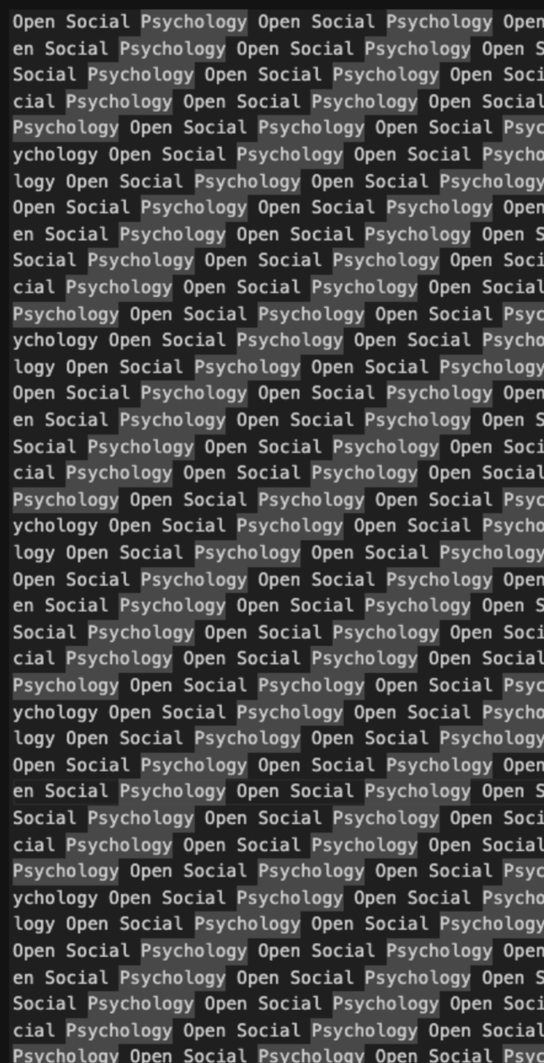
Open Social Psychology



Edited by Rima-Maria Rahal

Open Social Psychology

Rima-Maria Rahal

2025-08-27

Table of contents

Preface

Social psychology is built on a strong set of classical research paradigms and findings, featured in many of the textbooks, syllabi, online courses and teaching guides that aspiring psychologists study with and established psychologists use as teaching resources. However, the common body of knowledge that social psychology relies on is undergoing change. Modern research methods and changing attitudes towards permissible research practices bring about social psychological research that looks different today than it used. This book is dedicated to tracing some of these changes, and to offering a version of record of the changing perceptions and interpretations of classic social psychology in the light of it's contemporary counterpart. As such, this study book is a snapshot of how we see social psychology today.

Because it tends to be difficult to keep teaching and study materials up to date with emerging trends and debates, we see this study book as an addition to traditional educational resources in social psychology. It is published as an Open Educational Resource to aid the accessibility of this knowledge for all, and to be adapted to teachers' and learners' needs as they dive into what social psychology has to offer.

How this Book Came to Be

written by Flávio Azevedo and Rima-Maria Rahal

Social psychology is devoted to studying how individuals behave, think and feel within their social contexts. The field is therefore, by its very nature, set up for collaborative work. Leveraging the social context in which knowledge is generated is built in to the assumptions and interests that social psychology pursues. This fundamental attitude towards social embeddedness of knowledge is mirrored in the process by which this study book came to be.

It started by bringing together the work of students at Heidelberg University during the winter term of 2023. In the scope of classwork, they engaged with classical findings of social psychology, and discussed recent attempts to reengage with these classics. These works are the basis of the current book.

Researchers working on (areas related to) social psychology then revised these chapters. Through engaging the communities at the Big Team Science Conference 2024 (BTScon), the 2025 annual meeting of the Society for the Improvement of Psychological Science (SIPS) and the Framework for Open and Reproducible Research Training (FORRT), we found collaborators willing to contribute their knowledge and expertise to turning chapter drafts into an approachable and fact checked resource.

FORRT

An *International Bridge-Building Award* of the Society for Personality and Social Psychology (SPSP) to members of FORRT in 2026 facilitated three online hackathons with researchers from the Global South, as well as marginalized and liminal scholars.

The creation of this book was also supported by the German Reproducibility Network (GRN).

By co-creating educational content with diverse participants rather than relying solely on traditional authority figures, the process of writing this book explicitly built in diverse perspectives and lived experiences of groups who may otherwise not have access to such contribution opportunities. This process promoted cross-cultural scholarly exchange on replication and reproducibility in social psychology, and made this educational resource more inclusive by reflecting diverse perspectives.

In sum, this volume offers diverse perspectives on a shared target topic: Changing perceptions of classical social psychological research.

How to Use this Book

written by Melissa Engelbart and Rima-Maria Rahal

This book contains several types of resources: narrative text, definitions and questions for reflection, as well as references.

In fifteen chapters, we provide narrative summaries about classical research in social psychology and its modern follow-up. Often, this means we include new attempts to show the same finding (replication attempts) or meta-analytical work that brings together a lot of evidence from different sources regarding a certain hypothesis. Each chapter contains an overview of the classic study, a summary of important work thereafter, as well as a discussion of the evidence, experiments or analyses conducted. We then attempt to draw conclusions about the tested hypotheses.

Because this volume is targeted at students, we provide definitions of key terms, preceded by #definition and displayed like this:

> #definition Replication
>
> An attempt to find the same result as a previous study in a new data set.
>
> #definition Meta-Analysis
>
> An analysis that brings together evidence from several individual studies or experiments to estimate an overall effect across the available evidence.

We have aimed at providing a critical but neutral perspective to the classical and modern studies of social psychology discussed in the texts of this volume. To help you develop your own perspective and a well-reflected attitude towards this work, you will find guiding questions and suggestions that might prompt you to think more deeply about what you read throughout the book. The guiding questions cover topics such as the research and publication process itself and it's influence on research, the interpretation of data in general, as well as the experimental operationalization of theoretical questions. Moreover, to help you consider potential applications of the findings and theories discussed, these questions sometimes ask you to think of examples or consequences in real life.

You'll recognize these prompts by the preceding #yourturn. Here is an example of what these questions look like:

> #yourturn
>
> Do you think you might find such questions for reflection useful?

Finally, we have enabled the option to collaboratively annotate this work using hypothesis (note that this is how links are formatted in this book) in the online version. Your annotations will be visible to others, and others will be able to see yours, so that we can build a better learning experience using this book together.

To read up on the original research we cite in this book, such as from Vazire (2018), you can hover over or click on the references provided.

Feel free to make use of the resources in this book as you see fit. Our hope is that they will support you in building a well-reflected opinion about the existing body of knowledge in social psychology.

Introduction

written by Rima-Maria Rahal

The Role of Change for Scientific Discovery

Much of science capitalizes on change. It is the engine that drives progress and the expansion of knowledge (see Kuhn 1962; Popper 1959). Embracing change means taking established theories and challenging them to explore new directions. Changing perspectives, questioning the status quo, refining existing concepts, and adapting to new evidence provide the stuff that makes breakthroughs or new insights. In essence, change in science represents taking steps forward, toward greater insight and reality checks for the challenges we face. In other words, to push the boundaries of what we know, we must make change.

> #yourturn
>
> What instance of change regarding science have you recently heard about? Consider reports of breakthroughs you might have seen in the news or stories you saw on social media.

In the past decade, Open Science has made change, by transforming research practices to promote transparency, reproducibility, and collaboration in scientific endeavors. By fostering a culture of openness and collaboration, Open Science has brought about a paradigm shift in research methodologies, paving the way for more robust and reliable scientific discoveries (Munafò et al. 2017; Vazire, Schiavone, and Bottesini 2022). It is certainly no small feat to fundamentally reform how research is done, and yet we have seen significant change towards Open practices (Kidwell et al. 2016; Chambers 2019; Christensen et al. 2020).

> #definition Open Science
>
> An overhead term for a number of practices to make research more transparent, such as making the data a research is project is based on available to others.

Challenges of Making Change

Change can be a challenge because it disrupts established norms, habits, and power structures. This often means that individuals and groups might be hesitant to embrace change. Open Science as a reform to refocus on good research practice had to work with this difficulty of making change, where new methods, theories, or technologies often encounter skepticism and opposition from the scientific community. Open Science promotes transparency, data sharing, and collaborative research, which can expose flaws underlying previously held beliefs or reveal alternative interpretations. This shift can create debates about long-held ideas and established practices, which are scrutinized and potentially overturned. Established researchers may be reluctant to abandon familiar paradigms, and institutions may resist reallocating resources or altering well-known processes. Sometimes, inertia of traditional practices and fear of uncertainty can slow the adoption of innovative approaches, despite their potential to advance knowledge and solve pressing problems.

> #yourturn
>
> Consider a big change you have experienced. Was it easy to adapt to this change?

However, a questioning attitude and focus on methodological rigor and good practice also enhance the robustness and reliability of scientific conclusions by fostering an environment where continuous re-evaluation is encouraged. Thus, Open Science exemplifies how embracing change can lead to a more dynamic and resilient understanding of the world, even as it unsettles the familiar foundations of scientific consensus.

Change often implies the potential for a changed perception of what used to be, particularly in comparison to what is now. This is also the case in the scope of changes assosciated with Open Science. In particular, what were once considered unassailable facts can become contested or uncertain as new methodologies, data, and technologies challenge established knowledge. This is where our focus lies in this book: reporting on classical studies in social psychology and the change in how they are seen now, following a wave of additional research (often with an Open Science flavor).

> #yourturn
>
> "I was today years old when I found out …" What was the last long-held belief you had to give up?

In this spirit, when reading about the changes in perspective about classics in social psychology, there are two things to embrace:

On the one hand, revisiting classic social psychology studies is a demonstration of the profound impact they had on the field. Were they less important and less impactful, these studies would not draw continued debate, research interest and investment of resources. Therefore, reading classic studies can give readers

a sense of what matters to social psychological research, from hot topics to hot paradigms and research methods.

On the other hand, following the course of the academic debate about these claims, insights and phenomena allows us to hone our skills in accumulating insights and adjusting our perception of the currently held beliefs in this area of research. Put differently, tracing efforts to replicate, to conduct meta-analyses or to establish boundary conditions to the findings postulated in a certain study mostly reflects well-intentioned interest in assessing the validity of the claims of the original study, attempting to produce clarity about our collective knowledge about the phenomenon of interest. Reassessing classical studies might require change in opinions, calibration and reflection, but it can surely spark renewed trust in research and in its ability to refine and build our joint knowledge.

Chapter 1

Pygmalion Effect

written by Maja Düsenberg (original draft), and Jana Berkessel (revision)

1.1 The Classic

The Pygmalion Effect is a social psychological phenomenon that highlights how expectations can influence performance. It was first demonstrated by Robert Rosenthal and Lenore Jacobson in their seminal study, *Pygmalion in the Classroom* (Rosenthal and Jacobson 1968). The study investigated how teachers' expectations about their students' potential could shape the students' academic performance.

> #definition Pygmalion Effect
>
> The phenomenon in which higher expectations from others lead to improved performance.

In their experiment, Rosenthal and Jacobson (1968) told teachers that certain students were likely to be "growth spurters" who were expected to achieve significant academic improvement over the school year based on a fabricated test. The students were randomly selected and had no actual differences in ability compared to their peers.

The results were striking. The so-called "spurters" showed significantly higher gains in IQ scores from the pre-test to the post-test compared to their control peers. These changes could not be explained by retesting effects, familiarity with the test, or natural cognitive development due to aging. Instead, the findings highlighted the powerful role of teacher expectations in shaping student outcomes.

The study also explored moderating factors. Younger children demonstrated greater improvements, potentially due to their higher malleability to external

influences. Gender differences were observed, with girls showing greater increases in reasoning IQ and boys improving more in verbal IQ, aligning with their respective pretest strengths. Additionally, while not statistically significant, minority students appeared to benefit more from positive expectations, with "more Mexican-looking" boys (e.g., darker skin tones) showing particularly pronounced IQ gains. These results suggest that preconceived notions based on race and ethnicity may interact with expectation effects.

> #yourturn
>
> In which other situations could the Pygmalion effect play a role? Think about situations where your assumptions or expectations about others may influence their behavior—positively or negatively.

The implications of the Pygmalion effect extend far beyond the classroom. In organizational settings, for instance, research has shown that managers' high expectations for their employees can lead to improved performance through changes in behavior and increased self-efficacy (Eden 1990). Similar dynamics have been observed in therapeutic relationships, where therapists' beliefs about their clients' potential for progress influence treatment outcomes (Jenner 1990), and in healthcare, where nurses' confidence in their patients' recovery can affect health results (Learman et al. 1990). These examples illustrate how expectations have the power to shape behavior and performance in diverse domains.

> #yourturn
>
> Which other social psychological constructs are related to the Pygmalion effect?

The Pygmalion effect is closely related to two other psychological concepts: the self-fulfilling prophecy and self-efficacy. A self-fulfilling prophecy occurs when an initially false belief or expectation leads to behaviors that ultimately make the false belief come true. This concept aligns with the Pygmalion effect, as individuals may unconsciously alter their actions to align with the expectations placed upon them. Additionally, self-efficacy, or one's belief in their ability to succeed in specific situations, plays a key role in mediating the impact of expectations. When high expectations are communicated, they can enhance an individual's self-efficacy, reinforcing their motivation and performance. These interconnected mechanisms help explain how expectations shape outcomes across various domains.

1.2 The Aftermath

Thorndike (1968) and Snow (1969) offered early critiques of Rosenthal and Jacobson's (1968) Pygmalion study, challenging its methodology, data analysis, and conclusions. Thorndike (1968) focused on issues with data quality, pointing out inconsistencies such as the implausibly low IQ scores of some participants, which he described as rendering the testing meaningless. He argued that the

effects of the intervention were limited primarily to a small group of first- and second-grade students, raising concerns about the generalizability of the findings. Thorndike (1968) concluded that any observed effects might have been coincidental rather than genuinely linked to the intervention.

#yourturn

The Pygmalion effect often involves subconscious biases. How do you think societal stereotypes (e.g., gender, race) might influence the expectations we hold for others? Can you think of ways to address or mitigate these biases?

Snow (1969), similarly skeptical, critiqued the study's complex experimental design, highlighting incomplete data and methodological flaws, such as inadequate norms for the youngest children and those from lower socioeconomic backgrounds. He noted that 20% of the participants were not retested, an omission unaddressed in the analysis. Snow (1969) also questioned the mechanism of teacher influence, pointing out that teachers reportedly could not recall which students were identified as "bloomers," undermining the study's foundational premise. Snow (1969) concluded that the study's premature dissemination in popular media had harmed teachers, parents, and students by raising unrealistic expectations without robust evidence to support them.

#definition Special Issue

A collection of articles on a specific topic, typically published together in a single issue of an academic journal. Special issues are often edited by guest editors and aim to provide a comprehensive exploration of the chosen theme or field of study.

In 2018, the journal *Educational Research and Evaluation* published a special issue on the Pygmalion effect, just in time for its 50th birthday. In the Editorial, they summarize that despite warranted criticism of the early studies, research conducted over the past five decades has refined our understanding of the Pygmalion effect. Specifically, empirical studies have shown that teachers generally show a degree of accuracy in their expectations (Jussim and Harber 2005) but tend to favour students from affluent backgrounds over those from less privileged ones, while often holding lower expectations for special needs students (De Boer, Bosker, and Van der Werf 2010; Cameron and Cook 2013). Evidence on expectations related to student ethnicity and gender is more inconsistent, with some studies finding biases—such as lower expectations for ethnic minority students, boys in reading, and girls in mathematics—while others do not. Teacher expectations influence teaching behaviours, such as offering greater opportunities to learn, asking richer questions, and providing more targeted feedback for students with higher expectations (J. E. Brophy and Good 1970; C. M. Rubie-Davies 2007). These expectations can function as self-fulfilling prophecies, impacting student outcomes like performance, intelligence, and motivation. However, the magnitude of these effects varies significantly across studies (e.g., effect sizes ranging from d = .11, Raudenbush (1984); to d = .70, Rosenthal

and Rubin (1978)). Notably, students who are low achievers, from low-income families, or belong to ethnic minority groups appear more vulnerable to these effects, and some teachers are more likely than others to amplify these disparities (Madon, Jussim, and Eccles 1997; Christine M. Rubie-Davies et al. 2015).

> #definition Editorial
>
> An introductory article written by the editors of a special issue in an academic journal. It outlines the purpose, scope, and significance of the special issue, provides an overview of the included articles, and often highlights key themes, trends, or gaps in the research field.

Finally, the Editors underscore the need to view teacher expectations ecologically, considering the individuality of teachers and students, as well as the broader contexts of classrooms, schools, families, and communities. Teacher expectation effects are not universal; they vary by teacher practices, student vulnerability, and contextual factors like classroom composition. They also emphasize the importance of integrating teacher expectation findings into teacher education. Teaching future educators to avoid the negative effects of low expectations and to provide appropriately challenging learning opportunities could foster greater equity in student outcomes.

> #yourturn
>
> Can the Pygmalion effect apply to self-expectations? How might your own beliefs about your abilities influence your performance in a given task or goal?

1.3 Conclusion

The Pygmalion Effect is a social psychological phenomenon in which higher expectations from others lead to improved performance. This effect was first demonstrated by Robert Rosenthal and Lenore Jacobson in their seminal 1968 study, *Pygmalion in the Classroom* (Rosenthal and Jacobson 1968), which showed that teachers' beliefs about students' potential could significantly influence academic outcomes. While the original study laid the groundwork for understanding this phenomenon, the decades of subsequent research have added nuance to our understanding. Teacher expectations can indeed enhance or hinder students' academic achievements, but these effects are not uniform; they depend on various factors, including teacher practices, student background, and the context within which they operate.

Chapter 2

Ego Depletion

written by Hannah Baumgart (original draft) and Rima-Maria Rahal (revision)

2.1 The Classic

Ego depletion is a social psychological concept that describes the depletion of individuals' self-regulatory resources. Baumeister et al. (1998) were the first to demonstrate ego depletion effects in four different experimental settings: After having to engage in an act of self-control (compared to a control task that does not require self-control), willpower is used up and could not be deployed as effectively in a subsequent task.

> #definition Ego Depletion
>
> A concept that describes willpower as a limited resource that can be used up (depleted).

In Experiment 1, the focus was on the act of resisting a temptation, which requires self-control. Participants were randomly assigned to different food conditions, by which the independent variables were manipulated: Chocolate chip cookies and chocolate, radishes or no food at all (control group). Participants in the radish control condition were instructed to resist the tempting chocolates and instead eat several the radishes that were laid out next to the chocolate. In the chocolate condition, participants were asked to eat several cookies or chocolates, which were laid out next to the radishes – a task that was not supposed to require much self-control. The actual intention behind the experiment, to demonstrate ego depletion, was disguised with a cover story to make sure participants would not get suspicious. They were told the experiment was about taste perception.

> #yourturn

> Which other tasks in your daily life require more or less willpower?

In the no-food control condition, participants were not asked to taste any food, but worked on the rest of the experiment.

After the participants had completed the willpower task resisting the temptation of the foods presented to them, they had to complete questionnaires on mood and restraint. Then they had to work on "solving" a problem-solving task, which was actually unsolvable. Here, the time spent on trying to solve the problem before giving up was the dependent variable.

The results showed significant differences between the three conditions, with participants in the radish condition stopping earlier than those in the chocolate or no-food condition. In conclusion, it was suggested that craving chocolate but choosing to eat radishes depleted an internal resource, leaving individuals less able to persist while trying to solve the puzzles afterwards.

> #yourturn
>
> If willpower can be depleted, how can it be "refilled" or built up again?

2.2 The Aftermath

Since this study, several hundred follow-up studies, including several multi-lab studies that aimed to replicate the overall finding (Hagger et al. 2010; Vohs et al. 2021) and several meta-analyses (Hagger et al. 2010; Carter and McCullough 2014; Dang 2017; Blázquez, Botella, and Suero 2017) have been carried out.

> #definition Multi-Lab Study
>
> A research project in which researchers working at several different locations (laboratories) implement the same experimental design and then analyse the data together.

These studies yielded mixed results, with some concluding that it was highly unlikely that the ego depletion phenomenon does not exist (e.g., Hagger et al. 2010), while others failed to establish the effect despite relying on data from more than 2000 participants (e.g., Hagger et al. 2010). Publication bias has been argued to be high in the literature on ego depletion (Inzlicht, Gervais, and Berkman 2015), casting doubt on the effect.

> #definition Publication Bias
>
> A tendency for research in line with established theories or showing significant results to be more easily publishable than deviating research.

Continued research interest on ego depletion has brought forward varying hypotheses regarding circumstances under which the effect might be demonstrable

and robust. The meta-analysis on ego depletion conducted by Dang (2017) investigated only studies with sufficient initial effort exerted in the depleting, which was hypothesized to lead to the ego depletion effect. The study ensured that the depleting task required the use of self-control and excluded manipulations that were less clearly related to self-control, such as those based on social exclusion. Eight commonly used depletion tasks were assessed in the meta-analysis: attention essay, attention video, crossing out letters, emotion video, food trial, Stroop, thought suppression, and working memory.

> #yourturn
>
> Can you imagine what participants had to do in these tasks? Think about a version of each task that would drain self-control and one that would be less exhausting.

The results showed that two of these exhausting tasks, attention video and working memory, were not associated with significant changes in subsequent self-control. Emotion videos, on the other hand, appeared to be the most effective task and reduced subsequent self-control.

The overall analysis revealed a small to medium effect size for the ego depletion effect. Correcting for publication bias, this effect was not statistically significant when using the full sample of studies identified. However, a separate analysis for reliable depletion tasks, such as attention essay, emotion video and Stroop, showed the significant effect remained when attempting to correct for publication bias. This meta-analysis suggests that in special tasks, ego depletion might occur, but that it is difficult to generalize to other circumstances.

However, even in these special tasks, there is often no direct measure of the initial depletion of willpower involved: manipulation checks on whether willpower has been used up offer only an indirect measurement (Friese et al. 2018).

> #yourturn
>
> How could you objectiveley measure the amount of willpower available or drained?

2.3 Conclusion

The literature suggests a differentiated view on the potentially finite nature of willpower is necessary (for a detailed overview, read more in Friese et al. 2018). In the context of social psychological theories, the ego depletion effect can be seen as an important example of contradictory findings in research, where publication bias may play a role. Although several hundred studies on ego depletion have been published, we cannot be sure whether ego depletion exists or not.

> #yourturn

Do you think ego depletion exists?

The debate about ego depletion shows that individual findings should be re-assessed in several empirical demonstrations, including replication attempts that can provide a more realistic picture of the effect or construct. In this case, the original ego depletion effect may have been initially inflated due to publication bias. Following closer examination, it is less certain whether this effect indeed exists. The example of the ego depletion literature also shows the importance of examining the evidence closely, under the microscope, in order to ensure that it meets the quality criteria that are essential for assessing cumulative evidence of the overall effect.

Chapter 3

Social Facilitation

written by Dearbhaile Vaughan (original draft), Kate Grady (original draft), Cillian McHugh (revision), and Siobhán M. Griffin (revision)

3.1 The Classic

Social Facilitation is a theory that posits that one will perform better on a task when it is completed in the presence of others. In 1898, Norman Triplett demonstrated that when people complete a task in competition with another person, they perform better on the task compared to completing the task alone (Triplett 1898). This seminal experiment sparked a rich literature on the concept of "social facilitation," a term which was coined some 20 years later by Allport (1954) – the idea that the mere presence of others can lead to improvements in performance (Aronson, Wilson, and Akert 2005; Bond and Titus 1983).

> #definition Definition of Social Facilitation
>
> This theory proposes that the mere presence of others will positively affect performance on a task.

Triplett's seminal study (1898) was the result of perceived trends he observed in cyclists. Triplett noticed that in both paced and competitive settings, cyclists tended to cycle faster when accompanied by others. He had a multitude of theories as to why this was, including both physical and psychological hypotheses. One theory was that of the 'Encouragement Theory', where the presence of a friend would cheer on and "keep the [participant's] spirits up." Other theories included: 'Shelter Theory' – the lead cyclist creates shelter from the wind making it easier for those behind to cycle; 'Suction Theory' – a vacuum is created by "suction exertion" from the cyclist in front; and 'Theory of Hypnotic Suggestions' – that a hypnosis effect is created by the wheels of the bicycle in front, and this leads to better performance.

To test his theory, Triplett designed a lab-based study to examine if the presence of a competitor stimulates competition arousal, which he called "dynamogenic factors."

> #definition Definition of Dynamogenesis
>
> An increase in the mental or motor activity of an already functioning bodily system that accompanies any added sensory stimulation (Merriam-Webster).

For the experiment, two fishing reels were attached to a table to create a type of pulley system that moved a flag around a four-metre course. Children were invited to participate in this study. After a practice period to allow children to become accustomed to the machine, they completed six trials alternating between performing alone and performing in competition with another child. There were rest periods in between each trial to avoid the effects of fatigue. Performance was defined as the time taken to complete one trial (four laps of the course) as measured by a stopwatch. The results showed that children performed better (i.e., completed the laps faster) during the competition/together trials compared to the alone trials. However, some variation was noted where some children, described as "overstimulated," performed slower on the together trials.

> #yourturn
>
> Can you think of how the factors such as (i) age variability, (ii) potential differences in practice times, (iii) lack of clarity around the rest periods, and (iv) reporting on data from a subsample of 40 participants (out of 225) may have potentially affected the observed findings?

Triplett's findings and theory posited that competition stimulates performance (competitive coaction), but subsequent researchers focused on a broader application of this idea - that the mere presence of another person would improve performance, competitive or not.

3.2 The Aftermath

Subsequent research focused on social facilitation across a number of different social pressure contexts, including having an observer or audience present, having an evaluative observer or audience, a non-competing co-actor, and similar to Triplett's study - in the presence of a competing co-actor (Dashiell 1930). Some research has highlighted the importance of task complexity. For instance, Zajonc (1965) examined social facilitation in a sample of cockroaches, showing that social presence enhances performance on simple tasks but hinders performance on more complex tasks (completing a runway vs completing a maze). Based on behaviour theory (Hull 1943; Spence 1956), Zajonc postulated that "generalized drive" is what motivates habits. According to Zajonc's theory, having

other people around increases generalized drive, which makes it easier for habitual dominant responses to occur. While for more complex tasks the dominant response may not be the correct response (Bond and Titus 1983; Zajonc 1965)). However, a replication of Zajonc's study did not fully replicate this effect; in simple and complex tasks, the cockroaches performed more slowly when other cockroaches were present (Halfmann, Bredehöft, and Häusser 2020).

#definition Definition of Generalized Drive

The presence of others leads to an increase in generalized drive, thus facilitating habitualised dominant responses.

Although Zajonc (1965) believed that the mere presence of others is the necessary ingredient in producing social facilitation effects, other researchers disagreed. Cottrell et al. (1972; 1968) argued that social facilitation occurs when a third party is perceived to be observing the performance, but that mere presence (without observation) was not sufficient to produce social facilitation effects. It is the expectation of evaluation that increases drive, and thus influences performance (Bond and Titus 1983; Cottrell 1972; Weiss and Miller 1971).

There are a number of theoretical explanations to explain how, why, and when social facilitation effects occur (for reviews see Bond and Titus 1983; Seitchik, Brown, and Harkins 2017). Some key theories include:

- *Distraction-conflict theory:* the idea that the presence of others is distracting and takes up attention resources which may lead to cognitive overload, reducing attention on the task (Baron 1986; Sanders, Baron, and Moore 1978). This may result in dominant responses facilitating performance when the task is simple and requires attention to a small number of cues, but when the task is more complex or demands attention to a larger number of cues performance may be hindered.
- Muller and Butera's (2007) *Integrated distraction-conflict theory and Social comparison theory* (Festinger 1954); e.g., that people compare their own skills to other people's skills), and proposed that when in a co-action setting people can experience self-evaluation threat which may increase their attentional focus, in particular when a co-actor is seen to be superior, increasing drive and thus performance.

#definition Definition of Distraction-Conflict Theory

This theory states that attentional conflict, a type of response conflict regarding what attentional response one should make, can arise when the social presence of others (co-actors or an audience) is distracting, at least when the task is attention demanding. The actor may then be at risk of cognitive overload as a result of this conflict, which would ultimately lead to a limitation in their ability to focus on the task.

#definition Definition of Social Comparison Theory

According to the social comparison theory, people are motivated to assess their own beliefs and skills by comparing them to external images. These images can be comparisons to other people or a reference to physical reality. Individuals have a tendency to view images portrayed by others as accessible and realistic and subsequently make comparisons between themselves, other people, and these idealized images.

3.2.1 Practical Implications Arising from Triplett's Original Study

Research on social facilitation effects has highlighted its practical implications in real-world settings. For example, Anderson-Hanley, Arciero, and Snyder (2011) demonstrated that adults riding on stationary bikes with virtual reality equipment exercised more when a competitive fictional character was introduced compared to cycling alone (but only if they scored highly on self-reported competitiveness). Furthermore, people have been found to consume more food (Castro 1994), donate more money (Izuma, Saito, and Sadato 2010), and spend more money (Sommer, Wynes, and Brinkley 1992), when with other people compared to being alone. However, sometimes the presence of another is seen to have detrimental effects. For instance, an analysis of archival data demonstrated that learner drivers who took their driving test with another individual awaiting their test present were more likely to fail than those who took the test without an observer (Rosenbloom et al. 2007).

#yourturn

Can you think of a time when you performed worse on a task because there was another individual present? And can you think of a time when you performed better on a task when there was another individual present?

3.2.2 A Reanalysis of Triplett's Data

Since 1898, more advanced statistical methods are now at researchers' disposal. Strube (2005) reanalysed Triplett's (1898) data, exploring both within-person differences in alone vs. together conditions (within-subjects tests) and differences between people across the alone vs. together trials (between-subjects tests). This re-analysis demonstrated that in general performance in the competition trials was better than the alone trials (between-subjects test); however, this was not a statistically significant difference. Likewise, looking at within-participant variation, there was only a marginally significant effect for performing better on the competition trials compared to a person's alone trials.

3.2.3 Replication of the Original Study

A recent pre-registered study directly replicated Triplett's (1898) original experiment, addressing some of the limitations mentioned earlier – namely the small (and underpowered) sample size, standardization of experimental trials and rest periods, as well as examining if gender moderated the effects (McHugh et al. 2025). This analysis of >400 children aged 7-13 years, who were age- and gender-matched, demonstrated that participants completed the task quicker during the together trials compared to the alone trials. Gender moderated this effect, with females completing the task faster on average, and the social facilitation/competitive co-action effect was stronger for females. This replication provides support for Triplett's original findings.

3.3 Conclusion

Overall, it appears that in some settings the presence of another (whether evaluative or non-evaluative, or co-actor, competitor or observer) affects performance. Often the presence of another appears to facilitate performance or dominant response tendencies, but the conditions under which this occurs need further examination as sometimes the presence of another hinders performance. While Triplett focused on competitive coaction effects, which was later termed social facilitation (and gave rise to this literature), it is important to note that the theory of social facilitation relates to mere presence of another individual affecting performance. Triplett's (1898) experiment and the more recent replication (McHugh et al. 2025) are not able to disentangle if the effects on performance are truly due to mere presence of another person (i.e., social facilitation) or due to competition.

> #yourturn
>
> Why do you think it matters whether performance depends on mere presence of others or if others need to be co-actors and/or competitors?

More research is needed to fully understand what is driving the observed effects. Think back to the practical implications section of this chapter. If we know under what conditions mere presence affects performance (positively and negatively), or under what conditions having someone engaged in the same task as us (co-actor) or even competing against us, then this can help us design optimal environments for a range of performance-based activities, such as learning and exercise/sport.

Chapter 4

Intergroup Contact Theory

written by Vanessa Müller (original draft), Milica Ninković (revision), Raul Szekely (revision), and Lukas Wallrich (revision)

4.1 The Classic

In the mid-20th century, after the horrors of World War II and during fights against official racial segregation, social scientists began asking a deceptively simple question: If you bring members of conflicting groups into contact, will they start to get along? Opinions were divided. Some warned that interracial contact would only breed "suspicion, fear, resentment, disturbance, and at times open conflict" (Baker 1934, pg. 120; cited in Pettigrew and Tropp 2006; I. N. Brophy 1945). Others were more optimistic, suggesting that isolation allowed prejudice to "grow like a disease" (Brameld 1946, pg. 245; cited in Pettigrew and Tropp 2006) and that, under the right circumstances, interaction could lead to "mutual understanding and regard" (Lett 1945, pg. 35; cited in Pettigrew and Tropp 2006). This debate set the stage for one of social psychology's most influential ideas: intergroup contact theory.

One of the first real-world tests of these ideas came in the United States Merchant Marine shortly after World War II. In 1946, sociologist Norman Brophy surveyed white sailors now serving in newly desegregated ship crews (I. N. Brophy 1945). He created a "prejudice index" from interview questions and looked for patterns. Expected predictors of racial attitudes – such as where a sailor was born or how much education he had – turned out not to matter much. Instead, direct personal contact was the standout factor. Brophy found that white seamen who had never shipped with a Black crewmate scored highest in prejudice, whereas those who had taken four or more voyages with Black crewmates scored the lowest. In the cramped, cooperative environment of a ship – an "artificial society" where survival depended on teamwork – many sailors discovered they

could no longer "afford the luxury" of prejudice. And Brophy wasn't alone in this observation. Similar studies, mostly in the United States, showed more positive attitudes among White police officers who worked with Black colleagues (Kephart 1957), and White residents who lived in mixed buildings where they had the opportunity to interact with Black neighbours (Deutsch and Collins 1951). These early findings suggested that prejudice was not immutable, but could change with contact.

#definition Prejudice

A negative attitude toward a group and its members, often based on stereotypes rather than direct experience.

These patterns spurred social scientists to theorise why and when contact might reduce prejudice. In his landmark book *The Nature of Prejudice* (1954), the psychologist Gordon Allport proposed what has become known as the contact hypothesis: the idea that under appropriate conditions, interpersonal contact between members of different groups can be one of the most effective ways to reduce intergroup prejudice. Crucially, Allport (1954) did not claim that contact always works. Instead, he specified four optimal conditions that, in his view, were needed for contact to reduce prejudice:

1. **Equal Status:** The groups should have equal status within the contact situation.
2. **Common Goals:** The groups should strive towards a mutually beneficial outcome.
3. **Cooperation (Not Competition):** The interaction should require cooperative effort from members of different groups.
4. **Support of Authorities or Norms:** The contact experience should have the explicit or implicit support of authorities, law, or social norms (e.g., teachers who encourage intergroup exchange explicitly).

Allport (1954) hypothesised that when these conditions are met, contact encourages people to view one another as individuals and teammates, and thus perceive members of the "other" group as part of a shared "us" rather than a separate "them." Interpersonal contact could then reduce ignorance and anxiety, increase empathy and understanding, and ultimately chip away at prejudice (Allport 1954). On the other hand, Allport (1954) warned that contact in unfavourable circumstances could backfire.

#yourturn

Think about a common intergroup contact situation in your community (for example, students from different backgrounds meeting at university, or neighbours from different ethnic groups interacting). Does that situation meet Allport's four optimal conditions (equal status, common goals, cooperation, and supportive norms)? How might the presence or absence of these conditions be influencing how well the groups get along?

4.2 The Aftermath

Allport's (1954) formulation of the contact hypothesis was hugely influential. It inspired a wave of research from the 1950s onward as psychologists, sociologists, and others researched the power of contact in a variety of groups and settings, mostly in observational research. By the turn of the 21st century, the evidence base had become enormous – though somewhat scattered. Hundreds of studies across dozens of countries and intergroup contexts had examined intergroup contact in one form or another, and the contact hypothesis had become a cornerstone of social psychology. The overarching question remained: Does contact typically work to reduce prejudice, and under what conditions?

#definition Observational Research

A study design where researchers measure variables as they naturally occur, without manipulating them. Observational studies can reveal associations between variables but cannot, on their own, establish that one causes the other.

4.2.1 The Classic Meta-Analysis

By the early 2000s, it was challenging to see the big picture in contact research. To address this, psychologists Thomas Pettigrew and Linda Tropp (2006) conducted a landmark quantitative review. In 2006, they published a meta-analysis synthesising findings from 515 studies (covering 713 independent samples and over 250,000 participants) that had studied intergroup contact. Across this vast body of work, they found a consistent pattern: people who reported more positive contact with members of an outgroup also tended to report lower levels of prejudice toward that group.

#definition Meta-analysis

A statistical technique that combines the results of multiple independent studies to estimate an overall effect. Meta-analyses can reveal patterns across a large body of research, but the quality of their conclusions depends on the quality and comparability of the included studies.

Pettigrew and Tropp (2006) concluded that "intergroup contact can promote reductions in prejudice" (p. 751) and that "there is little need to demonstrate further contact's general ability to lessen prejudice" (p. 766), even in situations when not all optimal conditions are met. The average effect size was substantial by social science standards (Cohen's d 0.43). With this uplifting message, their meta-analysis has become one of the most-cited papers in social psychology, with over 13,000 citations to date.

However, most of the studies they synthesised were observational rather than experimental, meaning they measured naturally occurring contact rather than

manipulating it. While observational studies are valuable for spotting consistent relationships, they cannot, on their own, establish that contact caused the reduction in prejudice. For that, experiments are usually needed, and only 5% of the studies in the meta-analysis are true experiments. Pettigrew and Tropp (2006) acknowledged this limitation but advanced various arguments why their results still indicate causal effects. Most importantly, studies that used more rigorous methods (for example, longitudinal designs or experiments) tended to find larger effects of contact than weaker, correlational studies did.

Pettigrew and Tropp (2006) also aimed to assess whether the benefits of contact generalise – that is, does having a positive experience with, say, one Black teammate make a white person feel more positively toward Black people in general? Encouragingly, many studies did find evidence of generalisation: improved attitudes often extended beyond the specific individuals involved to the outgroup as a whole. For example, if a white student befriended a Latino roommate, not only might their attitude toward that roommate improve, but their overall attitude toward Latinos could become more favourable as well. This kind of generalisation is crucial if contact is to have a broad social impact, and the meta-analysis indicated that it often occurs.

For a time, Pettigrew and Tropp's (2006) comprehensive review seemed to settle the debate: Intergroup contact works. With so many studies and an authoritative meta-analysis affirming that contact typically reduces prejudice (even outside of perfect conditions), the contact hypothesis gained even more prominence. Textbooks began to state confidently that positive contact is a proven method to improve intergroup relations. However, the story didn't end there. Sceptics and careful scientists raised important questions and cautions that would spark the next wave of investigations. Most importantly: is the evidence causal? If we observe that people who have more friends from other groups also show lower prejudice, it's not always clear which way the arrow of causality points – does contact reduce prejudice, or do less-prejudiced people simply seek out more contact? Pettigrew and Tropp's (2006) analysis went a long way toward addressing this by showing that the best studies (including experiments) found stronger effects, but still, the bulk of studies in their database were not true experiments. Additionally, critics wondered about unpublished null findings: were there "file drawer" studies where contact had no effect that were never known, potentially making the published literature look overly rosy? These cautions set the stage for a new generation of research that aimed to more robustly test when and how contact works – and to probe its limits.

> #yourturn
>
> Why is it important to go beyond correlational evidence (where we simply observe relationships) when evaluating whether intergroup contact truly reduces prejudice? What kinds of studies or methods would give more convincing evidence of causation?

4.2.2 New Insights and Challenges: Refining the Theory

By the 2010s, researchers began responding to these methodological concerns, bringing fresh scrutiny to the study of intergroup contact. For instance, a review by Elizabeth Paluck, Seth Green, and Donald Green (2019) specifically re-evaluated the contact hypothesis from a rigorous causal perspective. They exclusively focused on studies that met a high bar for evidence: field experiments with random assignment to a contact condition versus a control condition, and outcome measures assessed after the contact experience was concluded. Out of the thousands of contact studies conducted over the decades, Paluck, Green, and Green (2019) found only 27 experiments that fit these strict criteria up to that point. (Notably, almost two-thirds of those 27 had been published after Pettigrew and Tropp's (2006) meta-analysis, reflecting the field's recent push for experimental work.)

> #definition Experiment
>
> A study where researchers deliberately manipulate one or more variables and randomly assign participants to different conditions. Random assignment helps ensure the groups are similar before the intervention, so differences in outcomes are more likely to be caused by the manipulation rather than by pre-existing differences.

The good news was that, overall, the evidence from these rigorously controlled studies still supported Pettigrew and Tropp's (2006) basic conclusion: intergroup contact "typically reduces prejudice." In their meta-analysis of the 27 experiments, Paluck, Green, and Green (2019) found that the average effect of being randomly assigned to a positive contact experience was a reduction in prejudice levels compared to the control groups, with Cohen's d 0.39, very similar to Pettigrew and Tropp's (2006) result. This helps rebut the idea that the contact-prejudice link was merely a selection effect; even when people were assigned to have contact, prejudice tended to go down, on average. However, the experimental evidence also revealed some important caveats. One striking finding was that contact's effectiveness varied considerably by context and target group. In particular, interventions aimed at reducing ethnic or racial prejudices (for example, between Israelis and Palestinians, or between white and Black Americans) tended to show weaker effects than interventions aimed at reducing prejudice toward other stigmatised groups (such as people with disabilities or members of an opposing political party). In other words, contact worked least well for some of the most historically entrenched divides like race and ethnicity. On the flip side, contact interventions addressing prejudices that might be less emotionally charged or less tied to deep-rooted group identities (for example, toward the disabled, or between fans of rival sports teams) produced relatively larger improvements on average.

Paluck and colleagues (2019) also highlighted critical gaps in the evidence. For example, they found an almost complete lack of field experiments focused on adult populations dealing with racial or ethnic prejudice – the context the con-

tact hypothesis had originally been about and arguably still one of the most important areas for policy. Additionally, very few studies had systematically tested Allport's (1954) optimal conditions by manipulating those factors to see which mattered most. The authors concluded that these gaps need to be filled before we can confidently advise policymakers to rely on contact to remedy societal prejudice. In short, their message was not "contact doesn't work" but rather "contact can work, but we need better evidence, especially on the toughest cases and the crucial conditions, to understand how to use it most effectively."

> #yourturn
>
> Intergroup contact seems to yield larger prejudice reductions for some kinds of group differences (for instance, attitudes toward people with disabilities) than for others (like attitudes between ethnic groups). Why do you think this might be? Consider the nature of prejudice or anxiety in each case. What factors could make prejudice based on race/ethnicity harder to change through contact compared to prejudice toward people with disabilities, and vice versa?

4.2.3 An Outstanding Modern Study: Contact on the Soccer Field in Post-ISIS Iraq

To illustrate both the strengths and limitations of intergroup contact in action, consider a modern field experiment that put Allport's (1954) hypothesis to a challenging test. Political Scientist Salma Mousa conducted a remarkable study in post-conflict Iraq, published in 2020, to see if positive contact could help heal rifts between deeply divided religious communities (Mousa 2020). The setting was Northern Iraq in the aftermath of the ISIS terror reign. In 2014, ISIS had overrun the region, committing atrocities including the displacement of almost the entire Christian population from certain towns. By 2016, after ISIS was defeated, many displaced Christian families began returning to their hometown of Qaraqosh, a historically Christian town that had been scarred by violence. These returning Christians carried intense distrust and resentment toward the local Muslims. The Christians feared that some Muslim neighbours had been complicit with ISIS, or at least did not suffer as they had, and rumours and grievances ran rampant. In turn, Muslim residents felt unwelcome and resented the suspicions. In this tense post-ISIS context, the two groups lived segregated lives, with social contact minimal and fraught. Prejudice and fear were high on both sides.

Mousa (2020) wondered if a carefully designed contact intervention could begin to rebuild trust and coexistence in this environment. She chose a grassroots approach: recreational soccer teams. Why soccer? Importantly, soccer in this context naturally met many of Allport's (1954) optimal conditions for positive contact. For one, players on a team share a common goal – to win matches – and must cooperate closely to do so (passing the ball, strategising, etc.). Team sports also tend to equalise status; when everyone puts on the same jersey, they

have equal status as teammates on the field. Additionally, Mousa (2020) worked with local organisations and community leaders (including church officials) to support and endorse the league, lending authority approval to this intergroup activity. In short, the intervention was deliberately structured to tick all of Allport's (1954) boxes.

Here's how the experiment worked. Mousa (2020) invited young Christian men in Qaraqosh who were interested in playing soccer to form teams in a new reconciliation soccer league. These men formed teams mostly with friends or neighbours, so initially, all-Christian teams. The twist was that Mousa (2020) then randomly assigned half of the league's teams to receive several Muslim players as additions to their roster (the other half of the teams remained all-Christian and served as a control group). The Muslim players were recruited from outside the town (from camps of displaced Muslims nearby) and chosen to be of similar skill level to the Christian players, so that they could genuinely contribute on the field without dominating or being token outsiders. In total, each "mixed" team got three Muslim teammates added. All teams – mixed and all-Christian alike – then played in the same 8-week amateur league, facing each other in matches. Importantly, every other aspect of the league was the same for everyone: all teams had the same equipment, schedule, and participated under the same community-endorsed conditions, with the only difference being whether your teammates included Muslims or not. This experimental setup meant that if differences emerged between players on mixed teams versus all-Christian teams, the only systematic explanation would be the experience of having (or not having) Muslim teammates.

At first, the intervention faced friction. Some Christian players were unhappy about Muslims joining their teams. In the early weeks, there were incidents of mistrust and even hostility – for example, a few Christian team members openly told the organisers "We don't want Muslims; they will ruin the league." Such remarks underscored just how deep the suspicion ran in this community; it wasn't an easy start. But as the season progressed and these young men practised and competed side-by-side, the tone began to shift. By about the mid-point of the season, signs of camaraderie had emerged. One small episode stood out: when some Christian players learned that their new Muslim teammates were struggling to afford taxi fare to the games (travelling from a distant displacement camp), the Christian players pooled money to help cover the cost so their teammates could make it to matches. On the field, teammates started to celebrate goals together and encourage one another. Over time, a shared team identity – we are the Lions, we are teammates – began to form, overlaying the previous religious divide. A Christian player, asked later about his experience, reflected that "I learned that Muslims could be friends of ours, even like brothers." The transformation was not instant or universal, but by the end of the league, many of the initial anxieties had given way to friendly competition and mutual respect on these mixed teams.

So, did this Allportian (Allport 1954) contact experience actually change at-

titudes or behaviors? Mousa's (2020) results were revealing. They showed both encouraging positive outcomes and clear limits. First, consider the effects within the context of the league itself – that is, how the Christian players felt and acted toward their Muslim teammates (and other Muslims in the league): The Christian players who had Muslim teammates ended up displaying significantly more positive behaviours toward Muslim peers compared to players on all-Christian teams. For example, at the end of the season, each team voted for a member of an opposing team to receive a sportsmanship award. Christians on mixed teams were more than 15 percentage points more likely to vote for a Muslim player (from another team) for this award than were Christians on all-Christian teams. This indicated greater esteem and fairness toward Muslim peers. Moreover, when sign-ups opened for a new season, the mixed-team Christians were much more willing to play on a mixed team again – they registered at higher rates for a subsequent mixed league – whereas many all-Christian team players declined to sign up once they heard teams might be mixed. Perhaps most impressively, about six months after the experiment, Mousa (2020) found that many of the mixed-team players were still regularly meeting up with their former Muslim teammates to practice together and maintain their friendship. In fact, roughly one-third of the mixed teams continued to meet socially for pick-up soccer games long after the official league ended, whereas almost none of the all-Christian teams chose to continue gatherings that included outgroup members. These findings show that meaningful friendships and trust did form through the contact intervention. By all accounts, prejudice had decreased, at least with respect to those specific Muslim teammates and other known Muslim players.

However, now consider what happened outside the context of the league – in the broader community and in attitudes toward Muslims in general. Here, the findings were more sobering: The positive effects of contact did not substantially generalise to Muslims beyond those directly encountered. In surveys and behavioural measures after the season, Christian participants who had played with Muslim teammates showed no significant change in their willingness to interact with unknown Muslims or visit Muslim communities compared to the control group. For instance, having had Muslim teammates generally did not make Christian players more likely to say they would patronise a restaurant in a nearby majority-Muslim city, nor did it increase their attendance at a mixed social event in town. When asked about broader attitudes, those who experienced contact did express somewhat stronger abstract support for coexistence or the idea that Christians and Muslims could be friends, but their core beliefs about Muslims as a group (for example, levels of trust toward Muslim strangers or stereotypes about Muslims) remained essentially as negative as before. In Mousa's (2020) own words, while the Christian players found it possible to trust and befriend specific Muslim individuals they got to know, extending trust to Muslim strangers outside that circle was "too much of an ask" in the aftermath of war. In short, the contact intervention succeeded in forging new cross-group friendships and improving attitudes toward those individuals, but

it largely failed to shift the participants' generalised feelings about the outgroup as a whole or their behaviour in other contexts.

This pattern – friendships without broad reconciliation – highlights a crucial challenge for intergroup contact theory. Mousa's (2020) study offers an inspiring proof-of-concept that even in a highly fraught, post-conflict setting, a well-designed contact program, featuring Allport's (1954) optimal conditions, can produce genuine goodwill and cooperation between former adversaries. The fact that young men who initially hated the idea of playing with "the other" ended up forming lasting bonds is powerful. It shows that under the right conditions, enemies can indeed become teammates, even friends. On the other hand, the limited scope of these changes tempers the optimism. The contact in this study changed how people felt about particular outgroup members, but not necessarily about the outgroup at large. From a policy or peacebuilding standpoint, that is a big limitation: improving one-to-one relationships is wonderful for those individuals, but it may not significantly mend the overall social fabric or reduce the kind of generalised fear that fuels wider conflict. Mousa's (2020) findings align with what many other studies have found and what is now a central puzzle in contact research – the generalization problem. How can we ensure that the effects of contact spread beyond the immediate participants and influence attitudes more broadly? If positive contact only affects the small circle of people directly involved, its ability to reduce community-wide prejudice or conflict is limited.

> #yourturn
>
> In the soccer study, Christian players clearly grew more accepting of the Muslim teammates they got to know personally, yet their attitudes toward Muslim strangers remained unchanged. Why do you think a positive experience with a few individuals might fail to generalise to the entire outgroup? What psychological factors might be at play? Can you think of any additional measures or tweaks to the intervention that might help encourage broader changes in attitudes or trust (for example, activities that mix the groups in other settings, discussions that address group stereotypes, etc.) to help bridge that gap?

Mousa's (2020) soccer experiment encapsulates both the promise and the limitations of intergroup contact. It provides a vivid example that contact can work – even under pretty challenging conditions, it built trust and friendship where there was initially fear and hostility. At the same time, it underscores that a single intervention, even a well-crafted one, is no panacea for deeply rooted prejudices. Especially in contexts of recent violence and trauma, biases may run so deep that it takes much more than a brief intervention to budge generalised attitudes. These nuanced outcomes have prompted researchers to investigate strategies to amplify and extend contact effects. How might we design contact interventions that not only improve attitudes toward the people directly involved, but also shift perceptions of the broader group? This remains

an active area of research.

In fact, as the field has progressed, experts have adopted a more cautious tone about what contact can realistically achieve on its own. In 2021, Paluck et al. (2021) published an extensive review of 418 prejudice-reduction experiments conducted between 2007 and 2019, a collection that included many contact-based interventions alongside other approaches. The results of this review were mixed and somewhat concerning. On one hand, many of the experiments reported at least some positive effects on attitudes, suggesting there are reasons for optimism. On the other hand, the authors uncovered "troubling indications of publication bias," meaning that studies showing big success were likely overrepresented in the literature, while those with null or tiny effects may not have been published. When they statistically accounted for this bias, the overall picture became less rosy. Furthermore, three-quarters of interventions in that review were very "light-touch" or brief, such as a short workshop, a single encounter, or a one-time media exposure. Not surprisingly, any positive changes from such brief interventions often faded over time or were quite limited in scope. In the relatively few cases where more intensive, long-term interventions were implemented (what the authors called "landmark studies"), the effects on prejudice tended to be modest at best. This included some multi-week educational programs, extended intergroup dialogues, and other sustained efforts – many showed only small improvements, highlighting how stubborn prejudices can be. Paluck-et-al_2021 concluded that new theoretical innovation is needed to achieve larger and more lasting impacts. They suggested that perhaps contact on its own is often too limited, and that combining contact with other approaches (or addressing larger structural issues in tandem) might be necessary to produce more substantial change. In their view, simply throwing diverse people together for a short period is rarely a magic fix; researchers need to think bigger about the mechanisms of change and consider multi-pronged solutions.

Most recently, the strongest tests of the contact hypothesis have been compiled in a 2025 meta-analysis by economist Matt Lowe (Lowe 2025). Lowe (2025) focused exclusively on the highest-quality studies: those that were pre-registered, randomised experiments on intergroup contact.

> #definition Pre-Registered Study
>
> A study in which the researchers publicly register their hypotheses, methods, and analysis plan before collecting data. Pre-registration helps increase transparency and credibility – it prevents researchers from changing their analyses or selecting results after seeing the data, which can lead to false-positive findings.

By zeroing in on these rigorously planned studies, Lowe (2025) aimed to eliminate biases introduced by practices like p-hacking or cherry-picking of data – practices that can inflate apparent effects.

> #definition p-hacking

> The practice of misusing data analysis to find patterns that can be presented as statistically significant, often by trying many variable combinations or statistical tests until something "significant" turns up. This can lead to unreliable conclusions because it capitalises on chance patterns in the data.
>
> #definition Cherry-Picking
>
> Reporting only the data, outcomes, or time frames that support one's hypothesis while ignoring or dismissing those that do not. This makes the story or articles simpler and might make them more publishable, but provides a distorted view of the evidence.

The findings are instructive. When considering only these methodologically pristine studies, the average effect of intergroup contact on prejudice outcomes was much smaller than earlier reviews had suggested, with d 0.1, a quarter of the effect size suggested by Pettigrew and Tropp (2006). In plain language, this means that the effect is statistically significant, but on average, quite modest. Lowe's (2025) meta-analysis also reinforces the now-familiar theme about specificity vs. generality: contact's benefits tend to be localised. People's attitudes and behaviours toward the particular individuals they met often improved more strongly than their attitudes toward the outgroup in general. Broad attitude change was much less common, with many studies finding little to no shift in generalised prejudice or policy views even when interpersonal warmth increased. That contact's effects often fail to generalise widely is now recognised as one of the central challenges in the field.

4.3 Conclusion

Seven decades after Allport (1954) first set out the contact hypothesis, it remains a cornerstone of prejudice-reduction research. From post-war merchant ships and integrated housing projects to modern field experiments, the idea has consistently shaped both science and policy. Its core message that prejudice is not fixed and can change through structured, positive interaction helped shift thinking away from segregation toward integration as a deliberate tool for improving relations.

The evidence, however, shows that contact is no cure-all. Gains are often local, improving attitudes toward specific individuals but failing to generalise to the wider group. Outcomes depend heavily on context, structure, and the quality of interaction. For that, Allport's (1954) optimal conditions (equal status, common goals, cooperation, and authority support) remain a useful guide, though further research is needed. Contact is also not always positive, and researchers have started taking that more seriously, as negative experiences can be as powerful, if not more so, than positive ones, though they are fortunately rare (Paolini et al. 2024). Contact also needs to be understood in context, as broader forces such as inequality, political division, and historical grievances can limit its impact.

Here, psychologists can fruitfully cooperate with other social science disciplines.

Studying the most meaningful forms of contact – deep, sustained relationships forged over years – poses particular challenges. Such relationships cannot be randomly assigned, develop slowly, and are difficult to measure without disrupting them. Creative, flexible designs are therefore needed, with interpretations that acknowledge their limitations. Current research focuses on making contact more effective and lasting. Promising approaches include pairing it with perspective-taking, cooperative learning in schools, norm-shaping media campaigns, or virtual-reality simulations of positive encounters. Some initiatives also embed contact in long-term community projects or redesign institutions, such as integrated workplaces or mentoring networks, so diverse cooperation becomes part of daily life. The challenge ahead is to move from showing that contact can work to understanding how to make it work consistently, at scale, and for the long term.

> #yourturn
>
> The IAT was designed to assess automatic associations people may hold unconsciously. However, if implicit and explicit attitudes are highly correlated, what are the implications for how we understand the relationship between conscious and unconscious mental processes?

Chapter 5

Implicit Association Test and Attitudes

written by Karolin Kessel (original draft), Bradley Baker (revision), Savannah C. Lewis (revision)

5.1 1. The Classic

How can we know what people truly think about a certain topic? This question is difficult to answer, given that people are sometimes motivated to misreport their true attitudes.

> #definition Definition of "attitude"
>
> The cognition, affect and behavioral tendencies towards a certain object.

For example, if a friend was very excited about a new band they discovered, you might feel like you don't want to burst their bubble of joy by telling them you don't enjoy the music as much as your friend. Because it is socially desirable to respond positively, to mirror your friends' liking of the band, you might misreport your true attitude to them.

> #yourturn
>
> Think back to a time you thought or felt differently from what you expressed publicly. Why did you not report the truth?

Researchers in social psychology have been working on ways to assess and measure people's attitudes towards a multitude of different topics. The Implicit Association Test (IAT), developed by Greenwald, McGhee, and Schwartz (1998),

is a psychological tool to measure implicit attitudes that people may not be aware of or may not openly express.

#definition Implicit Attitude

An enduring mental disposition toward something that is not consciously identified and of which a person may lack awareness.

The test works by measuring how quickly people process and respond to pairs of words or images. It relies on the idea that people respond faster when two concepts that are closely linked – or associated – in their mind (a congruent association) are paired than when the pairing feels mismatched or unrelated (an incongruent association, Jhangiani and Tarry 2022).

#definition Association

"A connection or relationship between two items (e.g., ideas, events, feelings) with the result that experiencing the first item activates a representation of the second" ("APA Dictionary of Psychology" 2018).

#definition Congruent Association

A mental relationship between two objects or concepts characterized by agreement or harmony.

#definition Incongruent Association

A mental relationship between two objects or concepts characterized by lack of harmony or misalignment.

Measuring these reaction times allows researchers to understand the strength of automatic associations between concepts. Being able to measure implicit attitudes provides a different perspective on how people feel than is available from simply asking people to report their attitudes, either because people may not realize their preferences or may not be willing to share them. The IAT is designed to reveal unconscious biases and remove bias that can be introduced by people simply giving answers they think are socially acceptable, rather than what they truly believe.

#definition Implicit Association Test

A reaction-time task that measures the strength of automatic associations between concepts (e.g., flowers and positivity) by comparing how quickly people classify paired categories. Faster responses indicate stronger underlying associations.

In the classic study by Greenwald, McGhee, and Schwartz (1998), participants completed a task called the Implicit Association Test (IAT) to measure automatic associations. They were asked to quickly sort words and pictures into four groups: two groups of objects (like flowers and insects) and two groups of feelings (pleasant and unpleasant words). The test had two main parts. In

one part, participants pressed the same key for flowers and pleasant words, and another key for insects and unpleasant words. In the other part, the pairings were switched: flowers with unpleasant words and insects with pleasant words. Participants had to respond as fast and accurately as possible. If someone naturally associates flowers with pleasantness, they will respond faster when those two categories are paired together.

> #yourturn
>
> Why would researchers not simply ask participants about their attitudes?

This approach is used to capture unconscious connections between concepts in memory, which in the original test format were aimed at assessing implicit stereotypes and prejudices, but have been used to identify a variety of subtle attitudes in various subject areas (Nosek and Smyth 2007).

> #yourturn
>
> Do you think that reaction times or spontaneous reactions are an appropriate measure of implicit cognitions such as stereotypes? Why or why not?

5.2 2. The Aftermath

In his study, "The Implicit Association Test: A Method in Search of a Construct," Ulrich Schimmack (2021) examines the power of the IAT in revealing individual differences in implicit social cognition.

> #definition Implicit Social Cognition
>
> The automatic, unconscious mental processes that influence how we perceive, evaluate, and interact with others.

The results show that there is insufficient evidence for the construct validity of the test (Schimmack 2021), in other words, that there is not enough proof that the IAT measures what it was intended to measure (implicit bias), rather than something else. This can be seen when scores from a test aren't related to other measures of the same concept in the expected ways. Based on examination of several multimethod studies, Schimmack found little or no evidence of discriminant validity compared to measures of explicit attitudes, making it unclear whether the test really captures a different type of attitude (implicit, rather than explicit). Problems with discriminant validity show up when a measure's scores are too similar to those of an established measure for a different concept, making it unclear whether the new measure is assessing something unique. That is, Schimmack (2021) raises questions regarding a lack of evidence that the IAT adequately measures individual-level differences in implicit associations and the extent to which the IAT measures something different from self-report measures of explicit associations.

#definition Multimethod Study

Research that employs two or more distinct methods.

#definition Discriminant Validity

The extent to which a test is unrelated to measures designed to assess theoretically distinct constructs.

#definition Construct Validity

The extent to which a test measures the theoretical construct or concept it is intended to measure.

Schimmack highlights that these deficiencies have been overlooked for many years and finds that explicit measures are more valid than the IAT in all areas. This means simply asking participants about their attitudes might indeed be the better measure of these attitudes than making them take the IAT. At the same time, Schimmack also argues that the IAT can be used as a complementary measurement tool to explicit measures for sensitive settings to reduce measurement errors by employing a multi-method measurement model. In other words, using both explicit measures and the IAT might be the best approach.

#yourturn

What are the pros and cons of using this kind of test in bias training?

5.3 3. Conclusion

The establishment of implicit association testing resulted in one of the most influential articles in personality and social psychology (Greenwald, McGhee, and Schwartz 1998), and established the foundation for a variety of new (social psychological) theories (Schimmack 2021). At the same time, the difficulties identified by Schimmack illustrate the extent to which social psychological theory formation is highly complex. Particularly when investigating the discrepancy between human thinking and socially desirable conformity, as well as its (uncertain) influence on behavior, precise (construct) differentiation and validity testing are essential in research. Extensive research has used and built on the IAT (Greenwald, Nosek, and Banaji 2003; Greenwald et al. 2009) and related approaches to measuring the strength of automatic associations or using implicit measures to bypass bias in research data due to socially desirable responding by study participants.

#definition Social Desirability

The tendency to want to be viewed positively by others, often by aligning with socially approved behaviors and attitudes.

#definition Socially Desirable Responding

> The act of providing inauthentic responses to better present oneself favorably according to current social norms.

However, Schimmack (2021) highlighted weaknesses of the IAT regarding construct and discriminant validity as a measure of implicit constructs. He emphasizes the significance of being cautious when making claims about subtle ideas based on the IAT and highlights the variation in the IAT's validity depending on the construct being measured. If the IAT is not measuring implicit attitudes (unconscious biases) or does not provide additional information beyond simply asking people about their attitudes, as suggested by Schimmack, then the test offers limited utility to researchers and calls into question findings that rely on the IAT.

> #yourturn
>
> The IAT was designed to assess automatic associations people may hold unconsciously. However, if implicit and explicit attitudes are highly correlated, what are the implications for how we understand the relationship between conscious and unconscious mental processes?

Chapter 6

False Consensus Effect

written by Marcel Zubrod (original draft), Jana Berkessel (revision), and Márton Kolozsvári (revision)

6.1 The Classic

The false consensus effect is a cognitive bias in which individuals overestimate the extent to which their own beliefs, preferences, and behaviors are shared by others.

> #definition Bias
>
> A systematic distortion of perception or judgment.

This psychological phenomenon was first systematically studied by Ross, Greene, and House (1977), who demonstrated that individuals tend to perceive their own choices and opinions as more common than they actually are. For instance, people who express a preference for a particular option are likely to assume that others would make the same choice, even when evidence suggests otherwise. This bias occurs because individuals use their own perspective as a reference point, leading to distorted judgments about the preferences, opinions and behaviors of others.

> #definition False Consensus Effect
>
> A cognitive bias where individuals overestimate the extent to which others share their beliefs, preferences, and behaviors.

In Study 1 of the original research by Ross, Greene, and House (1977), participants were presented with one of four short stories, each describing a fictional scenario with a behavioral choice to be made. After reading the assigned story, participants were asked to estimate the percentage of their peers who would choose one behavioral option over the other within the context of the story.

#yourturn

Can you think of a time when you assumed others thought or behaved the same way you did and it turned out to not be the case?

Following these percentage estimates, participants completed a questionnaire. First, they were required to indicate which behavioral option they personally would have chosen in the scenario. Next, they rated themselves on a personality scale. As part of the assessment, participants also evaluated the typical personality characteristics of someone their age and gender who would choose either behavioral option presented in the story.

The results revealed a consistent pattern: participants who chose a particular behavioral option tended to believe that "people in general" would likely make the same choice. Conversely, participants who rejected an option perceived that behavior as less likely for others. Across all four stories, participants' own choices strongly predicted their estimates of how the general population would behave.

Additionally, significant differences emerged in personality evaluations based on participants' own choices. For three of the four stories, participants rated the typical personality traits of those choosing their preferred behavioral option as less extreme than those who selected the alternative. These effects were statistically significant in three stories, while one story showed a weaker significance, and the fourth story showed no significant results.

#yourturn

Are there certain methodological choices that could enhance or reduce the magnitude of the false consensus effect? These could include, but are not limited to, the number of choices to choose from, the social setting, the controversiality of the choices and the order of choices. Do they increase or reduce the magnitude of the false consensus effect?

6.2 The Aftermath

A meta-analysis by Mullen et al. (1985) examined 23 studies and a total of 115 hypotheses related to the false consensus effect. The analysis demonstrated that tests for the false consensus effect were highly significant and produced a moderate effect size. Importantly, it identified specific methodological factors that influenced the magnitude of the effect. For instance, the number of behavioral decisions participants were asked to make, as well as the order in which decisions and consensus estimates were presented, significantly impacted the observed false consensus effect.

#definition Effect Size

> A quantitative measure of the magnitude of a phenomenon, used to assess the practical significance of research findings.

These findings suggested that subsequent studies should limit the number of behavioral decisions participants are required to make and prioritize consensus assessments before behavioral decisions, as those methodological peculiarities might maximize the observed extent of the false consensus effect in experimental settings.

The **self-presentation explanation** posits that individuals strategically align their behavior with perceived social norms. According to this theory, the false consensus effect should be more pronounced when individuals make their behavioral decision before estimating the consensus. Only in this sequence do participants have the chance to adjust the social norm (i.e., other people's behavior) to their own behavior. However, the meta-analysis by Mullen et al. (1985) found no statistical evidence supporting this prediction, suggesting that the false consensus effect does not vary as the self-presentation explanation would anticipate.

> #yourturn
>
> Which other mechanisms could explain the False Consensus Effect? How would you test those mechanisms?

mullen_false_1985 outlined several theoretical explanations for the false consensus effect. One explanation, **attributive projection**, suggests that individuals rely on cognitive biases to justify their belief that their own behavioral choices are rational and appropriate responses to the environment. Another perspective suggests that the false consensus effect can **protect a person's self-esteem**. It may help people feel better about themselves when they face failure or receive negative feedback about their personal characteristics. A third explanation focuses on **social environments**, noting that people tend to associate with others who share similar backgrounds, values, and interests. Using false consensus makes us associate with the others who are (often falsely) perceived to be similar, thus fulfilling the need for a sense of relatedness. This selective association reinforces the perception that their choices are widely shared. Finally, **cognitive availability** provides a more mechanistic account, proposing that the behaviors individuals have chosen—or would choose—are more easily recalled or imagined than alternative actions when theorising about the behavior of others, a phenomenon linked to the availability heuristic.

> #definition Availability Heuristic
>
> A mental shortcut where people estimate the likelihood of an event based on how easily examples come to mind, which can lead to overestimating rare but memorable occurrences.

Overall, the false consensus effect is often attributed to a psychological desire to see one's thoughts and actions as appropriate, normal, and correct. Together, these cognitive and motivational factors help explain why individuals

consistently overestimate the prevalence of their own opinions and behaviors, a phenomenon observed across numerous studies.

Recent research has refined our understanding of the false consensus effect, particularly by situating it in contemporary social and digital contexts. In a series of studies, Bunker and Varnum (2021) found that greater social media use was reliably associated with stronger false consensus effects across domains such as political attitudes, personality traits, and social motives. However, the size of these effects was consistently smaller than laypeople anticipated, suggesting a public overestimation of social media's distorting power. Luzsa and Mayr (2021) experimentally demonstrated that exposure to attitudinally congruent news feeds, especially those with high agreement and visible endorsement cues like "likes", leads individuals to overestimate public support for their own views. Interestingly, this inference was moderated by participants' interest in the topic, with highly engaged individuals showing more skepticism toward consensus cues.

Building on the political implications of false consensus, Steiner, Landwehr, and Harms (2025) found that individuals who overestimate how many others share their political preferences are more likely to express populist attitudes and to distrust political elites. Similarly, Weinschenk, Panagopoulos, and Linden (2021) showed that individuals' views about democratic norms, such as the peaceful transfer of power, were strongly linked to their perceptions of what others believe—indicating a false consensus bias, particularly among conservatives. Finally, Furnas and LaPira (2024) extended the scope of the false consensus effect to unelected political elites (e.g., lobbyists and journalists) demonstrating that this group's perceptions of public opinion systematically reflected their own views, suggesting egocentrism rather than ideological bias as the driving force.

Together, these studies demonstrate that the false consensus effect is a robust phenomenon with wide-ranging relevance from digital communication to political judgment and that it is shaped not only by cognitive mechanisms but also by the structural, technological, and ideological environments in which opinions are formed.

6.3 Conclusion

The body of research on the false consensus effect highlights its robustness as a psychological phenomenon while also revealing important complexities in how it comes about. Early experimental studies, such as those by Ross, Greene, and House (1977), demonstrated that individuals consistently overestimate the degree to which others share their beliefs and behaviors. Follow-up meta-analyses, like that of Mullen et al. (1985), confirmed the effect's significance and explored the methodological and contextual factors that influence its magnitude.

In the broader context of social psychology, the false consensus effect provides valuable insights into how cognitive biases and motivational factors shape human perception. Explanations for the effect, ranging from attributive projection and

ego defense to mechanisms like cognitive availability, underline the interplay between how individuals view themselves and how they perceive the social world around them.

However, as with many constructs in psychology, it is crucial to approach findings on the false consensus effect with careful scrutiny. Methodological variations can significantly impact the observed magnitude of the effect, and further research is needed to disentangle its underlying mechanisms. The enduring study of the false consensus effect is an example of the importance of revisiting and refining theoretical constructs to build a more comprehensive understanding of human cognition and behavior.

Chapter 7

Facial Feedback Hypothesis

written by Sophia Reitmayer (original draft), Patrícia Arriaga (revision), and Effy Zachou (revision).

7.1 The Classic

Does what your body does influence how you feel? This is a central question that the Facial Feedback hypothesis addresses. The idea is simple, and quite old. In fact, it echoes one of the earliest theories of emotions in modern psychology: the James-Lange theory of emotion (James 1884). This theory proposes that bodily changes precede and give rise to emotional experiences. In other words, perhaps what our body does informs what we feel.

> #yourturn
>
> Have you ever felt your heart beat faster when giving a presentation or walking into a room full of people, and then noticed yourself feeling nervous or fearful? These are examples of how bodily responses, such as a racing heart or sweating, might shape emotional experience, as suggested by James-Lange theory (James 1884).

Now think more specifically: have you ever noticed that frowning while concentrating made you feel more tense? Or that you felt more positive when you smiled, even without a clear reason? These are everyday examples of how facial expressions, as specific bodily reactions, might affect your emotional state, as proposed by the Facial Feedback Hypothesis.

The Facial Feedback Hypothesis can also be related to the work of Darwin (1872) and, later, Ekman (1992), as both suggested that facial expressions play a role in emotion. Ekman, for example, emphasized that certain facial expressions are universal and biologically innate. However, these theories are distinct, since unlike the Facial Feedback Hypothesis, neither Darwin nor Ekman proposed

that facial expressions causally influence the emotional experience itself. In contrast, the Facial Feedback Hypothesis suggests that the activation of facial muscles involved in an expression can modulate the subjective experience of emotion. This theory posits that the act of forming a facial expression, such as smiling, frowning, or furrowing the brow, can intensify, initiate, or modulate the corresponding emotional state, thereby establishing a bidirectional relationship between expression and affect. Thus, the act of smiling may actually make people feel happier.

> #yourturn
>
> Can you think of everyday situations where the Facial Feedback Hypothesis might apply? Try to go beyond smiling, by considering how other facial expressions might also shape your emotional experience, such as sadness, anger, fear, disgust.

The publication by Strack, Martin, and Stepper (1988) investigated this hypothesis in two studies. The authors tested whether adopting a facial expression typically associated with a specific emotion could influence people's emotional experience and their evaluation of external stimuli. More specifically, they investigated whether producing a smiling facial expression could lead to a more positive evaluation of cartoons and a more positive emotional state.

Strack, Martin, and Stepper (1988) conducted two studies using a new methodology designed to prevent a cognitive interpretation of facial action. In other words, the aim was to avoid participants consciously recognising their facial movements as expressions of specific emotions. This was important because one of the main concerns is the risk of demand characteristics, that is, the possibility that participants' awareness of the study's true purpose might influence their responses. To address this, they introduced a cover story, telling participants that the study focused on psychomotor coordination. This procedure became known as the "pen-in-the-mouth" paradigm, allowing for a more subtle manipulation of facial muscle activity.

In both studies, participants (N = 92, Study 1; N = 83, Study 2) used the same pen-in-the-mouth paradigm. In study 1, participants were assigned to three conditions. In one condition, participants were asked to hold a pen between their teeth in a way that would facilitate a facial configuration associated with smiling ("teeth" condition). In this condition, the way participants held the pen would activate the facial zygomaticus major muscles, which are typically involved in smiling.

> #definition Zygomaticus Major Muscles
>
> These bilateral facial muscles, when activated, raise the corners of the mouth in an upward and lateral direction, facilitating expressions such as smiling.

In a second condition, they were asked to hold the pen between their pursed lips ("lips" condition). In contrast to the "teeth" condition, this position engages

the orbicularis oris muscles, which may inhibit the activation of the zygomaticus major, making smiling more difficult.

> #definition Orbicularis Oris Muscles
>
> These are circular muscles around the mouth that close the lips and produce puckering, as in kissing or whistling.

The third condition included in study 1 served as a control group, as it did not involve any direct manipulation of the facial muscles. Instead, participants were asked to hold the pen with their non-dominant hand.

In study 1, the aim was to test whether facial manipulation influenced the evaluation of humorous stimuli (perceived funniness) and study 2 aimed at replicating the procedure but also differentiating the effects on cognitive and affective components of this response. Thus, after being assigned to one of the conditions, participants were presented with cartoons on various topics, ranging from neutral to humorous situations, and asked to rate how funny each cartoon was on a scale of 0 to 9 ("not at all funny" to "very funny"). Additionally, in study 2, the affective experience of amusement was measured by asking participants to indicate how amused they felt while viewing the cartoons, also using a 10-point scale (from 0 = "I felt not at all amused" to 9 = "I felt very much amused").

The results in study 1 showed differences in the ratings of the cartoons between the "teeth" and "lips" conditions. In the "teeth" condition, participants rated the cartoons as significantly funnier than in the "lips" condition, and the results of the control group fell between these two conditions. This suggests that activating the facial muscles involved in smiling can lead to a more positive perception of humorous stimuli, while inhibiting those muscles reduces this positive perception. In study 2, by introducing separate measures for cognitive and affective components, the authors showed that facial manipulation affected only the amusement experience without affecting the cognitive evaluation of funniness. This highlights the need to distinguish between these two components explicitly. According to the authors, the effects obtained in the perceived funniness of the cartoons in study 1 likely reflected a combination of affective and cognitive influences within a single global evaluation.

> #yourturn
>
> Why do you think Strack et al.'s (1988) publication was so influential? Are you fully convinced? Are there exceptions to the rule? How could facial-feedback be criticized?

Over the years, several questions have been raised, and both conceptual and direct replications of Strack et al.'s (1988) study have been conducted. For example: Are facial feedback effects stronger when people produce genuine, spontaneous smiles, compared to subtle and artificial manipulations like holding a pen in the mouth? Does facial feedback initiate emotional experiences, or does it merely amplify emotions that are already present? Also, although Strack et al. (1988) focused specifically on smiling, the Facial Feedback Hypothesis

suggests that other facial expressions may also contribute to shaping emotional experience.

7.2 The Aftermath

Strack et al.'s (1988) influential study has been the subject of debate in recent years, as several researchers have had difficulties replicating the original results. One of the attempts was the Registered Replication Report (RRR) by Wagenmakers et al. (2016). Despite coordination across 17 independent laboratories, the replication failed to reproduce original findings: participants did not rate cartoons as funnier when their facial muscles were configured into a smile. This null result raised doubts about how reliable the facial feedback hypothesis is. In response, Strack (2016) argued that small differences in the setting, especially the use of video cameras, may have affected the participants' responses. He suggested that being watched could make people more self-aware and stop the natural reactions needed for facial feedback to work. Later, Noah, Schul, and Mayo (2018) investigated this concern by examining whether the presence of video cameras could alter participants' behavior. In two experiments, they compared conditions with and without video monitoring. The results showed that the pen-in-mouth task influenced results only when participants were not being observed. This suggests that the facial feedback effect is influenced by whether people feel they are being monitored, and that subtle changes in the study design can affect the results.

More recently, Coles et al. (2022) contributed to this debate with the Many Smiles Collaboration, designed as a large-scale, pre-registered multi-lab project to test the facial feedback hypothesis through both direct and conceptual replications.

> #definition Conceptual Replication
>
> A study that aims to recreate the gist of a prior study without using an identical procedure. These studies often aim to explore boundary conditions, the influence of specific variables, or aim to broaden and extend a certain finding.

Conducted across 19 countries with data from 3,878 participants, their study used various methods to examine the reliability of facial feedback effects. Participants were asked to imitate prototypical or less prototypical facial expressions of happiness (facial mimicry paradigm) or to perform voluntary facial movements (voluntary facial action). In addition, the pen-in-the-mouth paradigm from Strack, Martin, and Stepper (1988) was used, in which participants held the pen either between their teeth or between their lips.

The results showed that, when present, the effects were small, supporting the idea that facial feedback contributes to emotion but is not its primary determinant. There was consistent evidence of emotional amplification in voluntary

smiling and mimicry tasks, while results for the pen-in-mouth task were less clear, even when avoiding video recording. It is worth noting that Strack was directly involved in this project, highlighting the project's collaborative effort to test the facial feedback hypothesis. The results suggest that different mechanisms may underlie the effects of each task. Rather than refuting the facial feedback hypothesis, Coles et al.'s (2022) findings frame it as a conditional and modest phenomenon, dependent on how facial expressions are elicited and on contextual factors such as participant awareness.

7.3 Conclusion

Attempts to replicate Strack et al.'s (1988) original findings have produced inconsistent results. Importantly, the Many Smiles Collaboration (Coles et al. 2022) did not provide clear evidence regarding the emotional amplification effect of the pen-in-mouth task used in Strack et al.'s (1988) study. However, this recent project broadened the scope of investigation by including additional paradigms, such as voluntary smiling and facial mimicry, which yielded small but consistent facial feedback effects.

Overall, the evidence suggests that facial feedback can influence emotional experience, but its effects are small, sensitive to context, and not consistent across all types of manipulations. These studies also highlight the importance of identifying the conditions under which facial feedback operates.

From the perspective of James-Lange theory, the findings remain consistent with the idea that bodily changes contribute to affective experience, though in a more limited and conditional way than originally assumed.

> #yourturn
>
> In light of these results, would you say that smiling more will make people feel happier?

In short, the relationship between facial expressions and emotions is complex. Such effects may occur, but they are usually small, context-dependent, and further research is still needed to determine when and how they emerge. Smiling alone is unlikely to serve as a simple route to happiness.

Chapter 8

Heat Priming-Hostile Perception Effect

written by Hannes Dieterle (original draft), and Patrícia Arriaga (revision)

8.1 1. The Classic

Does what you have recently seen, heard or read affect how you think, even if you do not realise it? This is a central question behind the concept of priming, which has been used to describe how subtle cues, like words related to temperature, might influence our thoughts, feelings, and behaviours.

The verb "to prime" means "to activate". In psychology, "priming" refers to the idea that exposure to a stimulus can activate mental representations, making it easier or faster to respond to that same stimulus later (direct priming), or to something related to it (indirect priming).

> #definition Definition of "priming"
>
> "Priming refers to facilitative effects of an encounter with a stimulus on subsequent processing of the same stimulus (direct priming) or a related stimulus (indirect priming)" (Tulving, Schacter, and Stark 1982, pg.336).

To test how such subtle verbal cues might affect person perception, DeWall and Bushman (2009) conducted an experiment (Study 2) in which they investigated the relationship between exposure to words associated with hot and cold temperatures and the subsequent evaluation of a fictitious person. The 72 undergraduate students who participated in this experiment were first randomly assigned to one of three groups, in which they were primed with temperature-related or neutral words.

Their task consisted of creating grammatically correct sentences from five scrambled words. In the "heat prime" and "cold-prime" groups, six of the 13 sentences contained words associated with heat or cold, respectively. The "neutral prime" group's task did not include any temperature-associated words; therefore, it served as the "control group".

A "control" group is often used as a baseline in experiments, allowing researchers to see whether the changes observed in the experimental groups are due to the manipulation, and not to other factors. In this study, the control group was created to test whether exposure to "hot" or "cold" words influenced how participants judged the fictitious person, compared to a group with no temperature cues.

Thus, the priming condition, with the three levels (heat, cold, and neutral), was the independent variable (IV) in this experiment. Subsequently, all participants read a text about a fictitious man named Donald, whose behaviour was described in an ambiguous but potentially hostile manner. Participants were asked to rate Donald's personality in four questions related to hostility traits. The responses to these four questions were combined into an index of hostile perception, which served as the dependent variable (DV).

> #yourturn
>
> Why did the researchers measure the perception of Donald's personality after participants were primed with the concepts of heat or cold, compared to the neutral control group?

The underlying assumption is that priming can increase the accessibility of specific personality-related concepts or trait descriptions in memory, which in turn may shape how ambiguous information about others is interpreted (Srull and Wyer 1979). Additionally, theoretical models such as the General Aggression Model (Anderson and Anderson 1998) integrate the temperature–aggression hypothesis, proposing that hot temperatures can serve as situational inputs that activate aggression-related thoughts and feelings. In Study 2, DeWall and Bushman (2009) tested the more specific hypothesis that exposure to heat-related words would increase hostile perceptions of an ambiguously described person, compared to both neutral and cold-related words.

To compare the groups, the authors adopted the null hypothesis significance testing approach (NHST, Brandt et al. 2014; Cumming 2014; Wasserstein and Lazar 2016), by comparing the mean scores on the hostility index across the three priming conditions. In this approach, a result is considered statistically significant when the probability of observing a difference is sufficiently low, typically less than 5% ($p < .05$), assuming that there is actually no real difference between groups (the null hypothesis). As is typical in psychological research, the authors used this threshold to determine whether the differences between group means were statistically significant.

DeWall and Bushman (2009) results showed that the "heat prime" group rated

Donald as significantly more hostile than both the "neutral" and the "cold" groups (heat vs. cold: $d = .67$, $p < .03$; heat vs. neutral: $d = .63$, $p < .05$). Moreover, no significant differences were found between the "cold" and the "neutral" groups ($p = .85$).

These findings suggest that exposure to heat-related words increased participants' tendency to perceive ambiguous behavior as more hostile, supporting the hypothesis that temperature-related concepts can activate hostility-related trait perception. The absence of a statistical difference between the "cold" and "neutral" groups ($p > .05$) further indicates that this effect was specific to heat-related priming, rather than a general effect of temperature-related concepts.

8.2 2. The Aftermath

In 2014, McCarthy conducted two replication studies of this experiment (McCarthy 2014).

> #yourturn
>
> What criteria should a replication meet in order to be relevant and helpful for examining the effect?

There are different types of replication studies, each with different criteria and goals, although both aim to test the same theoretical claims. Close replications aim to verify whether the original effect can be found again under the same conditions as the original study, using the same method. In contrast, conceptual replications test the generalisability of an effect across contexts and may rely on different operational definitions or use a different method (Brandt et al. 2014).

By following Brandt et al.'s (2014) definition and guidelines for close replication, McCarthy (2014) designed two studies aiming to reproduce the original procedures but using larger samples. He justified these attempts to replicate with three arguments. First, a single study is not sufficient to establish the reliability of an effect, and further testing is necessary. Second, the original study had a relatively small sample size, which can lead to unstable effect size estimates; and larger samples are required. Third, the original findings had already been widely cited, so it is important to verify whether they could be replicated before treating them as reliable knowledge.

In McCarthy's (2014) first replication study, involving 182 participants, participants were randomly assigned to one of three priming conditions (heat, cold, or neutral) and completed the same scrambled sentence task as in the original. After the priming task, they read the same ambiguous story about a fictitious man and rated his hostility using the same four items to measure hostile perceptions. The second replication, conducted online with 507 participants, used the same critical heat- and cold-related words as in the original experiment, but the scrambled sentences in which these words appeared were slightly different from

the original materials. Otherwise, the procedure closely followed the original study.

The results of these two replication studies did not support the original hypotheses. Donald's rated hostility did not differ significantly between the heat and the cold prime groups. Thus, the findings reported by DeWall and Bushman (2009) could not be replicated. Additionally, McCarthy (2014) conducted a meta-analysis combining the original study with the two replications. This analysis also indicated a non-significant effect of heat priming on hostile perceptions ($d = 0.18$, $p < .05$). Based on these results, McCarthy (2014) concluded that priming individuals with heat-related concepts does not reliably affect hostile perceptions of others, and that the original effect is likely non-existent or too weak to be considered meaningful.

8.3 3. Conclusion

McCarthy (2014) tried to replicate DeWall and Bushman's (2009) study twice and found no evidence that heat-related words increase hostile perceptions. Their meta-analysis combining the original and replication studies also showed a non-significant effect, suggesting the original finding is likely non-existent or too weak to be relevant.

This replication failure reflects a broader debate in psychology that social priming effects may be fragile and difficult to reproduce, particularly when it comes to temperature-related words and their relation to hostility. Moreover, these studies evaluated hostile perceptions rather than aggressive behaviour. Therefore, the findings do not directly challenge broader theoretical models such as the General Aggression Model (Anderson and Anderson 1998), which integrates the temperature–aggression hypothesis through a variety of situational, cognitive, and affective mechanisms. What the replications do suggest is that simple word-based priming of hot and cold temperature is unlikely to be a reliable predictor of person perception in terms of hostility.

Chapter 9

Hot Coffee Effect

written by Aslı Ay Arat (original draft), and Aswathi Surendran (revision)

9.1 The Classic

Our environment often exerts strong influences on us. For instance, seasonal changes in sunlight hours can seriously affect mood and wellbeing (e.g., seasonal affective disorder), and people tend to be more willing to make donations in December (also referred to as the Christmas effect).

> #yourturn
>
> How does your current environment affect you? Take a moment to reflect!

Much psychological research is interested in how environments shape human behavior, our thinking and what we feel (often also considered in interaction with person-specific variables). One subfield has dedicated research on embodied cognition – the idea that bodily states influence what and how we think and feel (Chabris et al. 2019). The assumption is that the environment acts on the mind, via the body.

> #yourturn
>
> Based on the principles of embodied cognition, how might working in a cluttered or messy room affect someone's ability to concentrate or study?

In a 2008 study, researchers Williams and Bargh (2008) worked on a related question. They wanted to know if temperature – a salient feature of the environment if you think about how often conversations are centered on the weather – affected how people are perceived. They hypothesized that "physical warmth

should activate concepts or feelings of interpersonal warmth" (Williams and Bargh 2008, pg. 3).

> #definition Definition of psychological warmth
>
> The sense that another person is friendly, kind, and has good intentions. It is one of the two central dimensions we rely on when forming first impressions, the other being competence. People tend to notice warmth quickly and often use it as a basis for deciding whether someone is trustworthy. Importantly, research suggests that experiences of physical warmth can subtly shape these social judgements (Williams and Bargh 2008; Fiske, Cuddy, and Glick 2007).

The researchers asked undergraduate subjects to hold either a warm cup of coffee or iced coffee in their hands while writing down information. The expectation was that the concepts of warmth (or coldness) would be primed due to the physical experience of the temperature of the coffee, making it more likely that a person was correspondingly perceived as warm (or cold).

> #definition Definition of "priming"
>
> "A change in how easily we recognise or produce something because of an earlier encounter with it. In other words, our previous experience with an item can make us faster or more accurate in responding to it later, even if we are not consciously aware of the connection" Tulving, Schacter, and Stark (1982).

In the first part of the study, participants were undergraduate students at Yale University. They were asked to hold either a warm cup of coffee or an iced coffee while evaluating a fictional individual described in a personality profile. Those who held the warm beverage rated the individual as significantly more "interpersonally warm" compared to those who held the cold beverage (Williams and Bargh 2008). This result was interpreted as evidence that the feeling of physical warmth can unconsciously bring to mind the idea of social warmth. In other words, holding something warm made people more likely to see the person in the profile as kind and friendly.

In a follow-up study, participants were asked to hold either a heated or a cold therapeutic pad, under the impression that they were simply evaluating the product. Afterwards, they were given a choice of reward for taking part in the study. They could either select a gift for themselves (such as a drink voucher or an ice cream certificate) or choose the same type of gift for a friend. The results showed a clear pattern. Those who had held the warm pad were more likely to pick the gift for a friend, while those who had held the cold pad tended to choose the gift for themselves. This finding suggests that physical warmth does not just influence how we see other people, but can also affect our own behaviour, making us act in a more generous or prosocial way.

> #yourturn

> What do you think: How are physical warmth and prosocial giving related?

The researchers expected that physical warmth would lead to more generous behavior because our early life experiences often connect warmth with comfort, safety, and care from others. For example, being held close by a caregiver usually involves both physical warmth and feelings of trust and affection. Over time, these experiences create an unconscious link between physical warmth and social warmth.

In the study, participants who were in the warm condition were more likely to make the generous, prosocial choice of giving a gift to a friend, rather than keeping it for themselves. Together, these findings suggest that physical warmth can unconsciously activate ideas of social closeness and trust. This supports the broader idea that common expressions like calling someone "cold" or "warm" are not just figures of speech but may reflect real psychological processes.

From this we could conclude that physical warmth can lead to perceiving others as "warmer" people and it also makes us "warmer" and more generous.

9.2 The Aftermath

Given the striking and intuitive appeal of the original findings, they received significant media attention and were cited widely. However, as concerns about replicability in social psychology grew, so did scrutiny of the hot coffee study. Scientists emphasized that the hypothesis that hot coffee makes you generous is worth testing again, as the original had several methodological flaws. A major limitation of the original research was the small sample size. The two studies included only 41 and 53 participants, respectively. Small samples increase the risk that results reflect random variation rather than a genuine effect, which reduces the reliability of the conclusions. In other words, findings based on so few participants may not hold up when tested with larger groups.

A second issue was that the participants were not representative of the wider population. They were all undergraduate students from one university in New York State. College students often share similar age, education level, and cultural background, which means the findings might not apply to older adults, children, or people from other places and backgrounds.

> #definition Definition of representativity
>
> The extent to which a study sample reflects a well-defined target population, such that the estimates or the interpretation of results can be generalised to that population (Rudolph et al. 2023).

Replication studies attempted to address these limitations by recruiting larger samples, with more than three times the original number of participants, and by selecting more diverse populations. These methodological improvements

provided stronger statistical power and greater external validity, allowing researchers to test whether the effect was robust beyond the narrow conditions of the original experiments.

Multiple high-powered replication attempts have since failed to reproduce the original effects. In 2014, Lynott and colleagues conducted a multi-lab replication of the first Williams and Bargh experiment. A multi-lab replication is when several independent research groups carry out the same study using a common protocol. This approach reduces the likelihood that the outcome is due to local conditions or the influence of a single research team. Across a larger and more diverse sample, Lynott et al. (2014) found no evidence that holding a warm object influenced social judgments.

In 2018, Chabris et al. (2019) attempted to replicate the findings of Williams and Bargh (2008) using more rigorous methodology. Their studies addressed several shortcomings of the original. The first studies used very small samples (41 and 53 people), which makes results unstable and prone to influences of chance. Chabris et al. tested much larger groups, giving their findings more statistical power. The original participants were all college students, limiting generalisability; the replication recruited a more diverse public sample. The original studies also took place in a lab with experimenters aware of conditions, raising concerns about artificiality and bias. Chabris et al. tested participants in a natural field setting and used double-blind procedures. Despite these improvements, and contrary to the original claims, they found no evidence that holding a hot or cold object influenced participants' judgments or generosity. In other words, the replication showed no evidence that physical warmth affected behaviour or perceptions (Chabris et al. 2019).

> #yourturn
>
> What could be the cause for a differing result?

The replication researchers, however, do not conclude that hot coffee does not make people generous. Instead, because they found null effects, they concluded that there was no evidence for such an effect. Using a different statistical approach, they found that the evidence actually favored the interpretation that there was no effect, and not the interpretation that they might have missed detecting the original effect.

Some later research has suggested that warmth effects might still exist, but only under specific conditions. A study by Citron and Goldberg (2014) found that physical warmth increased perceptions of interpersonal kindness only in neutral social contexts. When participants read about someone behaving negatively, the warmth manipulation had no effect. This suggests that the influence of physical warmth on social judgment is not universal, but shaped by the surrounding social context.

9.3 Conclusion

The Hot Coffee study (Williams and Bargh 2008) sparked fascination with the idea that physical sensations shape social judgments. However, over a decade of follow-up research has largely failed to replicate these findings reliably. While the metaphor of warmth remains powerful in language and intuition, its psychological effects appear to be fragile, context-sensitive, and not easily reproduced under stricter experimental controls.

This case illustrates an important lesson in psychological science: even intuitively satisfying findings must be rigorously tested, replicated, and interpreted within a broader theoretical and methodological context. The story of this study also reflects a broader shift in psychology: moving away from surprising, single-study findings toward replication, cumulative evidence, and methodological transparency.

> #yourturn
>
> Which study seems more convincing to you? Why?

More broadly, this debate encourages us to reflect on the role of the environment in shaping human behaviour. Findings like the Hot Coffee study suggest that seemingly minor physical cues might influence judgments and actions, but the difficulty in replicating these effects shows that such influences are neither simple nor uniform. Environmental factors may interact with individual differences, situational contexts, and cultural expectations in ways that make their effects less predictable than early studies implied. Our thoughts and actions could be shaped in subtle ways by the contexts we are in. The challenge for psychology is to determine which of these effects are robust, meaningful, and practically relevant, and which are not.

Summary

Add here

Take-Aways

Add here

Thanks

This book was made possible by the many helping hands and critical thoughts of the student authors involved in writing the individual chapters. In addition, Melissa Engelbarth's support with selecting and translating the chapters to include was invaluable.

References

Allport, Gordon W. 1954. "The Nature of Prejudice," The nature of prejudice, xviii, 537.

Anderson, C. A., and K. B. Anderson. 1998. "Temperature and Aggression: Paradox, Controversy, and a (Fairly) Clear Picture." In *Human Aggression: Theories, Research and Implications for Policy*, edited by R. Geen and E. Donnerstein, 247–98. Academic Press.

Anderson-Hanley, C., P. Arciero, and A. Snyder. 2011. "Social Facilitation in Virtual Reality-Enhanced Exercise: Competitiveness Moderates Exercise Effort of Older Adults." *Clinical Interventions in Aging* 6: 275–80. https://doi.org/10.2147/CIA.S25337.

"APA Dictionary of Psychology." 2018. https://dictionary.apa.org/. https://dictionary.apa.org/association.

Aronson, E., T. D. Wilson, and R. M. Akert. 2005. "Social Psychology."

Baker, Paul E. 1934. "Negro-White Adjustment in America." *The Journal of Negro Education* 3 (2): 194. https://doi.org/10.2307/2292313.

Baron, R. S. 1986. "Distraction-Conflict Theory: Progress and Problems." In *Advances in Experimental Social Psychology*, 19:1–40. Academic Press. https://doi.org/10.1016/S0065-2601(08)60211-7.

Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice. 1998. "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology* 74 (5): 1252–65. https://doi.org/10.1037/0022-3514.74.5.1252.

Blázquez, Desirée, Juan Botella, and Manuel Suero. 2017. "The Debate on the Ego-Depletion Effect: Evidence from Meta-Analysis with the p-Uniform Method." *Frontiers in Psychology* 8 (February). https://doi.org/10.3389/fpsyg.2017.00197.

Bond, C. F., and L. J. Titus. 1983. "Social Facilitation: A Meta-Analysis of 241 Studies." *Psychological Bulletin* 94 (2): 265–92. https://doi.org/10.1037/0033-2909.94.2.265.

Brameld, Theodore. 1946. *Minority Problems in the Public Schools: A Study of Administrative Policies and Practices in Seven School Systems*. Harper & Brothers.

Brandt, Mark J., Hans IJzerman, Ap Dijksterhuis, Frank J. Farach, Jason Geller, Roger Giner-Sorolla, James A. Grange, Marco Perugini, Jeffrey R. Spies,

and Anna Van 'T Veer. 2014. "The Replication Recipe: What Makes for a Convincing Replication?" *Journal of Experimental Social Psychology* 50 (January): 217–24. https://doi.org/10.1016/j.jesp.2013.10.005.

Brophy, Ira N. 1945. "The Luxury of Anti-Negro Prejudice." *Public Opinion Quarterly* 9 (4): 456–66. https://doi.org/10.1086/265762.

Brophy, J. E., and T. L. Good. 1970. "Teachers' Communication of Differential Expectations for Children's Classroom Performance: Some Behavioral Data." *Journal of Educational Psychology* 61: 365–74. https://doi.org/10.1037/h0029908.

Bunker, Cameron J., and Michael E. W. Varnum. 2021. "How Strong Is the Association Between Social Media Use and False Consensus?" *Computers in Human Behavior* 125 (December): 106947. https://doi.org/10.1016/j.chb.2021.106947.

Cameron, D. L., and B. G. Cook. 2013. "General Education Teachers' Goals and Expectations for Their Included Students with Mild and Severe Disabilities." *Education and Training in Autism and Developmental Disabilities* 48: 18–30. http://www.jstor.org/stable/23879883.

Carter, Evan C., and Michael E. McCullough. 2014. "Publication Bias and the Limited Strength Model of Self-Control: Has the Evidence for Ego Depletion Been Overestimated?" *Frontiers in Psychology* 5 (July). https://doi.org/10.3389/fpsyg.2014.00823.

Castro, J. M. de. 1994. "Family and Friends Produce Greater Social Facilitation of Food Intake Than Other Companions." *Physiology & Behavior* 56 (3): 445–55. https://doi.org/10.1016/0031-9384(94)90286-0.

Chabris, Christopher F., Patrick R. Heck, Jaclyn Mandart, Daniel J. Benjamin, and Daniel J. Simons. 2019. "No Evidence That Experiencing Physical Warmth Promotes Interpersonal Warmth: Two Failures to Replicate." *Social Psychology* 50 (2): 127–32. https://doi.org/10.1027/1864-9335/a000361.

Chambers, Chris. 2019. "The Registered Reports Revolution Lessons in Cultural Reform." *Significance* 16 (4): 23–27.

Christensen, Garret, Zenan Wang, Elizabeth Levy Paluck, Nicholas Swanson, David Birke, Edward Miguel, and Rebecca Littman. 2020. "Open Science Practices Are on the Rise: The State of Social Science (3S) Survey."

Citron, Francesca M. M., and Adele E. Goldberg. 2014. "Social Context Modulates the Effect of Physical Warmth on Perceived Interpersonal Kindness: A Study of Embodied Metaphors." *Language and Cognition* 6 (1): 1–11. https://doi.org/10.1017/langcog.2013.4.

Coles, Nicholas A., David S. March, Fernando Marmolejo-Ramos, Jeff T. Larsen, Nwadiogo C. Arinze, Izuchukwu L. G. Ndukaihe, Megan L. Willis, et al. 2022. "A Multi-Lab Test of the Facial Feedback Hypothesis by the Many Smiles Collaboration." *Nature Human Behaviour* 6 (12): 1731–42. https://doi.org/10.1038/s41562-022-01458-9.

Cottrell, N. B. 1972. "Social Facilitation." In *Experimental Social Psychology*, edited by C. G. McClintock. Holt, Rinehart & Winston.

Cottrell, N. B., D. L. Wack, G. J. Sekerak, and R. H. Rittle. 1968. "Social Facilitation of Dominant Responses by the Presence of an Audience and the

Mere Presence of Others." *Journal of Personality and Social Psychology* 9 (3): 245–50. https://doi.org/10.1037/h0025902.

Cumming, Geoff. 2014. "The New Statistics: Why and How." *Psychological Science* 25 (1): 7–29. https://doi.org/10.1177/0956797613504966.

Dang, Junhua. 2017. "An Updated Meta-Analysis of the Ego Depletion Effect." *Psychological Research* 82 (4): 645–51. https://doi.org/10.1007/s00426-017-0862-x.

Darwin, Charles. 1872. *The Expression of the Emotions in Man and Animals*. The Expression of the Emotions in Man and Animals. London, England: John Murray. https://doi.org/10.1037/10001-000.

Dashiell, J. F. 1930. "An Experimental Analysis of Some Group Effects." *The Journal of Abnormal and Social Psychology* 25 (2): 190–99. https://doi.org/10.1037/h0075144.

De Boer, H., R. J. Bosker, and M. P. C. Van der Werf. 2010. "Sustainability of Teacher Expectation Bias Effects on Long-Term Student Performance." *Journal of Educational Psychology* 102: 168–79. https://doi.org/10.1037/a0017289.

Deutsch, Morton, and Mary Evans Collins. 1951. *Interracial Housing; a Psychological Evaluation of a Social Experiment*. Interracial Housing; a Psychological Evaluation of a Social Experiment. Minneapolis, MN, US: University of Minnesota Press.

DeWall, Nathan C., and Brad J. Bushman. 2009. "Hot Under the Collar in a Lukewarm Environment: Words Associated with Hot Temperature Increase Aggressive Thoughts and Hostile Perceptions." *Journal of Experimental Social Psychology* 45 (4): 1045–47. https://doi.org/10.1016/j.jesp.2009.05.003.

Eden, D. 1990. *Pygmalion in Management: Productivity as a Self-Fulfilling Prophecy*.

Ekman, Paul. 1992. "An Argument for Basic Emotions." *Cognition and Emotion* 6 (3-4): 169–200. https://doi.org/10.1080/02699939208411068.

Festinger, L. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7 (2): 117–40.

Fiske, Susan T., Amy J. C. Cuddy, and Peter Glick. 2007. "Universal Dimensions of Social Cognition: Warmth and Competence." *Trends in Cognitive Sciences* 11 (2): 77–83. https://doi.org/10.1016/j.tics.2006.11.005.

Friese, Malte, David D. Loschelder, Karolin Gieseler, Julius Frankenbach, and Michael Inzlicht. 2018. "Is Ego Depletion Real? An Analysis of Arguments." *Personality and Social Psychology Review* 23 (2): 107–31. https://doi.org/10.1177/1088868318762183.

Furnas, Alexander C., and Timothy M. LaPira. 2024. "The People Think What I Think: False Consensus and Unelected Elite Misperception of Public Opinion." *American Journal of Political Science* 68 (3): 958–71. https://doi.org/10.1111/ajps.12833.

Greenwald, Anthony G., Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74 (6): 1464–80. https://doi.org/10.1037/0022-3514.74.6.1464.

Greenwald, Anthony G., Brian A. Nosek, and Mahzarin R. Banaji. 2003. "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm." *Journal of Personality and Social Psychology* 85 (2): 197–216. https://doi.org/10.1037/0022-3514.85.2.197.

Greenwald, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji. 2009. "Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity." *Journal of Personality and Social Psychology* 97 (1): 17–41. https://doi.org/10.1037/a0015575.

Hagger, Martin S., Chantelle Wood, Chris Stiff, and Nikos L. D. Chatzisarantis. 2010. "Ego Depletion and the Strength Model of Self-Control: A Meta-Analysis." *Psychological Bulletin* 136 (4): 495–525. https://doi.org/10.1037/a0019486.

Halfmann, E., J. Bredehöft, and J. A. Häusser. 2020. "Replicating Roaches: A Preregistered Direct Replication of Zajonc, Heingartner, and Herman's (1969) Social-Facilitation Study." *Psychological Science* 31 (3): 332–37. https://doi.org/10.1177/0956797620902101.

Hull, C. L. 1943. *Principles of Behavior*. Appleton-Century Crofts.

Inzlicht, Michael, Will Gervais, and Elliot Berkman. 2015. "Bias-Correction Techniques Alone Cannot Determine Whether Ego Depletion Is Different from Zero: Commentary on Carter, Kofler, Forster, & Mccullough, 2015." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2659409.

Izuma, K., D. N. Saito, and N. Sadato. 2010. "Processing of the Incentive for Social Approval in the Ventral Striatum During Charitable Donation." *Journal of Cognitive Neuroscience* 22 (4): 621–31. https://doi.org/10.1162/jocn.2009.21228.

James, William. 1884. "What Is an Emotion?" *Mind* os-IX (34): 188–205. https://doi.org/10.1093/mind/os-IX.34.188.

Jenner, H. 1990. "The Pygmalion Effect: The Importance of Expectancies." *Alcoholism Treatment Quarterly* 7 (2): 127–33.

Jhangiani, Dr Rajiv, and Dr Hammond Tarry. 2022. "11.1 Social Categorization and Stereotyping." *Principles of Social Psychology - 1st International H5P Edition*, January.

Jussim, L., and K. D. Harber. 2005. "Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies." *Personality and Social Psychology Review* 9: 131–55. https://doi.org/10.1207/s15327957pspr0902_3.

Kephart, William M. 1957. *Racial Factors and Urban Law Enforcement*. University of Pennsylvania Press. https://www.jstor.org/stable/j.ctv4w3vv0.

Kidwell, Mallory C, Ljiljana B Lazarević, Erica Baranski, Tom E Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, et al. 2016. "Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency." *PLoS Biology* 14 (5): e1002456.

Kuhn, Thomas. 1962. "The Structure of Scientific Revolutions." *International Encyclopedia of Unified Science* 2 (2).

Learman, L. A., J. Avorn, D. E. Everitt, and R. Rosenthal. 1990. "Pygmalion in the Nursing Home: The Effects of Caregiver Expectations on Patient

Outcomes." *Journal of the American Geriatrics Society* 38 (7): 797–803.

Lett, Harold A. 1945. *Techniques for Achieving Interracial Cooperation*. http://archive.org/details/Lett070.

Lowe, Matt. 2025. "Has Intergroup Contact Delivered?" *Annual Review of Economics* 17 (Volume 17, 2025): 321–44. https://doi.org/10.1146/annurev-economics-081324-091109.

Luzsa, Robert, and Susanne Mayr. 2021. "False Consensus in the Echo Chamber: Exposure to Favorably Biased Social Media News Feeds Leads to Increased Perception of Public Support for Own Opinions." *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 15 (1). https://doi.org/10.5817/CP2021-1-3.

Lynott, Dermot, Katherine S. Corker, Jessica Wortman, Louise Connell, M. Brent Donnellan, Richard E. Lucas, and Kerry O'Brien. 2014. "Replication of 'Experiencing Physical Warmth Promotes Interpersonal Warmth' By." *Social Psychology* 45 (3): 216–22. https://doi.org/10.1027/1864-9335/a000187.

Madon, S., L. Jussim, and J. Eccles. 1997. "In Search of the Powerful Self-Fulfilling Prophecy." *Journal of Personality and Social Psychology* 72: 791–809. https://doi.org/10.1037/0022-3514.72.4.791.

McCarthy, Randy J. 2014. "Close Replication Attempts of the Heat Priming-Hostile Perception Effect." *Journal of Experimental Social Psychology* 54 (September): 165–69. https://doi.org/10.1016/j.jesp.2014.04.014.

McHugh, C., S. M. Griffin, E. L. Kinsella, M. Quayle, B. Strunz, and T. Muldoon Orla. 2025. "A Replication of Triplett's 'Social Facilitation Experiment'." *Scientific Reports* 15. https://doi.org/10.1038/s41598-025-25608-x.

Mousa, Salma. 2020. "Building Social Cohesion Between Christians and Muslims Through Soccer in Post-ISIS Iraq." *Science* 369 (6505): 866–70. https://doi.org/10.1126/science.abb3153.

Mullen, Brian, Jennifer L Atkins, Debbie S Champion, Cecelia Edwards, Dana Hardy, John E Story, and Mary Vanderklok. 1985. "The False Consensus Effect: A Meta-Analysis of 115 Hypothesis Tests." *Journal of Experimental Social Psychology* 21 (3): 262–83. https://doi.org/10.1016/0022-1031(85)90020-4.

Muller, D., and F. Butera. 2007. "The Focusing Effect of Self-Evaluation Threat in Coaction and Social Comparison." *Journal of Personality and Social Psychology* 93 (2): 194–211. https://doi.org/10.1037/0022-3514.93.2.194.

Munafò, Marcus R, Brian A Nosek, Dorothy VM Bishop, Katherine S Button, Christopher D Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J Ware, and John Ioannidis. 2017. "A Manifesto for Reproducible Science." *Nature Human Behaviour* 1 (1): 1–9.

Noah, Tom, Yaacov Schul, and Ruth Mayo. 2018. "When Both the Original Study and Its Failed Replication Are Correct: Feeling Observed Eliminates the Facial-Feedback Effect." *Journal of Personality and Social Psychology* 114 (5): 657–64. https://doi.org/10.1037/pspa0000121.

Nosek, Brian A., and Frederick L. Smyth. 2007. "A Multitrait-Multimethod Validation of the Implicit Association Test." *Experimental Psychology* 54 (1):

14–29. https://doi.org/10.1027/1618-3169.54.1.14.

Paluck, Elizabeth Levy, Seth A. Green, and Donald P. Green. 2019. "The Contact Hypothesis Re-Evaluated." *Behavioural Public Policy* 3 (2): 129–58. https://doi.org/10.1017/bpp.2018.25.

Paluck, Elizabeth Levy, Roni Porat, Chelsey S. Clark, and Donald P. Green. 2021. "Prejudice Reduction: Progress and Challenges." *Annual Review of Psychology* 72 (Volume 72, 2021): 533–60. https://doi.org/10.1146/annurev-psych-071620-030619.

Paolini, Stefania, Meghann Gibbs, Brett Sales, Danielle Anderson, and Kylie McIntyre. 2024. "Negativity Bias in Intergroup Contact: Meta-Analytical Evidence That Bad Is Stronger Than Good, Especially When People Have the Opportunity and Motivation to Opt Out of Contact." *Psychological Bulletin* 150 (8): 921–64. https://doi.org/10.1037/bul0000439.

Pettigrew, Thomas F., and Linda R. Tropp. 2006. "A Meta-Analytic Test of Intergroup Contact Theory." *Journal of Personality and Social Psychology* 90 (5): 751–83. https://doi.org/10.1037/0022-3514.90.5.751.

Popper, Karl R. 1959. *The Logic of Scientific Discovery*. Hutchinson & Co.

Raudenbush, S. W. 1984. "Magnitude of Teacher Expectancy Effects on Pupil IQ as a Function of the Credibility of Expectancy Induction: A Synthesis of Findings from 18 Experiments." *Journal of Educational Psychology* 76: 85–97. https://doi.org/10.1037/0022-0663.76.1.85.

Rosenbloom, T., A. Shahar, A. Perlman, D. Estreich, and E. Kirzner. 2007. "Success on a Practical Driver's License Test with and Without the Presence of Another Testee." *Accident Analysis & Prevention* 39 (6): 1296–1301. https://doi.org/10.1016/j.aap.2007.03.015.

Rosenthal, R., and L. Jacobson. 1968. "Pygmalion in the Classroom." *The Urban Review* 3 (1): 16–20. https://doi.org/10.1007/bf02322211.

Rosenthal, R., and D. B. Rubin. 1978. "Interpersonal Expectancy Effects: The First 345 Studies." *Behavioral and Brain Sciences* 1: 377–86. https://doi.org/10.1017/S0140525X00075506.

Ross, Lee, David Greene, and Pamela House. 1977. "The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes." *Journal of Experimental Social Psychology* 13 (3): 279–301. https://doi.org/10.1016/0022-1031(77)90049-X.

Rubie-Davies, C. M. 2007. "Classroom Interactions: Exploring the Practices of High- and Low-Expectation Teachers." *British Journal of Educational Psychology* 77: 289–306. https://doi.org/10.1348/000709906X101601.

Rubie-Davies, Christine M., Elizabeth R. Peterson, Chris G. Sibley, and Robert Rosenthal. 2015. "A Teacher Expectation Intervention: Modelling the Practices of High Expectation Teachers." *Contemporary Educational Psychology*, Examining Innovations—Navigating the Dynamic Complexities of School-Based Intervention Research, 40 (January): 72–85. https://doi.org/10.1016/j.cedpsych.2014.03.003.

Rudolph, Jacqueline E., Yongqi Zhong, Priya Duggal, Shruti H. Mehta, and Bryan Lau. 2023. "Defining Representativeness of Study Samples in Medical and Population Health Research." *BMJ Medicine* 2 (1). https://doi.org/10

.1136/bmjmed-2022-000399.

Sanders, G. S., R. S. Baron, and D. L. Moore. 1978. "Distraction and Social Comparison as Mediators of Social Facilitation Effects." *Journal of Experimental Social Psychology* 14 (3): 291–303. https://doi.org/10.1016/0022-1031(78)90017-3.

Schimmack, Ulrich. 2021. "The Implicit Association Test: A Method in Search of a Construct." *Perspectives on Psychological Science* 16 (2): 396–414. https://doi.org/10.1177/1745691619863798.

Seitchik, A. E., A. J. Brown, and S. G. Harkins. 2017. "Social Facilitation: Using the Molecular to Inform the Molar." In *The Oxford Handbook of Social Influence*, 183–203. Oxford University Press.

Snow, R. E. 1969. "Unfinished Pygmalion [Review of the Book Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development, by r. Rosenthal & l. Jacobson]." *Contemporary Psychology* 14: 197–200. https://doi.org/10.1037/0010293.

Sommer, R., M. Wynes, and G. Brinkley. 1992. "Social Facilitation Effects in Shopping Behavior." *Environment and Behavior* 24 (3): 285–97. https://doi.org/10.1177/0013916592243001.

Spence, K. W. 1956. *Behavior Theory and Conditioning*. Yale University Press. https://doi.org/10.1037/10029-000.

Srull, Thomas K., and Robert S. Wyer. 1979. "The Role of Category Accessibility in the Interpretation of Information about Persons: Some Determinants and Implications." *Journal of Personality and Social Psychology* 37 (10): 1660–72. https://doi.org/10.1037/0022-3514.37.10.1660.

Steiner, Nils D., Claudia Landwehr, and Philipp Harms. 2025. "False Consensus Beliefs and Populist Attitudes." *Political Psychology* 00: 1–22. https://doi.org/10.1111/pops.70026.

Strack, Fritz. 2016. "Reflection on the Smiling Registered Replication Report." *Perspectives on Psychological Science* 11 (6): 929–30. https://doi.org/10.1177/1745691616674460.

Strack, Fritz, Leonard L. Martin, and Sabine Stepper. 1988. "Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis." *Journal of Personality and Social Psychology* 54 (5): 768–77. https://doi.org/10.1037/0022-3514.54.5.768.

Strube, M. J. 2005. "What Did Triplett Really Find? A Contemporary Analysis of the First Experiment in Social Psychology." *The American Journal of Psychology* 118 (2): 271–86.

Thorndike, R. L. 1968. "[Review of the Book Pygmalion in the Classroom, by r. Rosenthal & l. Jacobson]." *American Educational Research Journal* 5: 708–11. https://doi.org/10.3102/00028312005004708.

Triplett, N. 1898. "The Dynamogenic Factors in Pacemaking and Competition." *The American Journal of Psychology* 9 (4): 507–33. https://doi.org/10.2307/1412188.

Tulving, Endel, Daniel L. Schacter, and Heather A. Stark. 1982. "Priming Effects in Word-Fragment Completion Are Independent of Recognition Memory." *Journal of Experimental Psychology: Learning, Memory, and Cogni-*

tion 8 (4): 336–42. https://doi.org/10.1037/0278-7393.8.4.336.

Vazire, Simine. 2018. "Implications of the Credibility Revolution for Productivity, Creativity, and Progress." *Perspectives on Psychological Science* 13 (4): 411–17. https://doi.org/10.1177/1745691617751884.

Vazire, Simine, Sarah R Schiavone, and Julia G Bottesini. 2022. "Credibility Beyond Replicability: Improving the Four Validities in Psychological Science." *Current Directions in Psychological Science* 31 (2): 162–68.

Vohs, Kathleen D., Brandon J. Schmeichel, Sophie Lohmann, Quentin F. Gronau, Anna J. Finley, Sarah E. Ainsworth, Jessica L. Alquist, et al. 2021. "A Multisite Preregistered Paradigmatic Test of the Ego-Depletion Effect." *Psychological Science* 32 (10): 1566–81. https://doi.org/10.1177/0956797621989733.

Wagenmakers, E.-J., T. Beek, L. Dijkhoff, Q. F. Gronau, A. Acosta, R. B. Adams, D. N. Albohn, et al. 2016. "Registered Replication Report: Strack, Martin, & Stepper (1988)." *Perspectives on Psychological Science* 11 (6): 917–28. https://doi.org/10.1177/1745691616674458.

Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA Statement on p-Values: Context, Process, and Purpose." *The American Statistician* 70 (2): 129–33. https://doi.org/10.1080/00031305.2016.1154108.

Weinschenk, Aaron C., Costas Panagopoulos, and Sander van der Linden. 2021. "Democratic Norms, Social Projection, and False Consensus in the 2020 U.S. Presidential Election." *Journal of Political Marketing* 20 (3-4): 255–68. https://doi.org/10.1080/15377857.2021.1939568.

Weiss, R. F., and F. G. Miller. 1971. "The Drive Theory of Social Facilitation." *Psychological Review* 78 (1): 44–57. https://doi.org/10.1037/h0030386.

Williams, Lawrence E., and John A. Bargh. 2008. "Experiencing Physical Warmth Promotes Interpersonal Warmth." *Science* 322 (5901): 606–7. https://doi.org/10.1126/science.1162548.

Zajonc, R. B. 1965. "Social Facilitation." *Science* 149 (3681): 269–74. https://doi.org/10.1126/science.149.3681.269.

Glossary

Availability Heuristic

A mental shortcut where people estimate the likelihood of an event based on how easily examples come to mind, which can lead to overestimating rare but memorable occurrences.

Attitude

The cognition, affect and behavioral tendencies towards a certain object.

Bias

A systematic distortion of perception or judgment.

Cherry-Picking

Reporting only the data, outcomes, or time frames that support one's hypothesis while ignoring or dismissing those that do not. This makes the story or articles simpler and might make them more publishable, but provides a distorted view of the evidence.

Conceptual Replication

A study that aims to recreate the gist of a prior study without using an identical procedure. These studies often aim to explore boundary conditions, the influence of specific variables, or aim to broaden and extend a certain finding.

Contact Interventions (Chapter 12, Chapter 20)

Carefully tailored interventions that apply intergroup contact in real-world settings to try and reduce prejudice among social groups.

Construct Validity

The extent to which a test measures the theoretical construct or concept it is intended to measure.

Discriminant Validity

The extent to which a test is unrelated to measures designed to assess theoretically distinct constructs.

Dynamogenesis

An increase in the mental or motor activity of an already functioning bodily system that accompanies any added sensory stimulation.

Effect Size

A quantitative measure of the magnitude of a phenomenon, used to assess the practical significance of research findings.

Editorial

An introductory article written by the editors of a special issue in an academic journal. It outlines the purpose, scope, and significance of the special issue, provides an overview of the included articles, and often highlights key themes, trends, or gaps in the research field.

Ego Depletion

A concept that describes willpower as a limited resource that can be used up (depleted).

Experiment (Chapter12, Chapter 20)

A study where researchers deliberately manipulate one or more variables and randomly assign participants to different conditions. Random assignment helps ensure the groups are similar before the intervention, so differences in outcomes are more likely to be caused by the manipulation rather than by pre-existing differences.

False Consensus Effect

A cognitive bias where individuals overestimate the extent to which others share their beliefs, preferences, and behaviors.

Generalized Drive

The presence of others leads to an increase in generalized drive, thus facilitating habitualized dominant responses.

Implicit Association Test

A reaction-time task that measures the strength of automatic associations between concepts (e.g., flowers and positivity) by comparing how quickly people

classify paired categories. Faster responses indicate stronger underlying associations.

Implicit Attitude

An enduring mental disposition toward something that is not consciously identified and of which a person may lack awareness.

Implicit Social Cognition

The automatic, unconscious mental processes that influence how we perceive, evaluate, and interact with others.

Incongruent Association

A mental relationship between two objects or concepts characterized by lack of harmony or misalignment.

Intergroup Bias (Chapter 12, Chapter 20)

Tendency to favour one's own social group (ingroup) over other groups (outgroups), which often leads to negative attitudes or behaviours toward outgroup members.

Meta-analysis (Chapter 12, Chapter 20)

A statistical technique that combines the results of multiple independent studies to estimate an overall effect. Meta-analyses can reveal patterns across a large body of research, but the quality of their conclusions depends on the quality and comparability of the included studies.

Moderator

A variable that modifies the relationship between independent and dependent variables.

Multi-Lab Study

A research project in which researchers working at several different locations (laboratories) implement the same experimental design and then analyse the data together.

Multimethod Study

Research that employs two or more distinct methods.

Observational Research

A study design where researchers measure variables as they naturally occur, without manipulating them. Observational studies can reveal associations between variables but cannot, on their own, establish that one causes the other.

Pygmalion Effect

The phenomenon in which higher expectations from others lead to improved performance.

Prejudice

A negative attitude toward a group and its members, often based on stereotypes rather than direct experience.

Priming

A psychological phenomenon where exposure to one stimulus (e.g., a word, image, or idea) influences how you respond to a later stimulus, often without conscious awareness.

Representativity

The extent to which a study sample reflects a well-defined target population, such that the estimates or the interpretation of results can be generalised to that population (Rudolph, J. E., 2023).

Social Comparison Theory

According to the social comparison theory, people are motivated to assess their own beliefs and skills by comparing them to external images. These images can be comparisons to other people or a reference to physical reality. Individuals have a tendency to view images portrayed by others as accessible and realistic and subsequently make comparisons between themselves, other people, and these idealized images.

Social Desirability

The tendency to want to be viewed positively by others, often by aligning with socially approved behaviors and attitudes.

Socially Desirable Responding

The act of providing inauthentic responses to better present oneself favorably according to current social norms.

Special Issue

A collection of articles on a specific topic, typically published together in a single issue of an academic journal. Special issues are often edited by guest editors and aim to provide a comprehensive exploration of the chosen theme or field of study.

Treatment and Control Conditions (Chapter 12, Chapter 20)

In experimental design, treatment condition refers to the participants who are randomly chosen to undergo the intervention (e.g., to play in the mixed soccer team). Control condition refers to the participants who are subject to intervention-like treatment that lacks the critical aspect of the intervention (e.g., those allocated to play in the all-Christian soccer team). Here, the critical aspect is intergroup contact within the team.

Orbicularis Oris Muscles

These are circular muscles around the mouth that close the lips and produce puckering, as in kissing or whistling.

Zygomaticus Major Muscles

These bilateral facial muscles, when activated, raise the corners of the mouth in an upward and lateral direction, facilitating expressions such as smiling.