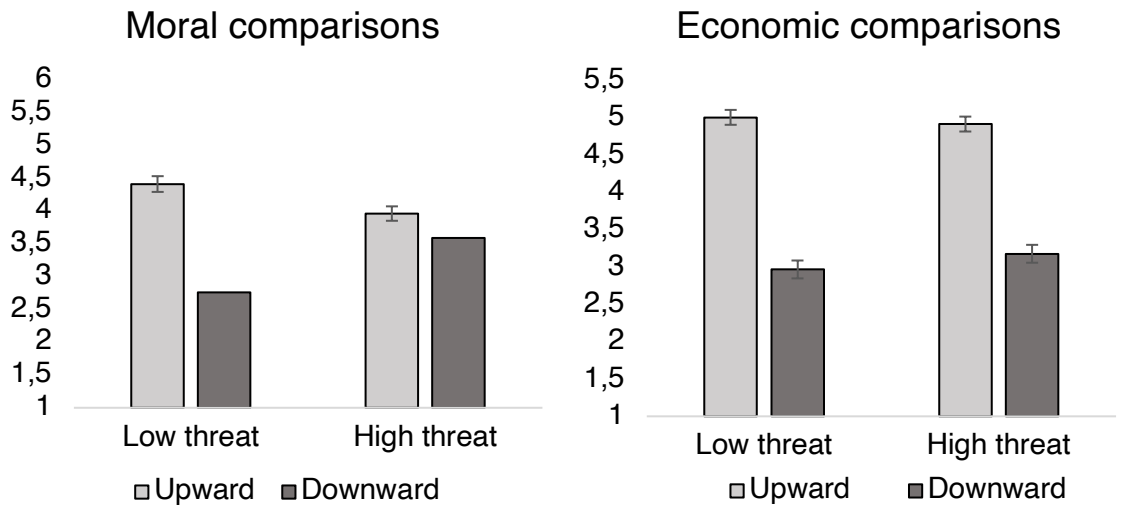


Figure 6. Experiment 6a: Threat increases the interest in downward comparisons for extreme comparisons in morality, but not for extreme comparisons in economics. Bars show mean interest (error bars show the standard error) in making upward and downward comparisons, depending on threat (high vs. low) and domain of those comparisons (morality vs. economics).

In the economic domain, however, interest in stories did not depend on threat, $F(1, 385) = 2.10$, $p = .149$, $\eta_p^2 = .005$, CI_{90} [<.001, .024]. Here, both participants in the low ($n = 197$) and high-threat ($n = 190$) condition were more interested in the stories of economic success than the stories of economic failure (see Table 4), $F(1, 385) = 205.84$, $p < .001$, $\eta_p^2 = .348$, CI_{90} [.287, .404] for the low-threat condition; $F(1, 385) = 145.46$, $p < .001$, $\eta_p^2 = .274$, CI_{90} [.214, .331] for the high-threat condition.

Table 4

Means and Standard Deviations for Interest in Upward and Downward Comparisons in Experiments 6a and 6b

	Low Threat		High Threat	
	M	***SD***	***M***	***SD***
Experiment 6a				
Morality				
Upward	4.41	1.61	3.96	1.52
Downward	2.76	1.84	3.59	1.84
Economics				
Upward	5.01	1.38	4.91	1.37
Downward	2.97	1.68	3.17	1.59
Experiment 6b				
Morality				
Upward	4.23	1.63	3.97	1.62
Downward	2.86	1.78	3.45	1.83
Athletics				
Upward	4.29	1.49	4.19	1.60
Downward	3.12	1.69	3.25	1.66

Looking at choice, the results were similar. As in previous experiments, participants in the high and low-threat condition chose differently, $\chi^2(1) = 7.71$, $p = .005$, $\phi = .10$, CI_{95} [.03, .17]. This was mainly due to the effect in the morality condition (see Table 2). Participants who were not threatened chose moral stories more often, while threatened participants did not, $\chi^2(1) = 9.38$, $p = .002$, $\phi = .15$, CI_{95} [.06, .26]. In the economic condition, threat did not change that participants overall chose stories of economic success more often (see Table 2), $\chi^2(1) = 0.76$, $p = .383$, $\phi = .04$, CI_{95} [-.06, .14]. But the moral and economic condition did not differ significantly, as there was no significant interaction of threat and domain, $b = -0.394$, $p = .275$, CI_{95} [-1.156, 0.347].

Experiment 6b

As predicted, a 2(Threat: low vs. high; between) × 2(Domain: morality vs. athletics) × 2(Comparison Standard: upward vs. downward, within) mixed ANOVA (marginally) showed the

predicted three-way interaction between threat, domain and comparison standard, $F(1, 748) = 3.48$, $p = .063$, $\eta_p^2 = .005$, CI_{90} [<.001, .016]. Again, we tested each domain separately. As expected, for morality we found the established interaction of threat and comparison standard, $F(1, 372) = 12.04$, $p = .001$, $\eta_p^2 = .031$, CI_{90} [.009, .066]. Participants in the low-threat condition ($n = 189$) were more interested in upward than downward comparisons (see Table 4), $F(1, 372) = 61.64$, $p < .001$, $\eta_p^2 = .142$, CI_{90} [.091, .196]. While participants in the high-threat condition ($n = 185$) also showed more interest in upward than downward comparisons (see Table 4), $F(1, 372) = 8.33$, $p = .004$, $\eta_p^2 = .022$, CI_{90} [.004, .052], this difference was descriptively 85% smaller.

In the athletics domain, however, participant's interest in upward or downward comparisons did not depend on threat, $F(1, 376) = 0.93$, $p = .335$, $\eta_p^2 = .003$, CI_{90} [<.001, .018]. In both the low-threat ($n = 190$), $F(1, 376) = 49.82$, $p < .001$, $\eta_p^2 = .117$, CI_{90} [.071, .169], and the high-threat condition ($n = 188$), $F(1, 376) = 32.01$ $p < .001$, $\eta_p^2 = .079$, CI_{90} [.040, .125], participants were more interested in upward comparisons than downward comparisons (see Table 4). The interaction of domain and threat did not translate to choice of stories (see Table 2), $b = 0.004$, $p = .988$, CI_{95} [-0.600, 0.611].

5.7.3 Discussion

Experiments 6a – b clearly demonstrate that the diagnosticity principle is reduced for moral comparisons relative to other social comparison domains. For both athletic and economic comparisons, participants clearly showed more interest in reading about extreme upward comparisons than extreme downward comparisons regardless of whether their self was threatened. This confirms the diagnosticity principle because it demonstrates that extreme comparisons were seen as non-diagnostic and therefore non-threatening. In contrast, interest in reading about extreme moral and immoral comparisons depended on whether the self was threatened, indicating that extreme moral comparisons were not seen as non-diagnostic. This suggest that because morality is so important and considered to be a pan*human* standard, even extreme upward comparisons are seen as relevant and therefore have the potential to be threatening. Although it is theoretically possible that even more extreme, more distant, or more different comparison standards might be considered non-diagnostic in morality (but it is difficult to imagine what this would be, concretely, given that war criminal or war

heroes are insufficiently extreme), the standard definitions of diagnosticity used in social comparison theory do not apply in morality.

5.8 General Discussion

Twelve studies, including one preregistered experience-sampling study, two preregistered experiments, and nine non-preregistered experiments, tested whether social comparisons are different in the moral domain. First, consistent with the notion of the centrality of morality for the self and social relationships, we found strong support in our experience sampling study that downward comparisons dominate in the domain of morality. In fact, downward comparisons are only more dominant when it comes to physical additions such as alcohol and cigarettes.

Second, we found that the in line with the idea that moderately upward moral comparisons are threatening (Tesser, 1988, 1991), people showed less interest in upward moral comparisons and more interest in downward comparisons when their moral self was threatened (Experiments 2a – b). Experiments 3a – c demonstrated that this effect is so robust that people are willing to incur substantial costs to avoid upward comparisons when threatened, such as time lost working on a longer task (Experiment 3a), a reduced likelihood of winning bonus money (Experiment 3b), or even a reduced direct monetary payment (Experiment 3c). Moreover, we demonstrated that this effect was stronger in the moral domain than in other social comparison domains: When threatened, people showed a stronger interest in downward and a weaker interest in upward comparisons in the moral domain than in two other commonly studied domains, athletic performance and economic success (Experiments 4 – b).

Third, we found that for moral comparison the diagnosticity principle is attenuated. This principle states that upward comparisons are only threatening when they are close or similar, and therefore diagnostic (Festinger, 1954; Tesser, 1988). In contrast to this principle, we found that even seemingly irrelevant comparisons were experienced as threatening for people's moral self. Even when upward comparisons were so extreme that they involved selling one's house and donating the money to the poor (Experiment 5a), or so distant in time and place that they involved historical figures on a different continent more than half a century ago (Experiment 5b), participants whose moral self was threatened still avoided them and instead showed more interest in downward comparisons.

Experiments 6a – b confirmed that this reduction of diagnosticity is a unique feature of moral comparisons: If the self was threatened, participants demonstrated a reduced interest in extreme upward moral comparisons (despite their low diagnosticity) but did not demonstrate a similar reduced interest in extreme upward comparisons in economic (Experiment 6a) or athletic (Experiment 6b) domains.

Together, these findings show that moral comparisons differ from other comparisons in three important ways: Compared to other types of comparisons, moral comparisons are especially likely to be directed downward, upward moral comparisons are particularly threatening, and moral comparisons depend less on the diagnosticity of the comparison. This suggests that moral comparisons are a unique form of social comparison.

5.8.1 Implications for Morality Research

The current research has important implications for research on morality. A recent meta-analysis on social comparisons (Gerber et al., 2018) found that people preferred upward comparisons overall (i.e., across domains), and that threat only slightly reduced this effect. In contrast to these findings, our studies demonstrate that the threat principle is rather robust when it comes to moral social comparisons. For example, in Experiment 2a, threat increased the choice of downward comparisons from 51% to 72% and in Experiment 2b from 53% to 67%. Demonstrating the strength of the threat principle in moral comparisons, participants in Experiments 3a – c showed more interest in downward comparisons and less interest in upward comparisons after threat despite strong incentives not to do so. And Experiments 4a – b confirmed that the threat principle operates more strongly in the moral domain than in other domains. As a demonstration with greater external validity, our experience-sampling study also showed that people make more downward comparisons in morality than in most other domains in everyday life (Study 1).

As we proposed in the Introduction, people might engage in stronger motivated downward comparisons in morality because morality is uniquely important to people (Monin, 2007; Strohminger & Nichols, 2014); they value morality in a way that they do not value other domains. We measured domain importance in Study 1 but did not find that the effect depended on it. However, as this was a correlational study, we cannot rule out that the importance of a domain might still play a role in how

threating upward comparisons in a particular domain are. Future research could test this idea by experimentally manipulating the importance of different domains, measuring threat, and testing whether threat mediates the effect on increased downward comparisons.

In the Introduction, we also mentioned that moral comparisons might be more threatening because of the importance that morality plays in the judgments of others (Brambilla & Leach, 2014; Brambilla et al., 2011; Goodwin et al., 2014; Hartley et al., 2016; Landy et al., 2016; Pizarro & Tannenbaum, 2012; Wojciszke et al., 1998). People's moral identity is influenced not only by their desire to see themselves as moral, but also by their desire to present themselves as moral to others (Aquino & Reed, 2002). Consequently, people may choose downward comparisons as an impression management strategy. Previous research found that people self-enhance more in morality than in other domains (Allison et al., 1989; Van Lange & Sedikides, 1998; Ybarra et al., 2012). Furthermore, people alter their decisions in moral dilemmas to impact the impression they convey to others (Rom & Conway, 2018). In line with this reasoning, it is likely that people also seek to present themselves as not at all immoral, even more strongly than in other domains. Downward comparisons can serve that purpose because they improve how people's own immoral behavior is perceived. Therefore, the perception by others may moderate the current effects. Future research could test this mechanism by examining whether people engage in more downward comparisons when these comparisons are made in public versus in private or when observers have more versus less strict moral standards.

Alternatively, the effect of threat might depend on the centrality of that particular domain to the self, with different domains being differently central to different people. This might explain why the overall comparison direction found in the meta-analysis by Gerber et al. (2018) was upward, as most dimensions are peripheral and not central to most participants. For example, most professors would consider academic achievement as central but athletic achievement as peripheral to their identity, while the opposite is the case for most athletes. In contrast, morality is likely to be central to (almost) all people's self-concept (Monin, 2007; Strohminger & Nichols, 2014), and therefore threat effects do not differ between participants. In fact, morality is so consistently central to the self that it even produces qualitatively different processes. For example, moral convictions are different from other convictions in that they have effects above their attitude strength (Skitka et al., 2005), and moral

trade-offs are considered taboo (Atran et al., 2007; Tetlock, 2003; Tetlock et al., 2000). Overall, our findings suggest that people go to great lengths to uphold their self-view or self-presentation as a moral person.

It is noteworthy that gaining time or even money was not sufficient to remove increased interest in downward comparisons after a threat (Experiments 3a – 3c). Here, an interesting parallel can be drawn between our findings and work on moral trade-off taboos (Atran et al., 2007; Tetlock, 2003; Tetlock et al., 2000). Specifically, earlier research has found that people believe that certain moral values tend to be sacred—for example, the welfare of their family, or their commitment to justice. This taboo is so strong that merely even considering to violate such moral values for economic gain is disquieting; people react to the idea of making such trade-offs with moral outrage (Atran et al., 2007; Tetlock, 2003; Tetlock et al., 2000). In our experiments, it was clear that some people recalled having violated important moral values. For example, one participant wrote about lying to a close friend, while another participant wrote about not helping her sister when she was in dire need of help; a third wrote about cheating on her husband. Indeed, after recalling such events, we found that participants were willing to forego money in order to choose downward comparisons. This avoidance of moral threats is in line with the idea that people avoid thinking about taboo trade-offs when these trade-offs are necessary (Tetlock, 2003). It indicates that upward comparisons in morality are psychologically quite "costly", as can be seen by people foregoing money to avoid such comparisons. While this does not necessarily mean that "morality has no prize", it might suggest that if moral values have already been violated, higher amounts of money might be needed to force people to see that others chose to act morally.

5.8.2 Implications for Social Comparison Research

The current research also has important implications for social comparison research. One of our main findings was that the diagnosticity principle is reduced in morality; comparisons that are seemingly irrelevant, extreme, and removed in time and space are still threatening. This finding is contrary to a basic principle of social comparison research (Festinger, 1954; Tesser, 1988, 1991), which indicates that only close or relevant comparisons are diagnostic, and therefore, threatening. Our

experiments suggest that the classic diagnosticity principle is indeed a prerequisite for threatening comparisons in other domains, such as athleticism or economic success, but not for morality.

In the Introduction, we suggested that people consider a much broader range of comparison standards to be relevant in morality because of the absolute nature of morality. When drawing comparisons, people usually consider the context to select an appropriate standard to judge against. For example, in sports, beginners are judged by different standards than are professional athletes. A statement such as "You are a good dancer for a beginner" reflects this selection of appropriate standards and associated comparison group (Biernat & Manis, 2007; Biernat et al., 1991). That is, being a good dancer will mean something different for a beginner and a professional dancer. In contrast, in the moral domain it seems less likely that people use such shifting standards. A sentence "You are a good moral person for a beginner" is absurd. Instead, it seems likely that people predominantly judge moral behavior on an absolute scale, making even extreme and very different comparisons relevant.

Another reason why people take the context less into account in morality than in other domains may be perceived achievability. Comparisons can be made for ability and opinion domains (Festinger, 1954; Gibbons & Buunk, 1999), but only ability domains include the possibility of downward or upward comparisons. In such ability domains, people are likely to consider very extreme or different comparisons irrelevant as there are external circumstances that preclude people from having the same ability. For example, professional athletes are not relevant comparison standards for beginners of a sport. Unlike other ability domains, morality comparisons are based, in part, on opinion (Monin, 2007). Therefore, people might perceive that willpower or effort is more important than ability to achieve a moral standard. Everyone can be an extremely trustworthy and helpful person if they invest the effort and have the motivation. Therefore, different or extreme comparison standards become relevant in morality, as people believe they or others could achieve such standards with enough motivation and effort. Future research should measure whether people believe it takes ability versus willpower to achieve moral behavior and test whether this moderates how relevant people consider traditionally non-diagnostic comparison standards in the moral domain.

5.8.3 Future Research

One question for future research is to identify whether there are other domains where standard definitions of diagnosticity do not apply. If morality differs from other domains with regard to the diagnosticity principle due to the absoluteness of judgment, it is possible that morality is truly the sole exception to this basic principle of social comparison. This is because people view that morality is different from judgments of other preferences precisely because of its absoluteness (even if morality might not actually be absolute, Hare, 1981; Theriault et al., 2017). However, if comparisons in morality depend less on the diagnosticity principle due to the perception that morality is based on will power and effort, then diagnosticity may be also less relevant in other domains that are viewed through the lens of willpower. For example, people often consider health, such as healthy eating, dieting, and exercise, to be based more on people's will power, and less on their ability (Thibodeau & Flusberg, 2017). Therefore, health might be one such domain in which diagnosticity is less relevant, similar to morality.

Although the diagnosticity principle is reduced when making moral comparisons, there might be contexts where diagnosticity still matters. For example, humanity might be a necessary prerequisite, as morality is a panhuman standard. In support of this idea, Experiment S1 in the supplemental materials found that aliens were not seen as diagnostic comparisons. Future research may want to test this in more detail.

Another interesting idea for future research is to test whether the threat of moral comparisons depends on the intentionality of a moral actor. Intentionality does not play an important role in other domains. After losing money, we do not want to hear about someone who gained money, regardless of whether the money was gained through intentional action or by a lucky chance. But we judge the moral character of people by their intentions and desires, that is, we consider whether people intended to cause the moral or immoral action they produced. People who unintentionally do something good, such as someone who accidentally pushes someone away from a speeding car, are not considered moral (Cushman, 2015; Pizarro & Tannenbaum, 2012; Pizarro et al., 2003). It would be interesting to see whether accidental moral or immoral targets are also threatening, that is, whether the diagnosticity

principle in morality, instead of being based on closeness and similarity, is dependent on the intention and causality of an action, making some moral or immoral actions irrelevant.

An interesting finding across our experiments is that participants in the non-threatened conditions showed more interest in reading about moral exemplars than in reading about immoral exemplars. This finding begs the question of why people compare themselves upwardly in the moral domain. After all, even if the self is not threatened, a downward comparison makes it easier to feel good about the self than an upward comparison. In addition, reading about other people's immorality may be more exciting because it involves forbidden acts that people are less likely to experience themselves. However, one reason people may be especially likely to use upward comparisons in morality is the belief in a just world (Furnham, 2003; Lerner, 1980). People are motivated to think of the world as a moral and just place where bad things happen only infrequently and only to people who deserve them. Reading descriptions of immoral acts may be inherently threatening, reminding people that we are all vulnerable to getting hurt by immoral others. Future research might test whether this explains the general tendency for upward comparisons.

An additional interesting finding in our experiments is that participants often did not show a higher interest in immoral than moral comparisons even in the threatened condition. This is in line with the results of Gerber et al. (2018). In contrast to that, in several experiments, participants did show a stronger tendency to choose downward than upward comparisons under threat. These results are supported by the findings in Study 1, in which participants chose more downward than upward comparisons in morality in their everyday lives. One reason for this disparity could lie in virtue signaling or impression management. As mentioned before, people's moral identity is determined not only by how moral they see themselves, but also by how moral they present themselves to others (Aquino & Reed, 2002). It is possible that when people only have to indicate their interest, there is a general trend to express more interest in upward comparisons to show that one is a virtuous person. But when people actually have to engage in the comparison (in the experiments or in real life), there is an overall shift to more downward comparisons.

Finally, it would be interesting to test whether our results extend to automatic comparison processes. For automatic comparison processes, people usually assimilate to moderate standards and

contrast away from extreme standards. This is the case because people are more likely to test for similarities with the former, and more likely to test for differences with the latter, as their first holistic assessment signals similarities or differences (Mussweiler, 2003; but see Barker et al., 2020, for more nuanced evidence). For morality, this first holistic assessment might differ. As an absolute standard, moral judgments declare something either moral or immoral, regardless of extremity. Therefore, as long as both the self and the comparison standard are moral (or both immoral), people's first holistic assessment might point to similarities—and therefore lead people to search for further similarities and then assimilate to the standard—even for extreme comparisons. In contrast to that, when the self is immoral and the comparison standard moral (or the opposite), even moderate standards might lead people to see differences and engage in further difference testing, until they then contrast away from the standard.

5.8.4 Strengths and Limitations

Our range of research paradigms offered many advantages in answering our research question. Our experience-sampling study offered high external validity while our controlled experiments provided high internal validity. Additionally, our experiments showed that participants engaged in strong motivated downward comparisons in private, a context where threats should not be particularly threatening. As Study 1 showed, these same effects occur in everyday life, where the threat to people's moral self is stronger due to the potential negative judgment by others.

All participants in our experiments were users of Amazon Mechanical Turk. Such MTurkers are more representative of the U.S. population than most convenience samples and studies conducted with MTurk samples have been shown to produce similar results as studies with more expensive, nationally representative samples (Berinsky et al., 2012; Mullinix et al., 2016). At the same time, we acknowledge that our studies were limited to one culture and that we did not test our hypotheses outside the United States. One particularly important cross-cultural difference that future research should test is independence–interdependence (Markus & Kitayama, 1991) We tentatively suggest that our results would generalize to more interdependent cultures because of the increased concerns around social approval and disapproval (Singelis, 1994).

Another potentially relevant cultural difference is that Western participants are more liberal than people in other parts of the world such as Asia, Africa, or South America (Haidt & Graham, 2007; Haidt & Joseph, 2006). Future research should test whether the effects hold for more conservative participants. Previous research has established that conservatives and liberals differ on which moral foundations they find important, with conservatives valuing five moral foundations and liberals valuing only two. Consequently, conservatives face a wider range of threats to the moral self (e.g. ingroup/loyalty, authority/respect, and purity/sanctity, Graham et al., 2009; Haidt & Graham, 2007), which might lead conservatives to search for more downward comparisons than liberals.

5.8.5 Conclusion

In a world that is filled with sinners and saints, people can often choose to focus on the sinners. This focus on the sinners and avoidance of saints is both similar to and dramatically different from other types of social comparisons. The current studies found that people tend to compare more downward in morality in their everyday life than in most other domains. Motivated downward comparisons after threat are stronger in morality than in many other domains, leading people to even reduce their utility to avoid upward comparisons. But when selecting comparison partners, people are less sensitive to the criterion of diagnosticity, making even extreme or distant comparison relevant and threatening. Overall, morality intensifies the threat principle, but upends the diagnosticity principle, leading people to avoid almost any saint and take refuge in comparisons to any sinner they can find.

Chapter VI: General Discussion

People orient themselves to their social environment in order to get feedback on their abilities and behavior, to distinguish themselves positively from others, or to be inspired by the success of others (Festinger, 1954a; Taylor & Lobel, 1989a; Wheeler, 1966; Wills, 1981a). In the already extensive literature on social comparison, however, there still remains a gap that needs to be filled. The whole spectrum of motivational outcomes across all possible social comparisons (from extreme downward to extreme upward comparisons) has not been covered. Previous known models focused mainly on perception and appraisal of comparison standards and have only limited applicability to motivational processes (Diel & Hofmann, 2019). This dissertation aims to fill this gap by developing a new framework that brought together two literatures: past work on self-regulation as a motivational process to reach personally endorsed goals and the literature on social comparison. In this manner, the motivational effort to achieve something worthwhile as a function of social comparison processes is best explained by integrating basic tenets of the self-regulation approach instead of purely focusing on principles of assimilation and contrast judgments. With this new look at social comparison and goal pursuit, the three motives - self-evaluation, self-enhancement and self-improvement - are no longer the ultimate goals of comparison processes but rather represent different stages of the self-regulation feedback loop. This way, people can flexibly decide when and how much effort they need to invest at a given point with close linkage to their social surroundings. In this dissertation, the theoretical framework was introduced and subsequently supported by empirical studies from both the lab and the field. First of all, Chapter 2 showed that the classical social-cognitive approach of social comparison judgments does not fully provide answers with regard to motivation and goal pursuit. For instance, the comparison to an extreme downward standard did not result in an increase in motivation and effort but rather resulted in the opposite (i.e. almost complete effort withdrawal). Chapter 3 introduced the full new framework of motivation, which was tested by means of an experience sampling study that included a heterogeneous sample and covered a wide range of comparison domains. The extensive study demonstrated for the first time that different discrepancy assessments between the self and a standard of comparison are associated with the predicted motivational patterns of pushing, coasting and disengagement. A negative discrepancy was related to a motivational increase (pushing), however,

this relationship was nonlinear in nature as motivation and effort decreased towards more extreme ends of upward comparison. At the same time, people were more likely to give up with increasing negative discrepancies (disengagement). More specifically, pushing effects were even more pronounced for people who perceived their comparison domain as important or who perceived to be in control over their comparison domain, which is in line with previous research on goal commitment (Fishbach & Ferguson, 2007; Kruglanski et al., 2002) and the attainability of upward standards (Lockwood & Kunda, 1997). So, if discrepancies were too large, the standard seems out of reach with detrimental effects on motivation and effort.

Of particular interest was also the motivational nature of downward comparison. So far, the literature had linked downward comparison to an increase in self-esteem and positive emotions (Morse & Gergen, 1970; Taylor & Lobel, 1989; Wills, 1981), but has not made comprehensive statements about possible motivational consequences, especially toward the more extreme ends of inferior comparison standards. In that sense, Chapter 2 revealed that a moderate downward standard was less inspiring than upward standards but more motivating than an extreme downward standard, which was simply perceived as different from the self and hardly motivating. In Chapter 3, a positive discrepancy (downward comparison) was related to a reduction of motivation and effort (coasting). This way, the positive affective consequence that had been linked to downward comparison may now be seen as a byproduct of the notion that a person has surpassed a standard of comparison.

On top of the motivational markers of pushing, coasting, and disengagement, Chapter 4 added a behavioral measure of effort investment, namely real-life performance scores in athletes. Contrary to our hypothesis, athletes did not show a bias towards upward comparison, but still, upward comparison was related to increased performance improvement. Also here, motivation in general was associated with upward comparison but started to decrease with more extreme upward comparisons. Again, disengagement was related to upward comparison, whereas coasting rose with downward comparison.

Of secondary interest but nevertheless indispensable are the emotional correlates of discrepancy assessments. In line with previous research, negative discrepancies are related to negative emotions (e.g. guilt, lower self-esteem; Carver & Scheier, 1990; Morse & Gergen, 1970) while positive discrepancies come with positive emotions (e.g. pride, higher self-esteem; Carver & Scheier, 1990,

2011; Morse & Gergen, 1970; Taylor & Lobel, 1989; Wills, 1981). However, a crucial difference between the motivational and emotional patterns remains: whereas motivation decreased with more extreme upward comparisons, negative emotions continued to increase. In other words, people disengage motivationally but not emotionally. Also, one would expect that positive emotions decrease towards the more extreme ends of downward comparisons as outperforming a very unsuccessful person might be of lower value than outperforming someone who is somewhat similar to the self (i.e. moderate downward standard). However, Chapter 3 and 4 show that the larger the positive discrepancy becomes the better people feel about themselves. The framework of social comparison-based emotions (Smith, 2000) may provide answers to why positive as well as negative emotions grow linearly with positive (or negative) discrepancies instead of dropping towards extreme ends. If a negative discrepancy appears to be unchangeable, negative emotions are aroused, and even more so if the standard of comparison is similar to the self, or if the domain of comparison is of high importance. Therefore, the greater the discrepancy (and thus the lower the perceived control), the stronger the negative feelings become. On the other hand, negative feelings are diminished if there is a feeling of control to work on one's deficits. With positive discrepancies, the perception of control grows, which in turn arouses positive feelings (Major et al., 1991). However, if the discrepancies are barely positive, the perception of control is diminished, which can even trigger negative feelings because there is a risk of undergoing the same misfortune (Major et al., 1991; Smith, 2000). This again emphasizes the notion that the experience of positive emotions following downward comparison is not the final goal (i.e. the fulfillment of the self-enhancement motive) but rather accompanies the realization of having made sufficient progress towards a goal.

Lastly, Chapter 5 addressed a rather unique domain of comparison, namely the domain of morality, where the motivation to be a morally good person deviates from other goal-related domains and thus from the previous framework. People who want to appear as morally good do not orient themselves towards moral role models, but are more inclined to focus on downward comparison to put themselves in a more favorable light.

In sum, viewing social comparison from a self-regulatory instead of an overly cognitive view gives a better understanding of the motivating but also discouraging nature of different comparison

standards. The motivational framework now illustrates when upward comparisons are inspiring (i.e. resulting in pushing) and when they tend to lead to the opposite (i.e. resulting in disengagement). It also fills the gap in the literature by putting the motivational potential of downward comparisons into a new perspective (i.e. coasting). However, not all domains may function according to the pushing, coasting, and disengagement principles. Rare exceptions, such as moral comparisons, are driven by different motivations with the ultimate goal to appear as a morally impeccable person.

6.1 Implications and Future Directions

Upward comparison can inspire people unless certain standards are perceived to be "too far off" relative to the individual, which in turn hampers motivation and evokes ill feelings. Still, people are frequently exposed to extremely successful others, especially through the constantly available social media (Appel et al., 2015). Thus, people may strive towards the unreachable success of others by overrating the relevance of an extreme upward comparison (Strahan et al., 2006). In some instances, people even show a trend toward continuous self-optimization that presumably prevents them from settling on moderate and attainable goals (e.g. striving for the perfect body, King et al., 2019). In this regard, Chapter 2 shows that people do not indicate a drop in motivation when comparing to an extremely superior person compared to a moderately superior person. Even though the standards' exercise performance went through the roof, people still felt encouraged by the extremely upward standard. This is in line with people's general tendency to believe they are better than the average (Alicke, 1985) and their proneness to set unrealistic high goals (Schwarzer, 1998; Taylor & Brown, 1998; Buehler et al., 1994), paired with positive illusions and false hopes (Polivy, 2000). Thus, people might also be inclined to look up to unrealistic high comparison standards who are far off from attainability. The constant pursuit of overly ambitious goals in addition to ineffective strategies to reach those goals and the anticipated happiness and joy, can actually result in the opposite (Ford & Mauss, 2014). Chapter 3 demonstrates the detrimental effects of extreme upward comparisons including feelings of "giving up" and a harsh increase in negative affect. Especially in relation to social media, people need to be aware that their personal goals influence their media exposure, which in turn reinforces their adherence to their goals. On the other hand, media exposure can even trigger goals about possible future selves that had not been there before, especially in relation to beauty ideals

(Knobloch-Westerwick et al., 2020). Hence, it is important to make social comparison choices that consist of comparisons to realistic and attainable standards in order to maximize motivational potential and minimize emotional suffering. For instance, instead of exposing the self to unrealistic standards on social media, people can change their social media input by following people who are perceived as inspiring but are still more similar and relevant to the self. Future research should examine to what extent the purposeful manipulation of one's social media exposure can benefit people. In this regard, intervention studies should foster more strategic media consumption. This way, social comparisons might facilitate goal pursuit by concentrating on the most beneficial role models.

After all, it is important that role models exist who have achieved extraordinary things that are likely to inspire others. Hence, another way to not permanently suffer from the pursuit of unrealistic and unattainable standards is to reduce the perceived discrepancy between the self and the outstanding success of others. In other words, in some domains the success of other people should become more attainable for the ordinary person. For example, women in leadership positions still seem to be the exception and those individuals may seem out of reach for the average female employee. Research has already shown that female role models are important and beneficial in domains where woman are still underrepresented (Buck et al., 2008). However, research also demonstrates that women were less able to identify with elite-female leaders in contrast to non-elite leaders, which in turn negatively influenced their own leadership aspiration (Hoyt & Simon, 2011). On the other hand if leadership success is perceived as attainable through the believe that the required skills are obtainable rather than fixed (Hoyt et al., 2012; Major et al., 1991) or by means of high leadership self-efficacy (Hoyt, 2013), assimilation to exceptional female role models are more likely. The introduction of a women's quota could lead to a reduction in the discrepancy between normal female employees and their career goals, by making the image of a woman in a management position more common and giving women the necessary perceived control to achieve their goal or successfully follow their standard of reference. Finally, perceived attainability increases women's interest in a career path that at the same time contradicts current stereotypes (Luong et al., 2020). Future research should focus on strategies to further minimize discrepancies between individuals and exceptional others especially in areas where certain groups are underrepresented.

At times, however, it might be worthwhile to pause the wheel of success pursuit and practice gratitude for what has been achieved so far. The risk of constant comparison is that the social environment will always remind someone of what has not been achieved (yet). As the "social ocean" we swim in is endless, there will always be someone better, and even if one area in life seems to be sufficient, there will be another area that needs fixing and calls for improvement. At the same time, people quickly get used to the positive aspects in their lives and thus, take their successes for granted which makes them want to strive for more (Lyubomirsky, 2010). Instead of constantly looking up, the comparison to less successful others helps to practice gratitude (Nicuță & Constantin, 2021). Hence, instead of following their self-improvement motivation, people may benefit from their motivation to occasionally rest and look back to what has been achieved so far (i.e. coasting motivation). Furthermore, the practice of self-compassion – treating oneself with kindness even in events of failure (Neff, 2003) - could reduce harmful upward comparisons on social media (Seekis et al., 2020). And paradoxically, practicing self-compassion after a personal failure can in turn boost people's motivation to self-improve (Breines & Chen, 2012). Thus, taking other people as reminders of one's own successes and simultaneously being grateful and self-compassionate can be a beneficial strategy for improving one's well-being without harming future self-improvement motivation. Again, intervention studies could investigate long-term effects of frequent downward comparison as a strategy to increase gratitude and self-compassion while minimizing detrimental social comparison effects.

Upward comparison can inspire people under certain conditions, for instance, to live a healthier lifestyle or to save money. On the other hand, in the domain of morality motivation is driven differently. People are biased toward downward comparison with the ultimate goal to distinguish their moral character positively from others. This might be problematic if the greater goal is to make people aware of their moral flaws in order to pursuit improved behaviors. For instance, with regard to environmentally friendly behavior, each individual is challenged to rethink their behavior in everyday life and to make better (often moral) decisions in several areas concerning the environment. One might assume that positive role models may reinforce this behavior. However, as shown in Chapter 5, upward comparisons of moral behavior are perceived as a great threat and therefore upward comparisons tend to be avoided to protect the self (e.g. if vegetarians frame their abstinence from meat

morally, they are more disliked by meat-eating persons, Cramwinckel et al., 2013). Instead, people rather focus on downward comparison in order to present themselves in a moral light and to presumably also justify their own misbehavior (e.g. self-licensing effect; Merritt et al., 2010). One way to prevent this, may be to not frame climate-targeting behavior in moral terms. For instance, if certain others promote everyday water savings, instead of purely addressing people's morality ("Too much water use hurts the environment"), an additional goal focus can be incorporated ("Saving water benefits the environment *and* you save money"). Another approach might be to first acknowledge people's moral behavior (in contrast to evoke threat) and then bring in an upward standard to encourage further improvements. Future studies addressing sustainable behavior change should take those findings into account and find alternative ways to change people's behavior without appealing to their morality.

Building on our research, future studies could investigate specific social comparison strategies people use over a period of time with regard to their long-term goals. In this regard, it would be interesting to see whether people tend to stick to a specific upward standard in relation to a goal as a source of inspiration and motivation, or also adapt these standards strategically. For example, if the discrepancy to a certain standard has been reduced, depending on a further goal, a new standard could be sought that pushes the person even further in the desired direction. Similarly, it might be worthwhile to investigate whether people who run the risk of giving up because the perceived discrepancy to a standard is too great, may spontaneously row back and measure themselves against a more moderate standard (i.e. to push motivation), rather than abandoning the goal and their initial motivation altogether. Moreover, future research could also distinguish between upward comparison regarding a final state ("I want to be like this person") or in relation to targeted strategies ("I want to act like this person to become better"), which would affect the perceived attainability of those standards. For instance, whereas the success of a certain other might be unattainable, the adoption of certain strategies that helped to achieve their success over the years might be not. And lastly, future studies can examine whether coasting effects could be buffered. For instance, the feeling of pride stemming from a positive discrepancy to a standard could be a new basis for future motivation. As past research shows, pride can facilitate perseverance on difficult tasks (Williams & DeSteno, 2008).

Hence, assessing a positive discrepancy to a standard would not necessarily result into resting on what has been achieved but could serve as a new motivational push in the desired direction by experiencing certain emotions.

6.2 Limitations

Some limitations have to be mentioned. Chapter 2 tested motivational outcomes of different social comparison standards who differed in comparison direction (upward vs. downward) and their extremity (moderate vs. extreme). One limitation to mention is that those effects were only investigated in the domain of fitness. While improving health may be a common goal for the majority of people, personal fitness, especially one trained by specific workouts, may not be of interest for many people who might suffer from physical impairments, do not have time for workouts or simply do not like working out. Testing motivational effects in a domain that is of relevance and importance to participants might reveal larger effects. Second, the outcome variable of motivation was based on self-report and thus only represent intention rather than behavior. Although I included a behavioral measure in the last study in Chapter 2, the manipulation did not have an influence on this measure. In general, implementing behavioral measures becomes more and more difficult as the number of online experiments compared to lab experiments rises due to the demand for larger samples. This way, online studies that often involve self-report measures are easier to conduct instead of the recruitment of a high number of participants for experiments in the laboratory (Sassenberg & Ditrich, 2019). Chapter 3 and 4 comprise of experience sampling studies that have the advantage over Chapter 2 to cover a wide range of comparison domains (Chapter 3) and even real-life behavior (Chapter 4). Still, the method of experience sampling comes with additional limitations. First of all, all analyses are strictly correlational and thus, no causality can be established. In this sense, it has not yet been fully resolved whether a comparison to a successful person simply fosters motivation or whether increased motivation in the first place leads people to orient themselves more towards successful people. What the study also missed to capture in this context are the different goals people may pursue in their everyday life, including information on how far they are from their goal or whether the goal has already been achieved in the past.

However, not all comparisons are intentional (Mussweiler et al., 2004; Ohmann et al., 2016) and in this sense, even an extensive experience sampling study might not have been able to detect especially those comparisons in a day that did not happen on purpose but rather occurred automatically. In particular, the effects of comparisons which happen spontaneously and which are harder to escape from are intriguing. Still, the study in Chapter 3 made an effort to capture some unintentional comparisons (e.g. people on the streets) people might be aware of later, which is still an advantage over experimental studies that test either selection studies (consciously deciding on a comparison standard) or reaction studies (being exposed to a random standard; for an overview, see Gerber et al., 2017).

Another limitation of the experience sampling study is the rather subjective judgment of discrepancy assessments between the self and the comparison standard. This way, the discrepancies are already relative in nature instead of being composed of absolute and objective differences between two people, which is easier to implement in experimental studies. This way, it is not always distinguishable, for instance, if an upward standard just exhibits an exceptionally good standing or if the target person's standing is simply very poor. Thus, future research should additionally capture the participant's self-standing relative to, for example, the population average. Another disadvantage of the self-measures is the vulnerability of social desirability (Scollon et al., 2009) or more favorable self-perceptions (Alicke & Sedikides, 2011) as people often overestimate themselves and their abilities. However, the advantage of experience sampling is that the within-person variance (the differences between one comparison in a day to another) generates the most power in comparison to the between-person variance and thus, social desirability effects stemming from the between-person variance can be buffered to some extent. Lastly, in order to keep items in the multiple daily questionnaires to a minimum, constructs are measured only on single items. Still, the reliability of these reduced items is compensated by the large degree of statistical power gained from the high number of repeated measures per participant (Bolger & Laurenceau, 2013).

In addition, Chapter 4 includes a behavioral measure of real-life performance in young athletes consisting of the difference between the performance at the beginning of the semester (i.e. swimming times) and the performance at the end of the semester. Unfortunately, not all participants were able to

provide both scores. Thus, the analysis of comparison direction as a predictor of performance improvement was based on a small sample. The result is to be treated with caution.

Finally, all studies include sample from Western-societies, that is the United States and Germany. Results might be culturally influenced as social comparison processes vary across different cultures (Baldwin & Mussweiler, 2018). Future research should draw on other cultural groups to replicate findings or rather reveal important differences.

6.3 Conclusion

Social comparisons are an integral part of everyday life, but the comparison to other people comes with merits and drawbacks – not only in the emotional but also in the motivational sense. This dissertation combines social comparison with key insights from self-regulation research to give a complete overview about motivational "gains and pains" of the whole spectrum of possible social comparisons. Detrimental effects are most likely to be driven by extreme and unattainable standards (disengagement). Thus, a better strategy is to select comparison standards according to similarity and accessibility, so that discrepancies that arise are easier to minimize (pushing). A positive discrepancy comes with emotional benefits but with the motivational tendency to temporarily rest on what has been already achieved (coasting). However, certain domains may deviate from this pattern: in the morality domain people are mostly motivated by downward comparisons. As implications, people should be aware of the impacts of other people concerning motivation and emotion and generally refrain from overachievements not only with regard to goal setting but also concerning comparison standards. After all, we need role models in our life, but these role models should not be beyond our reach.

References

Alicke, M. D. (2007). In defense of social comparison. *Revue Internationale de Psychologie Sociale, Tome, 20*(1), 11–29.

Alicke, M. D., & Sedikides, C. (2011). *Handbook of self-enhancement and self-protection*. The Guilford Press.

Allen, M. S., Knipler, S. J., & Chan, A. Y. C. (2019). Happiness and counterfactual thinking at the 2016 Summer Olympic Games. *Journal of Sports Sciences*, *37*(15), 1762–1769. https://doi.org/10.1080/02640414.2019.1592803

Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. *Social Cognition, 7*(3), 275-295. https://doi.org/10.1521/soco.1989.7.3.275

Ames, C. (1995). Achievement goals, motivational climate, and motivational processes. In G. C. Roberts (Ed.), *Motivation in sport and exercise* (pp. 161–176). Human Kinetics Publishers.

Appel, H., Crusius, J., & Gerlach, A. L. (2015). Social comparison, envy, and depression on Facebook: A study looking at the effects of high comparison standards on depressed individuals. *Journal of Social and Clinical Psychology*, *34*(4), 277–289. https://doi.org/10.1521/jscp.2015.34.4.277

Aquino, K., & Reed, A., II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423-1440. https://doi.org/10.1037/0022-3514.83.6.1423

Aspinwall, L. G., Hill, D. L., & Leaf, S. L. (2002). Prospects, pitfalls, and plans: A proactive perspective on social comparison activity. *European Review of Social Psychology*, *12*(1), 267–298.

Atkinson, J. W. (1964). *An introduction to motivation*. Van Nostrand.

Atran, S., Axelrod, R., & Davis, R. (2007). Sacred Barriers to Conflict Resolution. *Science, 317*(5841), 1039–1040. https://doi.org/10.1126/science.1144241

Baillargeon, R. (1991). Reasoning about the height and location of a hidden object in 4.5- and 6.5-month-old infants. *Cognition, 38*(1), 13-42. https://doi.org/10.1016/0010-0277(91)90021-U

Baldwin, M., & Mussweiler, T. (2018). The culture of social comparison. *Proceedings of the National Academy of Sciences*, *115*(39), E9067–E9074. https://doi.org/10.1073/pnas.1721555115

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269-278. https://doi.org/10.1037/1528-3542.6.2.269

Bardel, M.-H., Fontayne, P., Colombel, F., & Schiphof, L. (2010). Effects of match result and social comparison on sport state self-esteem fluctuations. *Psychology of Sport and Exercise*, *11*(3), 171–176. https://doi.org/10.1016/j.psychsport.2010.01.005

Barker, P., Dotsch, R., & Imhoff, R. (2020). Assimilation and contrast in spontaneous comparisons: Heterogeneous effects of standard extremity in facial evaluations. *International Review of Social Psychology, 33*(1), 11. https://doi.org/10.5334/irsp.402

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berger, J., & Pope, D. (2011). Can losing lead to winning? *Management Science*, *57*(5), 817–827. https://doi.org/10.1287/mnsc.1110.1328

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20*(3), 351-368. https://doi.org/10.1093/pan/mpr057

Bessem, B., de Bruijn, M. C., & Nieuwland, W. (2017). Gender differences in the electrocardiogram screening of athletes. *Journal of Science and Medicine in Sport*, *20*(2), 213–217. https://doi.org/10.1016/j.jsams.2016.06.010

Biernat, M., & Manis, M. (2007). Stereotypes and shifting standards: Assimilation and contrast in social judgment. In D. A. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology.* (pp. 75-97). Psychology Press.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology, 60*(4), 485-499. https://doi.org/10.1037/0022-3514.60.4.485

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive Longitudinal Methods An Introduction to Diary and Experience Sampling Research*. Guilford Press.

Bounoua, L., Cury, F., Regner, I., Huguet, P., Barron, K. E., & Elliot, A. J. (2012). Motivated use of information about others: Linking the 2×2 achievement goal model to social comparison propensities and processes. *British Journal of Social Psychology*, *51*(4), 626–641. https://doi.org/10.1111/j.2044-8309.2011.02027.x

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition, 32*(4), 397-408. https://doi.org/10.1521/soco.2014.32.4.397

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*(2), 135-143. https://doi.org/10.1002/ejsp.744

Brawley, A. M., & Pury, C. L. S. (2016). Work experiences on MTurk: Job satisfaction, turnover, and information sharing. *Computers in Human Behavior, 54*, 531-546. https://doi.org/10.1016/j.chb.2015.08.031

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*(1), 109–131. https://doi.org/10.1146/annurev.ps.40.020189.000545

Breines, J., & Chen, S. (2012). Self-compassion increases self-improvement motivation. *Personality & Social Psychology Bulletin*, *38*(9), 1133–1143. https://doi.org/10.1177/0146167212445599

Buck, G. A., Clark, V. L. P., Leslie-Pelecky, D., Lu, Y., & Cerda-Lizarraga, P. (2008). Examining the cognitive processes used by adolescent girls and women scientists in identifying science role models: A feminist approach. *Science Education*, *92*(4), 688–707. https://doi.org/10.1002/sce.20257

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the" planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, *67*(3), 366-381.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5. https://doi.org/10.1177/1745691610393980

Caltabiano, M. L., & Ghafari, N. (2011). The effect of social comparison processes, athletic prototypes, perceived behavioural control and perceived threat on physical activity. *International Journal of Health Promotion and Education*, *49*(2), 44–52. https://doi.org/10.1080/14635240.2011.10708208

Carver, C. S. (2003). Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition and Emotion*, *17*(2), 241–261. https://doi.org/10.1080/02699930302294

Carver, C. S. (2004). Self-regulation of action and affect. In R.F. Baumeister & K.D. Vohs (Eds.) *Handbook of self-regulation: Research, theory, and applications.* (pp. 13–39). The Guilford Press.

Carver, C. S., & Scheier, M. F. (1981). *Attention and Self-Regulation: A Control-Theory Approach to Human Behavior*. Springer New York. https://doi.org/10.1007/978-1-4612-5887-2

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality-social, clinical, and health psychology. *Psychological Bulletin*, *92*(1), 111–135. https://doi.org/10.1037/0033-2909.92.1.111

Carver, C. S., & Scheier, M. F. (1990). Principles of self-regulation: Action and emotion. In E.T. Higgins and R.M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior, Vol. 2.* (pp. 3–52). The Guilford Press.

Carver, C. S., & Scheier, M. F. (2005). Engagement, disengagement, coping, and catastrophe. In A.J. Elliot & C.S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 527–547). Guilford Publications.

Carver, C. S., & Scheier, M. F. (2011). Self-regulation of action and affect. In K.D. Vohs & R.F. Baumeister (Eds), *Handbook of self-regulation: Research, theory, and applications, 2nd ed* (pp. 3–21). Guilford Press.

Cerin, E., & Barnett, A. (2006). A processual analysis of basic emotions and sources of concerns as they are lived before and after a competition. *Psychology of Sport and Exercise*, *7*(3), 287–307. https://doi.org/10.1016/j.psychsport.2005.07.002

Chanal, J. P., Marsh, H. W., Sarrazin, P. G., & Bois, J. E. (2005). Big-fish-little-pond effects on gymnastics self-concept: Social comparison processes in a physical setting. *Journal of Sport and Exercise Psychology*, *27*(1), 53–70. https://doi.org/10.1123/jsep.27.1.53

Chapman, G. B., Colby, H., Convery, K., & Coups, E. J. (2016). Goals and social comparisons promote walking behavior. *Medical Decision Making*, *36*(4), 472–478. https://psycnet.apa.org/doi/10.1177/0272989X15592156

Cho, K. W. (2019). Exploring the dark side of exposure to peer excellence among traditional and nontraditional college students. *Learning and Individual Differences*, 73, 52–58. https://doi.org/10.1016/j.lindif.2019.05.001

Collins, R. L. (1996). For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin*, *119*(1), 51-69. https://doi.apa.org/doi/10.1037/0033-2909.119.1.51

Collins, R. L. (2000). Among the better ones. In J. Suls & L. Wheeler (Eds.), *Handbook of Social Comparison: Theory and Research* (pp. 159–171). Springer US. https://doi.org/10.1007/978-1-4615-4237-7_9

Committee, I. O. (2021). *The Olympic motto*. https://olympics.com/ioc/olympic-motto

Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin, 38*(7), 907-919. https://doi.org/10.1177/0146167212442394

Corcoran, K., Crusius, J., & Mussweiler, T. (2011). Social comparison: Motives, standards, and mechanisms. In D. Chadee (Ed.), *Theories in social psychology* (pp. 119–139). Wiley Blackwell.

Corcoran, K., & Mussweiler, T. (2009). The efficiency of social comparisons with routine standards. *Social Cognition*, *27*(6), 939–948. https://doi.org/10.1521/soco.2009.27.6.939

Coyne, S. M., McDaniel, B. T., & Stockdale, L. A. (2017). "Do you dare to compare?" Associations between maternal social comparisons on social networking sites and parenting, mental health, and romantic relationship outcomes. *Computers in Human Behavior*, *70*, 335–340. https://doi.org/10.1016/j.chb.2016.12.081

Cramwinckel, F. M., Van den Bos, K., & Van Dijk, E. (2015). Reactions to morally motivated deviance. *Current Opinion in Psychology, 6*, 150-156. https://doi.org/10.1016/j.copsyc.2015.08.007

Cramwinckel, F. M., Van den Bos, K., & Van Dijk, E. (2016). *Negative reactions to non-drinkers: Self- and other-directed aggression after confrontations with moral refusers.* Manuscript submitted for publication.

Cramwinckel, F. M., van Dijk, E., Scheepers, D., & van den Bos, K. (2013). The threat of moral refusers for one's self-concept and the protective function of physical cleansing. *Journal of Experimental Social Psychology*, *49*(6), 1049–1058. https://doi.org/10.1016/j.jesp.2013.07.009

Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes*, *69*(2), 117–132.

Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology, 6*, 97-103. https://doi.org/10.1016/j.copsyc.2015.06.003

Davidai, S., & Deri, S. (2019). The second pugilist's plight: Why people believe they are above average but are not especially happy about it. *Journal of Experimental Psychology: General*, *148*(3), 570–587. https://doi.org/10.1037/xge0000580

Davidai, S., Deri, S., & Gilovich, T. (2020). There must be more to life than this: The impact of highly-accessible exemplars on self-evaluation and discontent. *Self and Identity*, *20*(1), 72-93. https://doi.org/10.1080/15298868.2020.1779121

de Vries, D. A., & Kühne, R. (2015). Facebook and self-perception: Individual susceptibility to negative social comparison on facebook. *Personality and Individual Differences*, *86*, 217–221. https://doi.org/10.1016/j.paid.2015.05.029

Deri, S., Davidai, S., & Gilovich, T. (2017). Home alone: Why people believe others' social lives are richer than their own. *Journal of Personality and Social Psychology*, *113*(6), 858–877. https://doi.org/10.1037/pspa0000105

Diel, K., Grelle, S., & Hofmann, W. (2021). A motivational framework of social comparison. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspa0000204

Diel, K., & Hofmann, W. (2019). Inspired to perspire: The interplay of social comparison direction and standard extremity in the context of challenging exercising goals. *Social Cognition*, *37*(3), 247–265. https://doi.org/10.1521/soco.2019.37.3.247

Duda, J. L., & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology*, *84*(3), 290–299. https://doi.org/10.1037/0022-0663.84.3.290

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237–251. https://doi.org/10.3102/0013189X15584327

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040–1048. https://doi.org/10.1037/0003-066X.41.10.1040

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional and personality development* (pp. 1017–1095). John Wiley & Sons, Ltd.

Ecker, Y., & Gilead, M. (2018). Goal-directed allostasis: The unique challenge of keeping things as they are and strategies to overcome it. *Perspectives on Psychological Science*, *13*(5), 618–633. https://doi.org/10.1177/1745691618769847

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, https://doi.org/10.1177/1073191120957102

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Society for industrial and applied mathematics.

Elbe, A.-M., Beckmann, J., & Szymanski, B. (2003). Die Entwicklung des allgemeinen und sportspezifischen Leistungsmotivs von SportinternatsschülerInnen. *Psychologie und Sport*, *10*(4), 134–143.

Elbe, A.-M., Wenhold, F., & Beckmann, J. (2008). *SOQ: Fragebogen zur Erfassung der Leistungsorientierung im Sport*. www.bisp.de

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Fardouly, J., Pinkus, R. T., & Vartanian, L. R. (2017). The impact of appearance comparisons made through social media, traditional media, and in person in women's everyday lives. *Body Image*, *20*, 31–39. https://doi.org/10.1016/j.bodyim.2016.11.002

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Feltz, D. L., & Lirgg, C. D. (2001). Self-efficacy beliefs of athletes, teams, and coaches. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (2nd ed., pp. 340–361).

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Filby, W. C. D., Maynard, I. W., & Graydon, J. K. (1999). The effect of multiple-goal strategies on performance outcomes in training and competition. *Journal of Applied Sport Psychology*, *11*(2), 230–246. https://doi.org/10.1080/10413209908404202

Finlay, W. M. L., & Lyons, E. (2000). Social categorizations, social comparisons and stigma: Presentations of self in people with learning difficulties. *British Journal of Social Psychology, 39*(1), 129-146. https://doi.org/10.1348/014466600164372

Fishbach, A., & Ferguson, M. J. (2007). The goal construct in social psychology. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles, 2nd ed.* (pp. 490–515). New York, NY, US: The Guilford Press.

Fogarty, G. J., & Else, D. (2005). Performance calibration in sport: Implications for self-confidence and metacognitive biases. *International Journal of Sport and Exercise Psychology*, *3*(1), 41–57. https://doi.org/10.1080/1612197x.2005.9671757

Ford, B., & Mauss, I. (2014). The paradoxical effects of pursuing positive emotion: When and why wanting to feel happy backfires. In J. Gruber & J. T. Moskowitz (Eds.), *Positive emotion: Integrating the light sides and dark sides* (pp. 363–381). New York, NY, US: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199926725.001.0001

Forstmann, M., Burgmer, P., & Mussweiler, T. (2012). The mind is willing, but the flesh is weak. The effects of mind-body dualism on health behavior. *Psychological Science*, *23*, 1239–1245.

Fulford, D., Johnson, S. L., Llabre, M. M., & Carver, C. S. (2010). Pushing and coasting in dynamic goal pursuit: Coasting is attenuated in bipolar disorder. *Psychological Science*, *21*(7), 1021–1027. https://doi.org/10.1177/0956797610373372

Furnham, A. (2003). Belief in a just world: Research progress over the past decade. *Personality and Individual Differences, 34*(5), 795-817. https://doi.org/10.1016/S0191-8869(02)00072-7

Garcia Bengoechea, E., & Strean, W. B. (2007). On the interpersonal context of adolescents' sport motivation. P*sychology of Sport and Exercise, 8*(2), 195–217. https://doi.org/10.1016/j.psychsport.2006.03.004

Gaudreau, P., Blondin, J.-P., & Lapierre, A.-M. (2002). Athletes' coping during a competition: Relationship of coping strategies with positive affect, negative affect, and performance–goal discrepancy. *Psychology of Sport and Exercise, 3*(2), 125–150. https://doi.org/10.1016/S1469-0292(01)00015-2

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65*(2-3), 263-297. https://doi.org/10.1016/S0010-0277(98)00002-X

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 225-277). Cambridge University Press. https://doi.org/10.1017/CBO9780511983689.008

Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin, 144*(2), 177–197. https://doi.org/10.1037/bul0000127

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology, 76*(1), 129–142. https://doi.org/10.1037/0022-3514.76.1.129

Gilbert, D. T., Giesler, R. B., & Morris, K. A. (1995). When comparisons arise. J*ournal of Personality and Social Psychology, 69*(2), 227-236 . https://psycnet.apa.org/doi/10.1037/0022-3514.69.2.227

Gilbert, P. (2014). The origins and nature of compassion focused therapy. B*ritish Journal of Clinical Psychology, 53*(1), 6–41. https://doi.org/10.1111/bjc.12043

Gill, D. L., & Deeter, T. E. (1988). Development of the sport orientation questionnaire. *Research Quarterly for Exercise and Sport, 59*(3), 191–202. https://doi.org/10.1080/02701367.1988.10605504

Goethals, G. R., & Darley, J. M. (1977). Social comparison processes: Theoretical and empirical perspectives. In J. M. Suls & R. L. Miller (Eds), *Social comparison theory: An attributional approach* (pp. 259–278). Washington, DC, USA: Hemisphere.

Gollwitzer, P. M. (1990). Action phases and mind-sets. In E. Higgins & R. M. Sorrention (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 53-92). Guilford Press.

Gollwitzer, P. M. (1996). The volitional benefits of planning. In P. M. Gollwitzer & J. A. Bargh (Eds.), *Linking cognition and motivation to behavior* (pp. 287–312). The Guilford Press.

Goodwin, G., Piazza, J., & Rozin, P. (2013). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*(1), 148-168. https://doi.org/10.1037/a0034726

Gotwals, J., & Wayment, H. A. (2002). Evaluation strategies, self esteem, and athletic performance. *Current Research in Social Psychology, 8*(6), 84-100.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Chapter Two - Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. In P. Devine & A. Plant (Eds.), *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. (2009). Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology, 96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Gröpel, P., Wegner, M., & Schüler, J. (2016). Achievement motive and sport participation. *Psychology of Sport and Exercise, 27*, 93–100. https://doi.org/10.1016/j.psychsport.2016.08.007

Gruder, C. L. (1971). Determinants of social comparison choices. *Journal of Experimental Social Psychology, 7*(5), 473–489. https://doi.org/10.1016/0022-1031(71)90010-2

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Haidt, J., Graham, J., & Joseph, C. (2009). Above and below left–right: Ideological narratives and moral foundations. *Psychological Inquiry, 20*(2–3), 110–119. https://doi.org/10.1080/10478400903028573

Haidt, J., & Joseph, C. (2006). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3, pp. 367-391). Oxford University Press.

Hakmiller, K. L. (1966). Threat as a determinant of downward comparison. *Journal of Experimental Social Psychology, 1*, 32–39. https://doi.org/10.1016/0022-1031(66)90063-1

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*(7169), 557-559. https://doi.org/10.1038/nature06288

Hare, R. M. (1981). *Moral thinking: Its level, method, and point*. Oxford University Press.

Hartley, A. G., Furr, R. M., Helzer, E. G., Jayawickreme, E., Velasquez, K. R., & Fleeson, W. (2016). Morality's centrality to liking, respecting, and understanding others. *Social Psychological and Personality Science, 7*(7), 648-657. https://doi.org/10.1177/1948550616655359

Harwood, C. G., & Swain, A. B. J. (1998). Antecedents of pre-competition achievement goals in elite junior tennis players. *Journal of Sports Sciences, 16*(4), 357–371. https://doi.org/10.1080/02640419808559364

Hauke, N., & Abele, A. E. (2020). Two faces of the self: Actor-self perspective and observer-self perspective are differentially related to agency versus communion. *Self and Identity, 19*(3), 346–368. https://doi.org/10.1080/15298868.2019.1584582

Head, M. L., & Bruchmann, K. (2018). The effects of health frame and target relevance in appearance social comparisons. *Journal of Applied Social Psychology, 49*(1), 27-35. https://psycnet.apa.org/doi/10.1111/jasp.12562

Heckhausen, H. (1977). Achievement motivation and its constructs: A cognitive model. *Motivation and Emotion, 1*, 283-329. https://doi.org/10.1007/BF00992538

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective likert scales?: The reference-group effect. *Journal of Personality and Social Psychology, 82*(6), 903–918. https://doi.org/10.1037/0022-3514.82.6.903

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life.* Thousand Oaks, CA, USA: Sage. https://dx.doi.org/10.4135/9781412984201

Hellstedt, J. C. (1990). Early adolescent perceptions of parental pressure in the sport environment. J*ournal of Sport Behaviour, 13*(3), 135-144.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*(12), 1280–1300. https://doi.org/10.1037/0003-066X.52.12.1280

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology*, *30*, 1-46. https://doi.org/10.1016/S0065-2601(08)60381-0

Higgins, E. T. (2012). Regulatory focus theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 483–504). SAGE Publications Ltd. https://doi.org/10.4135/9781446249215.n24

Hill, D. M., & Shaw, G. (2013). A qualitative examination of choking under pressure in team sport. *Psychology of Sport and Exercise, 14*(1), 103–110. https://doi.org/10.1016/j.psychsport.2012.07.008

Hofmann, W., Friese, M., Müller, J., & Strack, F. (2011). Zwei Seelen wohnen, ach, in meiner Brust. Psychologische und philosophische Erkenntnisse zum Konflikt zwischen Impuls und Selbstkontrolle. *Psychologische Rundschau*, *62*(3), 147-166. https://doi.org/10.1026/0033-3042/a000086

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science, 4*(2), 162–176. https://doi.org/10.1111/j.1745-6924.2009.01116.x

Horn, T. S., & Harris, A. (2002). Perceived competence in young athletes: Research findings and recommendations for coaches and parents. In F. L. Smoll & R. E. Smith (Eds.), *Children and Youth in Sport: A Biopsychosocial Perspective* (2nd ed.). Kendall-Hunt.

Hox, J. J. (2010). Multivariate multilevel regression models. Multilevel Analysis: Techniques and Applications (2nd ed.). New York, NY: Routledge. https://doi.org/10.4324/9781315650982-10

Hoyt, C. L. (2013). Inspirational or self-deflating: The role of self-efficacy in elite role model effectiveness. *Social Psychological and Personality Science, 4*(3), 290-298. https://doi.org/10.1177/1948550612455066

Hoyt, C. L., Burnette, J. L., & Innella, A. N. (2012). I can do that: The impact of implicit theories on leadership role model effectiveness. *Personality & Social Psychology Bulletin, 38*(2), 257–268. https://doi.org/10.1177/0146167211427922

Hoyt, C. L., & Simon, S. (2011). Female leaders injurious or inspiring role models for women? *Psychology of Women Quarterly, 35(*1), 143–157. https://doi.org/10.1177/0361684310385216

Huguet, P., Dumas, F., Marsh, H., Wheeler, L., Seaton, M., Nezlek, J., Suls, J., & Regner, I. (2009). Clarifying the role of social comparison in the big-fish-little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97*(1), 156-170. https://doi.org/10.1037/a0015558

Jones, M., Meijen, C., McCarthy, P. J., & Sheffield, D. (2009). A theory of challenge and threat states in athletes. *International Review of Sport and Exercise Psychology*, *2*(2), 161–180. https://doi.org/10.1080/17509840902829331

Jostmann, N. B., & Koole, S. L. (2009). When persistence is futile. Action orientation and the psychology of goal disengagement. In G. B. Moskowitz & H. Grant (Eds.), The psychology of goals (pp. 337–361). Guilford Press.

Kant, I. (1974/1929). *Kritik der praktischen Vernunft [Critique of practical reason]* (9th ed.). Meiner. (Original work published 1929)

Keltner, D., & Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition and Emotion, 17*(2), 297-314. https://doi.org/10.1080/02699930302297

Kenny, D. A., & Judd, C. M. (2019). The unappreciated heterogeneity of effect sizes: Implications for power, precision, planning of research, and replication. *Psychological Methods, 24*(5), 578-589. https://doi.org/10.1037/met0000209

King, V., Gerisch, B., & Schreiber, J. (2019). ". . . To really have everything completely perfect": On the psychodynamics of contemporary forms of body optimization. *Psychoanalytic Psychology, 37*(2), 148-157. https://doi.org/10.1037/pap0000287

Klatt, S., & Noël, B. (2020). Regulatory focus in sport revisited: Does the exact wording of instructions really matter? *Sport, Exercise, and Performance Psychology, 9*(4), 532–542. https://doi.org/10.1037/spy0000195

Klinger, E. (1975). Consequences of commitment to and disengagement from incentives. *Psychological Review, 82*(1), 1–25. https://doi.org/10.1037/h0076171

Knobloch-Westerwick, S., Robinson, M. J., Willis, L. E., & Luong, K. T. (2020). Beauty or Business Queen: How Young Women Select Media to Reinforce Possible Future Selves. *Communication Research, 47*(3), 323–345. https://doi.org/10.1177/0093650215626978

Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes. *Journal of Research in Personality, 46*(2), 184–194. https://doi.org/10.1016/j.jrp.2012.01.006

Kotabe, H. P., & Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science, 10*(5), 618–638. https://doi.org/10.1177/1745691615593382

Kreft, I. G., & De Leeuw, J. (1998). *Introducing multilevel modeling.* London, UK: Sage Publications. https://doi.org/10.4135/9781849209366

Kruglanski, A. W., Shah, J. Y., Fishbach, A., & Friedman, R. (2002). A theory of goal systems. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 331–378). Academic Press.

Kuhl, J. (2000). Chapter 5 - A Functional-Design Approach to Motivation and Self-Regulation: The Dynamics of Personality Systems Interactions. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 111–169). Academic Press. https://doi.org/10.1016/B978-012109890-2/50034-2

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science, 14*(5), 402-408. https://doi.org/10.1111/1467-9280.01454

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In *Flow and the foundations of positive psychology* (pp. 21–34). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_2

Lai, M. H. C. (2020). Bootstrap confidence intervals for multilevel standardized effect size. *Multivariate Behavioral Research*. Advance online publication. https://doi.org/10.1080/00273171.2020.1746902

Landy, J., Piazza, J. R., & Goodwin, G. P. (2016). When it's bad to be friendly and smart: The desirability of sociability and competence depends on morality. *Personality and Social Psychology Bulletin, 42*(9), 1272 - 1290. https://doi.org/10.1177/0146167216655984

Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion*. Springer. https://doi.org/10.1007/978-1-4899-0448-5

Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical Turk. *Behavior Research Methods, 47*(2), 519-528. https://doi.org/10.3758/s13428-014-0483-x

Liu, Q.-Q., Zhou, Z.-K., Yang, X.-J., Niu, G.-F., Tian, Y., & Fan, C.-Y. (2017). Upward social comparison on social network sites and depressive symptoms: A moderated mediation model of self-esteem and optimism. *Personality and Individual Differences, 113*, 223–228. https://doi.org/10.1016/j.paid.2017.03.037

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*, 705. https://doi.org/10.1037/0003-066X.57.9.705

Lockwood, P., Chasteen, A. L., & Wong, C. (2005). Age and regulatory focus determine preferences for health-related role models. *Psychology and Aging, 20*(3), 376–389. https://doi.org/10.1037/0882-7974.20.3.376

Lockwood, P., Crane, C., McShane, K., & Dolderman, D. (2005). The impact of positive and negative fitness exemplars on motivation. *Basic and Applied Social Psychology, 27*(1)*,* 1–13. https://doi.org/10.1207/s15324834basp2701_1

Lockwood, P., Jordan, C. H., & Kunda, Z. (2002a). Motivation by positive or negative role models: Regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology, 83*(4), 854–864. https://doi.org/10.1037/0022-3514.83.4.854

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology, 73*, 91–103. https://doi.org/10.1037/0022-3514.73.1.91

Lockwood, P., & Kunda, Z. (1999). Increasing the salience of one's best selves can undermine inspiration by outstanding role models. *Journal of Personality and Social Psychology, 76(2)*, 214. https://psycnet.apa.org/doi/10.1037/0022-3514.76.2.214

Lockwood, P., & Kunda, Z. (2000). Outstanding role models: Do they inspire or demoralize us? In A. Tesser, R. Felson, & J. Suls (Eds.), *Psychological perspectives on self and identity* (pp. 147–171). American Psychological Association. https://doi.org/10.1037/10357-006

Lockwood, P., Sadler, P., Fyman, K., & Tuck, S. (2004). To do or not to do: Using positive and negative role models to harness motivation. *Social Cognition, 22*, 422–450. https://doi.org/10.1521/soco.22.4.422.38297

Lockwood, P., Wong, C., McShane, K., & Dolderman, D. (2005). The impact of positive and negative fitness exemplars on motivation. *Basic and Applied Social Psychology, 27*, 1–13. https://doi.org/10.1207/s15324834basp2701_1

Luong, K. T., Knobloch-Westerwick, S., & Niewiesk, S. (2020). Superstars within reach: The role of perceived attainability and role congruity in media role models on women's social comparisons. *Communication Monographs, 87*(1), 4–24. https://doi.org/10.1080/03637751.2019.1622143

Lyubomirsky, S. (2010). *Hedonic Adaptation to Positive and Negative Experiences.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195375343.013.0011

Mahler, H. I., Kulik, J. A., Gerrard, M., & Gibbons, F. X. (2010). Effects of upward and downward social comparison information on the efficacy of an appearance-based sun protection intervention: A randomized, controlled experiment. *Journal of Behavioral Medicine, 33*, 496–507. https://doi.org/10.1007/s10865-010-9279-3

Major, B., Testa, M., & Blysma, W. H. (1991). Responses to upward and downward social comparisons: The impact of esteem-relevance and perceived control. In J. Suls & T. A. Wills (Eds), *Social comparison: Contemporary theory and research* (pp. 237–260). Erlbaum

Markman, K. D., & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. Personality and Social Psychology Review, 7, 244–267.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224-253. https://doi.org/10.1037/0033-295x.98.2.224

Marsh, H. W., Hey, J., Roche, L. A., & Perry, C. (1997). Structure of physical self-concept: Elite athletes and physical education students. *Journal of Educational Psychology, 89*(2), 369–380. https://doi.org/10.1037/0022-0663.89.2.369

Marsh, H. W., & Redmayne, R. S. (2016). A multidimensional physical self-concept and its relations to multiple components of physical fitness. J*ournal of Sport and Exercise Psychology, 16*(1), 43–55. https://doi.org/10.1123/jsep.16.1.43

McCullagh, P., & Weiss, M. R. (2001). Modeling: Considerations for motor skill performance and psychological responses. In R. N. Singer, H. A. Hausenblas, & C. Janelle (Eds.), *Handbook of Sport Psychology* (pp. 205–238). Wiley.

McGraw, A. P., Mellers, B. A., & Tetlock, P. E. (2005). Expectations and emotions of Olympic athletes. *Journal of Experimental Social Psychology*, *41*(4), 438–446. https://doi.org/10.1016/j.jesp.2004.09.001

Memmert, D., Plessner, H., & Maaßmann, J. (2009). Zur Erklärungskraft der „Regulatory Focus" Theorie im Sport. *Zeitschrift Für Sportpsychologie, 16*(3), 80–90. https://doi.org/10.1026/1612-5010.16.3.80

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. S*ocial and Personality Psychology Compass, 4*(5), 344–357. https://doi.org/10.1111/j.1751-9004.2010.00263.x

Midgley, C., Thai, S., Lockwood, P., Kovacheff, C., & Page-Gould, E. (2020). When every day is a high school reunion: Social media comparisons and self-esteem. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org/10.1037/pspi0000336

Milyavskaya, M., & Werner, K. (2018). Goal pursuit: Current state of affairs and directions for future research. *Canadian Psychology, 59*(2). https://doi.org/10.1037/cap0000147

Minson, J. A., & Monin, B. (2012). Do-gooder derogation: Disparaging morally motivated minorities to defuse anticipated reproach. *Social Psychological and Personality Science, 3*(2), 200-207. https://doi.org/10.1177/1948550611415695

Monin, B. (2007). Holier than me? Threatening social comparison in the moral domain. *Revue Internationale de Psychologie Sociale, 20*(1), 53–68.

Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology, 95*(1), 76-93. https://doi.org/10.1037/0022-3514.95.1.76

Morgenroth, T., Ryan, M. K., & Peters, K. (2015). The motivational theory of role modeling: How role models influence role aspirants' goals. *Review of General Psychology, 19*, 465–483. https://doi.org/10.1037/gpr0000059

Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality and Social Psychology, 16*(1), 148–156. https://doi.org/10.1037/h0029862

movisens GmbH. (2016). *movisensXS.(*Version 0.7.4162) [Mobile App]. movisens GmbH, Karlsruhe, Germany

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2016). The generalizability of survey experiments. *Journal of Experimental Political Science, 2*(2), 109-138. https://doi.org/10.1017/XPS.2015.19

Mussweiler, T. (2001). "Seek and ye shall find": Antecedents of assimilation and contrast in social comparison. *European Journal of Social Psychology, 31*, 499–509. https://doi.org/10.1002/ejsp.75

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110*, 472–489. https://doi.org/10.1037/0033-295X.110.3.472

Mussweiler, T., & Epstude, K. (2009). Relatively fast! Efficiency advantages of comparative thinking. *Journal of Experimental Psychology: General, 138*(1), 1–21. https://doi.org/10.1037/a0014374

Mussweiler, T., Rüter, K., & Epstude, K. (2004). The man who wasn't there: Subliminal social comparison standards influence self-evaluation. *Journal of Experimental Social Psychology, 40*(5), 689–696. https://doi.org/10.1016/j.jesp.2004.01.004

Myers, T. A., & Crowther, J. H. (2009). Social comparison as a predictor of body dissatisfaction: A meta-analytic review. *Journal of Abnormal Psychology, 118*(4), 683-698. https://doi.org/10.1037/a0016763

Neff, K. (2003). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity, 2*(2), 85–101. https://doi.org/10.1080/15298860309032

Nesi, J., & Prinstein, M. J. (2015). Using social media for social comparison and feedback-seeking: Gender and popularity moderate associations with depressive symptoms. *Journal of Abnormal Child Psychology, 43*(8), 1427–1438. https://doi.org/10.1007/s10802-015-0020-0

Nicuță, E. G., & Constantin, T. (2021). Take Nothing for Granted: Downward Social Comparison and Counterfactual Thinking Increase Adolescents' State Gratitude for the Little Things in Life. *Journal of Happiness Studies*. https://doi.org/10.1007/s10902-021-00382-5

Nielsen, K. S., & Hofmann, W. (2021). Motivating sustainability through morality: A daily diary study on the link between moral self-control and clothing consumption. *Journal of Environmental Psychology, 73*, 101551. https://doi.org/10.1016/j.jenvp.2021.101551

Ohmann, K., Mussweiler, T., & Kedia, G. (2016). Immediate relativity: EEG reveals early engagement of comparison in social information processing. *Journal of Experimental Psychology: General, 145*(11), 1512–1529. https://doi.org/10.1037/xge0000222

Park, Y., & Park, S. W. (2017). Goal orientations and social comparison: The role of different motivations in affiliation preferences. *Motivation and Emotion, 41*(5), 617–627. https://doi.org/10.1007/s11031-017-9634-6

Pila, E., Stamiris, A., Castonguay, A., & Sabiston, C. M. (2014). Body-related envy: A social comparison perspective in sport and exercise. *Journal of Sport and Exercise Psychology, 36*(1), 93–106. https://doi.org/10.1123/jsep.2013-0100

Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91-108). American Psychological Association. https://doi.org/10.1037/13091-005

Pizarro, D. A., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise. *Psychological Science, 14*(3), 267-272. https://doi.org/10.1111/1467-9280.03433

Polivy, J. (2001). The false hope syndrome: Unrealistic expectations of self-change. International Journal of Obesity, 25 Suppl 1, S80-4. https://doi.org/10.1038/sj.ijo.0801705

Polivy, J., Herman, C. P., Younger, J. C., & Erskine, B. (1979). Effects of a model on eating behavior: The induction of a restrained eating style. *Journal of Personality, 47*(1), 100–117. https://doi.org/10.1111/j.1467-6494.1979.tb00617.x

Project, J. (2018). Jamovi (Version 0.9)[Computer Software]. hhttps://www.jamovi.org

Pyszczynski, T., Greenberg, J., & LaPrelle, J. (1985). Social comparison after success and failure: Biased search for information consistent with a self-serving conclusion. *Journal of Experimental Social Psychology, 21*(2), 195–211. https://doi.org/10.1016/0022-1031(85)90015-0

R Core Team. (2017). *R: A language and environment for statistical computing.(*Version R Foundation for Statistical Computing, Vienna, Austria R-project.org/

Richter, M., Gendolla, G. H., & Wright, R. A. (2016). Three decades of research on motivational intensity theory: What we have learned about effort and what we still don't know. In A. J. Elliot (Ed.), *Advances in motivation science* (pp. 149–186). San Diego, CA, US: Elsevier Academic Press.

Rogers, T., & Feller, A. (2016). Discouraged by peer excellence: Exposure to exemplary peer performance causes quitting. *Psychological Science, 27*(3), 365–374. https://doi.org/10.1177/0956797615623770

Rom, S. C., & Conway, P. (2018). The strategic moral self: Self-presentation shapes moral dilemma judgments. *Journal of Experimental Social Psychology, 74*, 24-37. https://doi.org/10.1016/j.jesp.2017.08.003

Rosenthal, B., & McSweeney, F. K. (1979). Modeling influences on eating behavior. Addictive Behaviors, 4(3), 205–214. https://doi.org/10.1016/0306-4603(79)90029-7

Ruggieri, S., Ingoglia, S., Bonfanti, R. C., & Lo Coco, G. (2020). The role of online social comparison as a protective factor for psychological wellbeing: A longitudinal study during the covid-19 quarantine. *Personality and Individual Differences, 171*, 110486. https://doi.org/10.1016/j.paid.2020.110486

Ryska, T. A. (2001). Self-esteem among intercollegiate athletes: The role of achievement goals and competitive orientation. *Imagination, Cognition and Personality*, *21*(1), 67–80.

https://doi.org/10.2190/D6R7-KGFM-HALC-UPB7

Sassenberg, K., & Ditrich, L. (2019). Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science, 2*(2), 107-114. 251524591983878. https://doi.org/10.1177/2515245919838781

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science, 20*(4), 523-528. https://doi.org/10.1111/j.1467-9280.2009.02326.x

Scheier, M. F., & Carver, C. S. (1988). A model of behavioral self-regulation: Translating intention into action. *Advances in Experimental Social Psychology, 21*, 303-346. https://doi.org/10.1016/S0065-2601(08)60230-0

Schmalbach, B., Zenger, M., Spina, R., Steffens-Guerra, I., Kliem, S., Michaelides, M., & Hinz, A. (2017). Gain +1 or Avoid -1: Validation of the German Regulatory Focus Questionnaire (RFQ). *BMC Psychology, 5*(1), 40. https://doi.org/10.1186/s40359-017-0207-y

Schwarz, N., & Bless, H. (2007). Mental construal processes: The inclusion/exclusion model. In D. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology* (pp. 119-141). Philadelphia, PA: Psychology Press

Schwarzer, R. (1998). Optimism, goals, and threats: How to conceptualize self-regulatory processes in the adoption and maintenance of health behaviors. *Psychology and Health, 13*(4), 759–766.

Scollon, C. N., Prieto, C.-K., & Diener, E. (2009). Experience sampling: promises and pitfalls, strength and weaknesses. In E. Diener (Ed.), *Assessing well-being: The collected works of ed diener* (pp. 157–180). New York, NY, US: Springer. https://doi.org/10.1007/978-90-481-2354-4_8

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology, 84*(1), 60–79. https://doi.org/10.1037/0022-3514.84.1.60

Sedikides, C., & Strube, M. J. (1997). Self evaluation: To thine own self be good, to thine own

self be sure, to thine own self be true, and to thine own self be better. *Advances in Experimental Social Psychology, 29*, 209–269. https://doi.org/10.1016/S0065-2601(08)60018-0

Seekis, V., Bradley, G., & Duffy, A. (2020). Does a Facebook-enhanced mindful self-compassion intervention improve body image? An evaluation study. *Body Image, 34*, 259–269. https://doi.org/10.1016/j.bodyim.2020.07.006

Shakya, H., Christakis, N., & Fowler, J. (2015). Self-comparisons as motivators for healthy behavior. *Obesity, 23*(12), 2477-2484. https://doi.org/10.1002/oby.21201

Sheldon, J. P. (2003). Self-evaluation of competence by adult athletes: Its relation to skill level and personal importance. *The Sport Psychologist, 17*(4), 426–443. https://doi.org/10.1123/tsp.17.4.426

Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1-83). University of Chicago Press.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The 'big three' of morality (autonomy, community, divinity) and the 'big three' explanations of suffering. In A. M. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119-169). Taylor & Frances/Routledge.

Sidani, J. E., Shensa, A., Escobar-Viera, C. G., & Primack, B. A. (2020). Associations between comparison on social media and depressive symptoms: A study of young parents. *Journal of Child and Family Studies*. https://doi.org/10.1007/s10826-020-01805-2

Simonsohn, U. (2014, 15.07.). No-way interactions. *DataColada - Thinking about evidence and vice versa*. datacolada.org/17

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin, 20*(5), 580-591. https://doi.org/10.1177/0146167294205014

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88*(6), 895-917. https://doi.org/10.1037/0022-3514.88.6.895

Smith, R. H. (2000). Assimilative and contrastive emotional reactions to upward and downward social comparisons. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 173–200). Kluwer Academic Publishers. https://doi.org/10.1007/978-1-4615-4237-7_10

Snijders, T. A. (2005). Power and sample size in multilevel linear models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1570-1573). Chichester, UK: Wiley. https://doi.org/10.1002/0470013192.bsa492

Stoeber, J., Otto, K., Pescheck, E., Becker, C., & Stoll, O. (2007). Perfectionism and competitive anxiety in athletes: Differentiating striving for perfection and negative reactions to imperfection. *Personality and Individual Differences, 42*(6), 959–969. https://doi.org/10.1016/j.paid.2006.09.006

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*(3), 437–446. https://doi.org/10.1037/0022-3514.73.3.437

Strahan, E. J., Wilson, A. E., Cressman, K. E., & Buote, V. M. (2006). Comparing to perfection: How cultural norms for appearance affect social comparisons and self-image. *Body Image, 3*(3), 211–227. https://doi.org/10.1016/j.bodyim.2006.07.004

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition, 131*(1), 159–171. https://doi.org/10.1016/j.cognition.2013.12.005

Summerville, A., & Roese, N. J. (2008). Dare to compare: Fact-based versus simulation-based comparison in daily life. *Journal of Experimental Social Psychology, 44*(3), 664–671. https://doi.org/10.1016/j.jesp.2007.04.002

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. J*ournal of Personality, 72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193. https://psycnet.apa.org/doi/10.1037/0033-2909.103.2.193

Taylor, S. E., & Lobel, M. (1989). Social comparison activity under threat: Downward evaluation and upward contacts. *Psychological Review, 96*(4), 569–575. https://doi.org/10.1037/0033-295X.96.4.569

Taylor, S. E., Wayment, H. A., & Carillo, M. (1996). Social comparison, self-regulation, and motivation. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 3–27). Guilford Press.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology, 4*(4), 525–547. https://doi.org/10.1177/0959354394044004

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 181–227). Academic Press. https://doi.org/10.1016/S0065-2601(08)60227-0

Tesser, A. (1991). Emotion in social comparison and reflection processes. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 115–145). Lawrence Erlbaum Associates, Inc.

Tetlock, P. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences, 7*(7), 320–324. https://doi.org/10.1016/S1364-6613(03)00135-9

Tetlock, P., Kristel, O., Elson, S., Green, M., & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*(5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853

Theriault, J., Waytz, A., Heiphetz, L., & Young, L. (2017). Examining overlap in behavioral and neural representations of morals, facts, and preferences. *Journal of Experimental Psychology: General, 146*(11), 1586-1605. https://doi.org/10.1037/xge0000350

Thibodeau, P. H., & Flusberg, S. J. (2017). Lay theories of obesity: causes and consequences. *Adiposity: Epidemiology and treatment modalities* (pp. 23-36). Intech.

Thomassen, T. O., & Halvari, H. (1996). Achievement Motivation and Involvement in Sport Competitions. *Perceptual and Motor Skills*, 83(3_suppl), 1363–1374. https://doi.org/10.2466/pms.1996.83.3f.1363

Thornton, D. A., & Arrowood, A. J. (1966). Self-evaluation, self-enhancement, and the locus of social comparison. *Journal of Experimental Social Psychology, 1*, 40–48. https://doi.org/10.1016/0022-1031(66)90064-3

Thürmer, J. L., Scheier, M. F., & Carver, C. S. (2019). On the mechanics of goal striving: Experimental evidence of coasting and shifting. *Motivation Science, 6*(3), 266-274. https://doi.org/10.1037/mot0000157

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*(12), 455-460. https://doi.org/10.1016/j.tics.2008.10.001

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Van Lange, P. A. M., & Sedikides, C. (1998). Being more honest but not necessarily more intelligent than others: Generality and explanations for the Muhammad Ali effect. *European Journal of Social Psychology, 28*(4), 675-680. https://doi.org/10.1002/(SICI)1099-0992(199807/08)28:4<675::AID-EJSP883>3.0.CO;2-5

Vogel, T., & Genschow, O. (2013). When Do Chronic Differences in Self-Regulation Count?: Regulatory Focus Effects in Easy and Difficult Soccer Tasks. *Journal of Sport and Exercise Psychology, 35*(2), 216–220. https://doi.org/10.1123/jsep.35.2.216

Walton, C. C., Baranoff, J., Gilbert, P., & Kirby, J. (2020). Self-compassion, social rank, and psychological distress in athletes of varying competitive levels. *Psychology of Sport and Exercise, 50*, 101733. https://doi.org/10.1016/j.psychsport.2020.101733

Want, S. C. (2009). Meta-analytic moderators of experimental exposure to media portrayals of women on female appearance satisfaction: Social comparisons as automatic processes. *Body Image, 6*(4), 257–269. https://doi.org/10.1016/j.bodyim.2009.07.008

Weiss, M. A. (1995). Children in sport: An educational model. In S. M. Murphy (Ed.), *Sport psychology interventions* (pp. 39–69). Human Kinetics Publishers.

Wenhold, F., Elbe, A.-M., & Beckmann, J. (2008). AMS-Sport Langversion. Allgemeiner Fragebogen zum Leistungsmotiv im Sport. www.bisp.de

Wenhold, F., Meier, C., Elbe, A.-M., & Beckmann, J. (2008). Informationen zum Fragebogen AMS-Sport auf dem Internetportal Sportpsychologie des BISp. www.bisp.de

Wergin, V. V., Zimanyi, Z., Mesagno, C., & Beckmann, J. (2018). When suddenly nothing works anymore within a team – causes of collective sport team collapse. *Frontiers in Psychology, 9,* 2115. https://doi.org/10.3389/fpsyg.2018.02115

Wetherall, K., Robb, K. A., & O'Connor, R. C. (2019). An Examination of Social Comparison and Suicide Ideation Through the Lens of the Integrated Motivational-Volitional Model of Suicidal Behavior. *Suicide and Life-Threatening Behavior, 49*(1), 167–182. https://doi.org/10.1111/sltb.12434

Wheeler, L. (1966). Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology, 1*, 27–31. https://doi.org/10.1016/0022-1031(66)90062-X

Wheeler, L. & Miyake, K. (1992). Social comparison in everyday life. *Journal of Personality and Social Psychology, 62*, 760-773. https://doi.org/10.1037/0022-3514.62.5.760

White, J. B., Langer, E. J., Yariv, L., & Welch, J. C. I. (2006). Frequent social comparisons and destructive emotions and behaviors: The dark side of social comparisons. *Journal of Adult Development, 13*(1), 36–44. https://doi.org/10.1007/s10804-006-9005-0

Williams, L. (1994). Goal orientations and athletes' preferences for competence information sources. *Journal of Sport and Exercise Psychology, 16*(4), 416–430. https://doi.org/10.1123/jsep.16.4.416

Williams, L. A., & DeSteno, D. (2008). Pride and perseverance: The motivational role of pride. *Journal of Personality and Social Psychology, 94*(6), 1007–1017. https://doi.org/10.1037/0022-3514.94.6.1007

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592-598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin, 90*(2), 245–271. https://doi.org/10.1037/0033-2909.90.2.245

Wills, T. A. (1991). Similarity and self-esteem in downward comparison. In J. Suls & T. A. Wills (Eds.), S*ocial comparison: Contemporary theory and research* (pp. 51–78). Erlbaum

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin, 24*(12), 1251-1263. https://doi.org/10.1177/01461672982412001

Wood, J. V. (1996). What is social comparison and how should we study it? *Personality and Social Psychology Bulletin, 22*, 520–537. https://doi.org/10.1177/0146167296225009

Wright, R. A., & Brehm, J. W. (1989). Energization and goal attractiveness. In L. A. Pervin (Ed.), *Goal concepts in personality and social psychology* (pp. 169–210). Erlbaum

Wrosch, C., Scheier, M. F., Carver, C. S., & Schulz, R. (2003). The importance of goal disengagement in adaptive self-regulation: When giving up is beneficial. *Self and Identity, 2*, 1–20. https://doi.org/10.1080/15298860309021

Xiang, P., Liu, J., Li, W., & Guan, J. (2020). Students' ability beliefs about running in physical education: Qualitative findings from longitudinal data. *European Physical Education Review, 27(2)*, 348-365. https://doi.org/10.1177/1356336X20949902

Ybarra, O., Park, H., Stanik, C., & Lee, D. (2012). Self-judgment and reputation monitoring as a function of the fundamental dimensions, temporal perspective, and culture. *European Journal of Social Psychology, 42*(2). https://doi.org/10.1002/ejsp.854