

How Can People Commit Atrocities When They Follow Orders?

From qualitative interviews with former
genocide perpetrators to neuroscience research

Emilie A. Caspar

Moral & Social Brain Lab, Department of
Experimental Psychology, Ghent University (Belgium)

Résumé : La capacité des êtres humains à obéir aux ordres, même si ceux-ci
sont jugés atroces, n'a plus à être prouvée. Comme l'a souligné Howard Zinn,
« D'un point de vue historique, les événements les plus terribles – tels que la
guerre, les génocides et l'esclavage – ont résulté non pas de la désobéissance,
mais de l'obéissance » (Zinn, 1997). Mais la question de savoir « comment » les
individus peuvent commettre des atrocités lorsqu'ils suivent des ordres reste
ouverte. Les études de Milgram sont célèbres pour avoir montré une forte
soumission aux ordres de l'expérimentateur. Mais ces études n'ont révélé que
les facteurs situationnels et sociaux favorisant l'obéissance. Elles n'ont pas
permis de comprendre comment tant de personnes ont accepté d'envoyer des
chocs douloureux et potentiellement mortels à une autre personne. Dans cet
article, j'opposerai les entretiens qualitatifs réalisés avec d'anciens génocidaires
au Rwanda et la recherche expérimentale en neurosciences pour tenter de
faire émerger un élément de réponse à cette question critique. Je soutiendrai
que le dépassement des frontières entre différentes disciplines scientifiques est
la clé d'une meilleure compréhension de la façon dont l'obéissance modifie
la prosocialité. Je défendrai également l'idée provocatrice que les chercheurs
devraient saisir l'opportunité de quitter leur laboratoire pour étudier la
question de l'obéissance.

Abstract: The capacity for humans to obey orders, even atrocious ones,
no longer needs to be proven. As Howard Zinn famously pointed out,
“Historically, the most terrible things—war, genocide, and slavery—have
resulted not from disobedience, but from obedience” (Zinn, 1997). However,
the question of “how” people can commit atrocities when they follow orders
remains open. Milgram’s studies famously revealed a strong compliance

with the orders of the experimenter, but such studies only revealed the situational and social factors supporting obedience. They did not allow us to understand how so many people accepted to administer painful and potentially deadly shocks to another person. In the present article, I will contrast qualitative interviews conducted with former genocide perpetrators in Rwanda and experimental research in neuroscience to attempt to provide a partial answer to this critical question. I will argue that transcending the boundaries of different scientific disciplines is key to better understanding how obedience alters prosociality. I will also defend the provocative idea that researchers should seize the opportunity to leave their laboratories to study the issue of obedience.

It was late in the morning on a day in August 2021. Our small team had just arrived in a small rural village in Rwanda after a two-hour drive on a crowded road filled with trucks. We had come to meet François, a volunteer working for Prison Fellowship Rwanda, a local association dedicated to working with former genocide perpetrators and survivors to promote reconciliation. We brought along our portable electroencephalograms and two large boxes containing electronic devices, including laptops, electrodes, keyboards, screens, and more. We set up our equipment in the village church, as it was the only place with electrical outlets, which were essential for our machines to function. This was quite an unusual scene in this rural village, comprised of mud houses with straw roofs.

After setting up our equipment, I took a short walk with François, who explained how he had escaped the genocide by fleeing to Burundi with his family just before the massacres began. During our conversation, we suddenly crossed paths with an elderly man. I barely glanced at him, but he appeared peaceful, with a slow walk and a cane for support. I remember that he was dressed rather neatly, wearing a shirt with a pullover. François abruptly interrupted his narrative and said to me, “You see this man? He killed 13 people during the genocide.”

It was the first time I was about to interview former genocide perpetrators, and I was in this village to meet them. However, the casualness with which that sentence was spoken was disconcerting. But in many locations in Rwanda, this is to some extent common. It is estimated that about 120,000 men, mostly civilians, participated in the genocide, and many of them are still alive and have now been released from prisons [Clark 2005].

1 From neuroscience to qualitative interviews, and the other way around

By training, I am a neuropsychologist. In this regard, I surely never predicted that I would find myself conducting qualitative interviews with former genocide perpetrators in Rwanda and Cambodia at some point in my career. I initially approached the question of obedience in a rather conventional manner by familiarizing myself with Milgram's work and the work of his contemporaries. However, I felt that something was strongly missing to get a broader understanding of how obeying orders blurs morality and allows to strongly harm another person.

Milgram's work was crucial to establish with an experimental approach the situational factors supporting obedience [Milgram 1974]. For instance, it has been shown that obedience decreases if the experimenter is not physically present in the room or if two experimenters provide opposing views regarding the morality of the experiment. Milgram's studies were complemented by qualitative and quantitative data from people who refused ongoing orders during genocides and risked their lives. They explained that biographical availability, socialization, and situational contexts were critical for making rescue possible [Fox & Nyseth Brehm 2018]. However, such studies were not sufficient to explain how some people comply with orders to harm others, as they did not offer detailed explanations about the mechanisms that alter prosociality under authoritative pressure.

Milgram introduced the "agentic state theory," which posits that people obey authority figures even when causing harm to others by shifting from an autonomous state to an agentic state, where they view themselves as instruments following orders [Milgram 1974]. Some of Milgram's participants' reports seemed to support this theory, as they attributed their actions to external authority. However, concerns arose about its validity, and it lacked systematic experimental validation [Haslam, Reicher *et al.* 2015a]. Additionally, it was stated that this theory could not explain variations in obedience across Milgram's different experiments [Haslam, Reicher *et al.* 2015a, Haslam & Reicher 2017]. Later, a team of researchers proposed the "engaged followership model," building on the social identity theory [Tajfel & Turner 1979], emphasizing participants' identification with the experimenter and their mission [Haslam, Reicher *et al.* 2015a]. In different studies, the authors observed a correlation between identification and obedience, but they were limited to imagined scenarios or virtual environments [Haslam, Reicher *et al.* 2014, 2015b]. However, other research involving real behaviors has yielded mixed results, with some supporting the theory and others suggesting other factors influencing obedience [Caspar 2021, Caspar, Gishoma *et al.* 2022]. In summary, while identification with an authority figure may increase obedience, the engaged followership model appears insufficient alone to fully explain the widespread obedience to harmful orders.

In complement to previous experimental research, my aim was to better understand the mechanisms that explain how people’s morality and prosociality can change under the influence of obedience. Leveraging my expertise in neuropsychology and neuroscience, I wanted to offer a fresh perspective on the question of obedience to harmful orders. To achieve this, I developed a novel experimental approach that involved no deception, utilized real harmful behavior, and was adapted for neuroimaging and electrophysiological recordings [Caspar 2021, Caspar, Christensen *et al.* 2016]. The methods and results are described later in this paper, revealing specific neurocognitive processes altered by obedience to authority. My intention was further to use this experimental approach not only with classic WEIRD (Western, Educated, Industrialized, Rich & Democratic) individuals but also with diverse populations recruited worldwide—an approach not commonly seen in neuroscience.

However, I soon realized that addressing the question of obedience to harmful orders exclusively in the laboratory would be insufficient. This is especially true when the participants are typical university students. How can we theorize about the mental processes of people obeying orders that have tragic consequences without engaging with the individuals involved, without talking to those who committed such atrocities? To comprehensively understand human behavior, a holistic approach is necessary. We must transcend the boundaries of individual disciplines to fill in the missing pieces. Consequently, I have begun to consider studying and interviewing genocide perpetrators and military personnel. This work has not only offered novel insights for my experimental research but has also provided a more global overview and understanding of my results.

In the present paper, I will describe the interviews conducted in Rwanda with former perpetrators released from prison and discuss their main findings. I will also describe how the question of obedience to harmful outcomes has been approached in neuroscience and highlight the novel insights it has provided. I will occasionally mention work conducted in social psychology that follows Milgram’s studies, although it will not be the central focus of this work, as it has been extensively described elsewhere. I will conclude by arguing that combining these diverse approaches can lead us to a more comprehensive understanding of human obedience.

2 Qualitative interviews with former genocide perpetrators

Qualitative interviews offer a valuable source of information by providing insights directly from individuals involved in acts of killing, shedding light on the motivations behind such actions. However, there are several limitations

and difficulties that must be acknowledged, and it is important to mention them before presenting the results of these interviews.

First, the elements reported in the interviews are not objectively verifiable [Anderson 2017b], [Fujii 2010]. As they rely on what the interviewees accept to share, some of the provided responses can be, purportedly or not, false, deformed, attenuated or not complete. For instance, even if some individuals have been judged for their crimes, they may also keep hidden other crimes in order to prevent additional conviction in Court. Others may deny their crimes as part of a psychological process aiming at rebuilding a positive image of themselves [Anderson 2017b]. Others again may deform their responses, especially when they have to depict the terrible acts they did, because they want to appear nicer in front of the interviewer. This effect, well-known in social psychology, is called the social desirability bias [Nederhof 1985], which refers to the tendency of individuals to show favorable images of themselves and to act in a way that is thought to be appropriate during social interactions. A second reason is that most of the times, interviews are conducted years, even decades, after the dramatic event. The memory of the interviewees could thus be altered.

As an additional challenge for the generalizability of the findings, interviews with perpetrators of genocide are relatively rare, for different reasons. Some scholars have attributed this scarcity to the psychological challenges associated with confronting unspeakable atrocities while attempting to comprehend the decisions made by those involved [Anderson 2017a]. Conducting such interviews necessitates significant psychological and emotional preparation, and not everyone is willing to hear such atrocities. Further, in many countries where genocides have occurred or are ongoing, these tragic events may not be officially recognized, making it nearly impossible for researchers to reach these populations for interviews. In other cases, the perpetrators may no longer be alive. Despite the numerous ongoing mass atrocities occurring worldwide, the options for conducting interviews remain limited.

In addition, recruiting former genocide perpetrators to conduct interviews may be a particularly hard task, especially for foreigners. Former genocide perpetrators may be reluctant to openly discuss their actions, particularly with people they do not know. In Cambodia for instance, I faced this challenge, also because very few perpetrators have faced judgment for their actions during the Khmer Rouge regime between 1975 and 1979 [Ciorciari & Heindel 2014]. They thus do not have any reasons to speak to foreigners about their potential crimes.

This section presents a summary of the results of the interviews conducted with 49 former genocide perpetrators released from prison in Rwanda. A more detailed analysis of the interviews conducted in Rwanda and Cambodia is available in another scientific publication [Caspar In press] and in a book [Caspar 2024]. Here, I will primarily provide a brief overview of the interviewees' profiles and their stated reasons for participating in the genocide.

To begin with some demographic information, all of the respondents in Rwanda had been tried during the *Gacaca* Trials, which were popular tribunals reinstated in 2002 to prosecute those responsible for the genocide [Clark 2010]. On average, they had spent nine years in prison for their crimes. The mean
 5 age of our respondents was 60 years, with ages ranging from 41 to 79 years. The interviews took place in Kinyarwanda, and I was assisted by two research assistants from the University of Rwanda. To gain their trust, we worked with local volunteers from Prison Fellowship Rwanda who went door to door in several villages to ask the targeted individuals if they would accept to
 10 participate in our studies and to be interviewed regarding their actions during the genocide.

There were three main categories of crimes for which our respondents had been convicted. The crimes included group attacks (19/49 respondents), murder (29/49 respondents) and looting (11/49 respondents). Some of our
 15 respondents had been convicted for several of those crimes, and sometimes being in a group attack also included having murdered people. For instance, one of the respondents indicated:

“The crime I committed against the Tutsis, I was taught to go and kill the Tutsis and I went to grab a machete and slaughter
 20 them and looted their cows.” [Caspar In press]

This respondent was categorized for murder and looting. Another respondent reported:

“It is Genocide. I committed murder; I went into group attacks and killed. I killed four people, but I killed two by myself and
 25 the other two I killed them with the help of others that we were together at time.” [Caspar In press]

This respondent was categorized for murder and group attacks.

When asked why they had committed these crimes, several categories of answers emerged, but the main reason provided was obedience to authority.
 30 Indeed, approximately 70% of our respondents reported that they did so because they were following the orders of the government. They indicated for instance:

“It was caused by the government that ruled at that time, eeh it encouraged us to kill people [...] that’s it,” or “It is bad
 35 leadership that instructed us to kill people and become animals, even though we were not animals. Yes, it is the leadership that did this, not us,” or even “I committed the crime because of the bad government that was there at that time. It was not me as they instructed us to kill,” or again “Because of being ordered around and coerced by the government; we had no individual thoughts or
 40 feelings.” [Caspar In press]

This finding aligns with results observed by previous scholars. For example, during interviews conducted by Anderson [Anderson 2017b], over sixty percent of respondents also reported that the genocide resulted from obedience to authority. Straus also noted that many perpetrators he interviewed justified their actions by claiming they were following orders [Straus 2007].

The second reason provided by the former perpetrators was another form of social influence, that is, conformity to group [Cialdini & Goldstein 2004]. During the genocide against Tutsis, an armed militia, the *Interahamwe* (translated as “Those who attack together”), lead the killings and influenced others to join them. Of the 49 perpetrators interviewed, 9 (resulting in a percentage of 18.36%) indicated that *Igitero* (pl. *Ibitero*, meaning “group attacks”) was the reason for participating. Some examples include for instance:

“My armed group killed 7 people. Sorry, it was actually rather 7 children and their mother, making it 8 in total,” or “I joined group attacks to run after the Tutsis and went on patrol that killed many Tutsis,” or “Literally it was because of the pressure they put on us to go into the group attacks. So I went into one. Ehh, although I didn’t kill anyone, I did go in a group attack, and it killed people.”

Such reports, and the fact that they displace the responsibility of their individual actions on the group, are consistent with those obtained by other academics. One of the Rwandan perpetrators interviewed by Anderson [2017b] for instance reported: “I did not kill anybody. [...] The group I was in killed those six victims.”

While some people reported that they killed for obeying orders or because of the influence of the group, it appears that some individuals were afraid for their lives if they did not join. About 10% of our respondents indeed indicated that they felt forced to participate in the massacres. For instance, one interviewee reported:

“For the government that was in place at the time, the killings were almost law, which is why people were afraid for their lives. and I chose to do what they told me.”

and another one reported

“At that time, when you didn’t collaborate with others you were likely to be in danger, so it was in a manner of protecting myself. They even killed my dad because they found a person hiding in our home and butchered our cow, so it was a way of protecting myself.”

This is consistent with some reports indicating that the *Interahamwe* went door by door to tell the men to join the killings [Fujii 2008]. During that period, killing Tutsis was, in a way, mandated by law, and refusal to participate could

have had consequences, although this was not the case everywhere. It has been observed that in some districts or villages men were initially reluctant to participate and were then coerced by authorities to join the massacres. Conversely, in other areas, many men volunteered to join; some even competed
5 for the chance to be involved [[Jones 2002](#)].

3 “Just following orders,” over and over again

In the scientific literature, it has been described that participation in genocide can be understood at two levels: structural and individual [Loyle 2009]. Structural-based theories provide explanations rooted in the unique institutions, culture, and circumstances of a society at a specific time. In contrast, individual-based theories propose reasons for participation that are specific to each participant.

In the context of Rwanda, a structural explanation for the genocide would include, among other factors, a deeply ingrained culture of obedience [Hilker 2012]. Many scholars and journalists have reported that deference to authority played a significant role in explaining the genocide in Rwanda, as a substantial number of people followed orders from authorities to commit acts of violence [Paluck & Green 2009], [Prunier 1998], [Fujii 2008], [McDoom 2021]. It was therefore expected that the authority argument used by the genocide perpetrators would be that spread. Consequently, in Rwanda, individual motivations for participation may be intertwined with this structural element. Furthermore, many former perpetrators were imprisoned together, potentially leading to the creation or sharing of a common narrative to explain their participation [Mironko 2004].

However, the justification of “just following orders” has been used in several documented wars and genocides beyond Rwanda. In Western cultures, the most infamous example of justifying mass atrocities in the context of obedience is certainly the phrase “Befehl ist Befehl” (translated as “An order is an order”) used by senior Nazi officers during the Nuremberg Trials. Some authors, however, argue that this “just following orders” argument should be considered an alibi rather than a justification [Mandel 1998], particularly within the context of a trial. Moving beyond a Western- and trial-centered perspective to a more global approach, it is interesting to note that similar justifications can be found—though not exhaustively—in cases like the genocides in Indonesia and in Cambodia. Specifically, in the case of the former Khmer Rouge members, all of those interviewed claimed they were “just following orders,” despite never having feared prosecution in a trial [Caspar In press].

It is interesting to note that all these events occurred at different times in human history, in different countries, on different continents, involving people from very diverse cultures. Is it possible that the commonly cited justification of “just following orders” offered by perpetrators, might be rooted in a shared aspect of human cognition? Could this inclination be, in part, a reflection of a neurological trait shared among all members of our species? This reflection does not rule out the possibility that the “just following orders” argument can also serve as an alibi. It suggests that beyond being a *post hoc* justification,

obeying orders can also, at the moment of the act, influences how the brain processes information and computes decisions.

4 A novel experimental approach to study (dis)obedience to authority

Previous research conducted following Milgram's studies predominantly adopted a method heavily influenced by Milgram's original experimental design, albeit with minor variations. My aim was to develop a novel experimental task that would address the ethical, interpretational, and methodological challenges associated with Milgram's studies while striving for ecological validity.

The ethical concerns linked to Milgram's studies are frequently cited, particularly the high levels of stress experienced by participants and the use of a cover story involving deception [Baumrind 1964], [Miller 1986], [Perry 2013]. While subsequent variants using immersive virtual reality have alleviated some of these concerns [Slater, Antley *et al.* 2006], the transparency of the simulated scenario may not fully capture decision-making in a real-world context. Other variations, which adapt Milgram's initial methods, such as the 150-volt approach [Burger 2009], may appear to align with current ethical standards but still present interpretational and methodological challenges. Researchers employing cover stories face the challenging task of determining whether participants genuinely believed in the fictitious narrative, leaving room for doubt in the interpretation of results.

Recent analyses of explicit reports from Milgram's participants suggest that there is limited reliable evidence to confirm their belief in the cover story [Griggs & Whitehead 2015], [Haslam, Reicher *et al.* 2014], [Perry 2013]. Some authors have postulated that observable signs of stress during the experiments, such as trembling hands and nervousness, indicate participants believed they were actually harming another human being [Blass 1999a]. However, this interpretation has been challenged by virtual reality studies demonstrating that participants can exhibit physiological stress responses even in obviously fictitious experimental scenarios [Slater, Antley *et al.* 2006]. These divergent interpretations of Milgram's studies highlight the difficulties in interpreting results when cover stories are used [Kelman 1967]. A real scenario in which participants make decisions with actual consequences for another person emerges as a more robust solution to mitigate interpretation challenges. Lastly, methods based on the original Milgram paradigm, such as virtual reality versions [Slater, Antley *et al.* 2006] or the 150-V method [Burger 2009] are not adapted to neuroimaging and electrophysiological measurements. To achieve a satisfactory signal-to-noise ratio, multiple identical trials must be recorded. In Milgram-like experimental approaches, the trials are few in number and all distinct from one another.

In our novel approach [Caspar 2021, Caspar, Christensen *et al.* 2016], participants were tested in pairs, and care was taken to ensure that they did not know each other prior to the experiment. Within each pair, one individual took the role of the “agent,” while the other took on the role of the “victim.” In the middle of the experiment, they switched roles, ensuring a fully reciprocal procedure. This approach aimed to prevent our participants from exclusively embodying the role of the perpetrator, as seen in Milgram’s studies, which could lead to psychological distress [Zimbardo 2007]. Furthermore, unlike Milgram’s studies, both participants in our setup were real participants, with none serving as a confederate. This choice allowed us to avoid the use of cover stories and the associated interpretational challenges.

During the experiment, the two participants sat face to face at a table with a keyboard between them. The keyboard featured two buttons: one labeled “SHOCK” and the other labeled “NO SHOCK.” Participants were informed that, when acting as the agent, they were responsible for pressing these buttons. Pressing the “SHOCK” button would administer a real, calibrated electric shock of painful intensity to the “victim” while increasing the agent’s remuneration by £0.05. Conversely, pressing the “NO SHOCK” button would withhold the electric shock, and the agent would not receive the monetary incentive.

The pain threshold was determined before starting the experimental procedure. Two electrodes were attached to each participant’s left hand, connecting them to a constant current stimulator (Digitimer DS7A), a device designed for delivering electrical stimulations. This non-invasive procedure induced momentary pain and was used to determine a painful yet tolerable threshold. Once established, this threshold remained consistent for participants throughout the entire experimental process when they were in the role of the victim. Through this procedure, we ensured that participants were aware of the pain they could potentially inflict on each other and were willing to experience it themselves. Importantly, no participants reported enduring long-term effects associated with this procedure, even during post-session debriefings.

In one experimental condition, agents were informed that they had complete freedom to decide which buttons to press during 60 trials (referred to as the free-choice condition). In the other experimental condition, agents were informed that they would receive instructions from the experimenter to either send or refrain from sending a shock on each trial (referred to as the coerced condition). It is important to note that the term “coercion” here is used in a relative sense, rather than an absolute one. In strict terms, “coercion” refers to the use of force to compel someone to do something against their will, which is ethically unacceptable in experimental research. In our context, “coercion” simply describes an experimental situation where individuals follow orders from an experimenter to administer painful stimulation to another individual. Crucially, we never instructed our participants that they must obey these orders, and the experimenter did not provide any incentives for participants

to comply, as was done in Milgram’s studies. Unlike Milgram-like studies, where delivering a shock is associated with a mistake made by the learner, participants in both our free-choice and coerced conditions were not given any reasons for administering shocks.

When I created this experimental protocol together with my collaborator, I thought that no one would ever administer genuine, painful shocks to another person for a mere £/€0.05. I was proven wrong, as participants administered approximately 28 out of 60 shocks to the “victim” in the free-choice condition across various studies conducted with the same paradigm [Caspar, Christensen *et al.* 2016, Caspar, Vuillaume *et al.* 2017, Caspar, Beyer *et al.* 2021, Caspar, Lo Bue *et al.* 2020]. Furthermore, I was equally convinced that no one would obey my orders within this experimental protocol. Firstly, because these shocks caused actual physical harm to another person, and participants were fully aware of this aspect of the task. Secondly, Milgram’s studies [1963, 1974], were so (in)famous that it seemed likely participants, especially university students, had heard about them and would refuse to comply with orders. Additionally, I did not employ any prods, such as those used in Milgram’s studies, to encourage participants to continue. Once again, I was proven wrong. To date, my collaborators and I have tested approximately 1,500 volunteers using this procedure, and only about 54 of them disobeyed my orders with this experimental procedure: 43 for prosocial reasons (i.e., they refused to administer an electric shock to another individual), four due to contradiction (i.e., they systematically pressed the opposite button regardless of the order), and seven for antisocial reasons (i.e., administering shocks even when instructed not to). It should be noted that we have created small variants of the aforementioned paradigm, which allow for a higher rate of disobedience [Caspar 2021, Caspar, Gishoma *et al.* 2022], [Tricoche, Rovai *et al.* 2024], thus offering the possibility to study the neuro-cognitive processes that support resistance to immoral orders. We have now initiated data collection in Rwanda, with this task involving a higher rate of disobedience. We directly compared former genocide perpetrators and genocide rescuers to investigate whether they rely on similar neural underpinnings associated with resistance to immoral orders.

Of course, questions may arise regarding the conceptual aspects of the task, including its ecological validity and its ability to predict real-life immoral acts like mass murder. This issue is inherent to any laboratory approach, as critics have often questioned whether the results of psychology’s laboratory experiments can be generalized [Holleman, Hooge *et al.* 2020]. This is probably even more the case for research aimed at understanding real-life mass atrocities, which are highly complex and multifactorial behaviors. Originally, I create this paradigm to create a moral dilemma for participants (i.e., causing physical harm to someone in exchange for money) in a setting allowing neuroimaging techniques. Currently, our strongest evidence of the paradigm’s ecological validity lies in the very low resistance to immoral orders observed in Rwanda with this task [Caspar, Gishoma *et al.* 2022], compared to other

countries. This aligns with the academic emphasis on deference to authority as a contributing factor in explaining the 1994 genocide in Rwanda [Prunier 1998]. As mentioned, I am currently delving deeper into this topic, by studying individuals who have genuinely saved lives during genocides, often referred to as “the Righteous,” (*les Justes*, in French) to determine whether they too would resist immoral orders in this experimental paradigm. However, it is essential to recognize a significant limitation of a laboratory setting when it comes to reflecting real-life mass atrocities. The richness of the real-world context is challenging to fully capture within the confines of a laboratory. Moreover, the behaviors being examined are vastly different, ranging from administering a painful shock to another person to taking the life of another human being. That is why I strongly believe that for a better understanding of mass atrocities, we must combine experimental approaches to study the influence of various variables on specific processes with qualitative approaches that involve engaging with individuals who have directly experienced such atrocities outside the laboratory setting. Being aware of these potential limitations, I will now describe in more depth the main results obtained with a neuroscience approach to obedience.

5 The obedient brain: How does following the orders of an authority impact cognition

In recent years, neuroscience has begun to explore the concept of obedience, offering fresh and complementary insights into how following can influence brain function and create conditions favorable to moral transgressions. The objective here is not merely to determine the factors that lead to varying levels of obedience but rather to unravel the neural mechanisms behind compliance with orders. The neurocognitive processes investigated in existing research are closely associated with prosocial decision-making, thereby illuminating how obedience can lead to antisocial behaviors. I have separated these processes into two categories: one focuses more on the feeling of responsibility and being the author of one’s own actions (i.e., the sense of agency), and the other focuses on moral emotions (i.e., empathy for pain and the interpersonal feeling of guilt).

5.1 Sense of agency & feeling of responsibility

In the neuroscience literature, the sense of agency refers to an individual’s subjective awareness or feeling that they are the agent or initiator of their own actions and behaviors [Gallagher 2000]. It involves the perception that one’s actions are intentional and under their own control, leading to a sense of

responsibility for those actions. This concept is central to understanding how individuals attribute their actions to themselves and distinguish them from external events or influences.

The sense of agency is not a unitary phenomenon as it involves different
 5 conscious experiences of authorship, in particular the judgement of agency
 and the feeling of agency [Synofzik, Vosgerau *et al.* 2008]. The judgment
 of agency relates to the reflective, *post hoc* evaluation of agency, while the
 feeling of agency is the immediate, pre-reflective, and subjective sense of being
 the agent during the actual execution of an action. The judgment of agency
 10 is classically studied with explicit questions after performing specific actions.
 These questions may inquire about the perceived level of control, responsibility,
 or authorship over their actions. Participants may be required to provide
 numerical ratings or rankings that indicate their perceived level of agency,
 or to describe their subjective experience of agency during specific actions.
 15 However, explicit reports are known to be sensitive to social desirability
 and cognitive dissonance [Yoshie & Haggard 2013]. The feeling of agency
 is typically measured with implicit methods, which restricts the influence of
 biases, one of the most spread being the temporal binding paradigm [Haggard,
 Clark *et al.* 2002]. In this task, participants are asked to judge the perceived
 20 time interval between their voluntary action (e.g., pressing a button) and
 a subsequent sensory event (e.g., a flash of light or a sound). The results
 generally suggest that when comparing the perceived time intervals between
 a voluntary action and a tone versus an involuntary movement (e.g., induced
 by neuromodulation) and the same tone, the former is perceived as shorter.
 25 This implies that the execution of voluntary actions tends to create an illusion
 of time passing more quickly between the action and its consequence. This
 phenomenon is likely influenced by striatal dopaminergic activity, known for its
 role in time perception [Meck 2006] and in transmitting information from the
 basal ganglia to frontal motor areas [Nachev, Kennard *et al.* 2008], which are
 30 critical brain regions associated with the sense of agency [Haggard & Whitford
 2004], [Kühn, Brass *et al.* 2013]. The sense of agency is highly associated with
 the feeling of responsibility. However, the later relates more to an individual's
 mental representation of authorship within social and moral contexts [Balconi
 2010]. Researchers have typically measured this aspect using explicit questions
 35 posed to participants [Beyer, Sidarus *et al.* 2017, Caspar, Beyer *et al.* 2021],
 such as asking individuals whether they feel accountable for the outcomes of
 their actions.

By using the paradigm described previously, in a seminal study we
 evaluated how obeying orders would influence the feeling of responsibility
 40 and the sense of agency [Caspar, Christensen *et al.* 2016]. After pressing
 the SHOCK or the NO SHOCK buttons, agents could hear a tone, displayed
 after a few hundred milliseconds. They were instructed to report in ms
 the estimated action-tone intervals. They had to perform this estimation
 task in both the free-choice and the coerced condition. At the end of each

experimental condition participants had to report how responsible they felt on a scale ranging from “0” (not responsible at all) to “100” (entirely responsible’).

Importantly, the action-tone delays were exactly similar in the two experimental conditions, unknowingly to the participants. Yet, we observed that obeying orders lead to the interval between action and tone being perceived as longer than when the action was freely chosen. Participants indeed reported that more milliseconds elapsed between their action and the resulting tone in the coerced condition compared to the free-choice condition. This result implies that participants felt less agency over the outcomes of their actions when they were told what to do, compared to when they decided for themselves. This result has now been replicated many times [Caspar, Vuillaume *et al.* 2017, Caspar, Cleeremans *et al.* 2018, Caspar, Lo Bue *et al.* 2020, Pech & Caspar 2023], also in different contexts and by other research teams [Akyuz, Marien *et al.* 2023], [Barlas 2019], and does suggest that following someone else’s instructions diminishes the experience of authorship. We also observed that people systematically reported feeling less responsible in the coerced condition compared to the free-choice condition.

In a Magnetic Resonance Imagery Study (MRI) [Caspar, Beyer *et al.* 2021], we observed that resilience against agency reducing effects of coercion is associated with recruitment of volitional processes reflected in medial frontal gyrus (MFG) activity. A similar effect was found using explicit responsibility ratings: the more participants reported that they experienced feeling responsible over their actions in the coerced condition, the more activity was observable in the medial frontal gyrus. The cluster we found to correlate with both temporal binding and responsibility effects lies at the anterior border of the pre-SMA, in a region associated with voluntary action selection [Karch, Mulert *et al.* 2009], [Rushworth 2008]. These findings suggest that volitional processes during action planning and execution help to preserve a strong sense of agency under coercion.

5.2 Empathy for pain

When we witness someone suffering, be it physically or emotionally, we are likely to feel their pain, a phenomenon that is referred to as “empathy.” Empathy is a psychological construct that refers to the ability to understand and imagine what others feel. There is an extensive literature that have relied on the use of self-report questionnaires on one’s own estimated level of empathy, or on the use of subjective pain rating scales to capture the empathic experience [Gerdes, Segal *et al.* 2010]. But roughly two decades ago, neuroscientists have started to document what is happening in the brain when we experience empathy for others and have studied to what extent these brain activations may be related to prosociality.

In the case of empathy for pain, many experimental studies have shown that seeing another individual in pain triggers an empathic response in the

brain of the observer. Such vicarious activations mostly occur in the anterior cingulate cortex and in the insula, key brain regions that allow to understand the emotional component of what it feels to experience pain [Keysers & Gazzola 2014], [Singer, Seymour *et al.* 2004], [Timmers, Park *et al.* 2018].
 5 Because of these shared activations, we are likely not to inflict pain to others, because we understand how they would feel [Decety 2011].

There is a large consensus in the scientific community to agree that our capacity to feel and imagine the pain of others is largely automatic [Singer, Seymour *et al.* 2004], [Zaki & Mitchell 2013] and deeply engrained
 10 into our biology, likely because of the presence of mirror neurons [Gallese, Keysers *et al.* 2004, Gallese & Goldman 1998]. Yet, past experimental research has shown that empathy, despite being an inner and relatively automatic process, is also context-dependent [Zaki 2014]. Empathy for others' pain is reduced when the person receiving pain played unfair [Singer,
 15 Seymour *et al.* 2006], or when we observed the pain of an outgroup individual [Caspar, Pech *et al.* 2023], [Pech & Caspar 2022], [Cikara, Bruneau *et al.* 2011], [Han 2018], [Hein, Silani *et al.* 2010]. Empathy can also be increased if we share similar experiences with the person in pain [Hodges, Kiel *et al.* 2010], if we receive a reward [Klein & Hodges 2001], or
 20 if we consider the observed pain from a first-person perspective [Bucchioni, Fossataro *et al.* 2016].

We conducted two MRI studies to understand if our empathy for the pain of others could be blurred or diminished when people follow orders compared to when they acted freely [Caspar, Ioumpa *et al.* 2020, 2022]. Such finding would
 25 indeed be crucial to explain why obeying orders can diminish our aversion to hurt others. To measure empathy for pain, a critical aspect is that participants must see the pain being delivered. As agents and victims were not facing each other as agents were located in the MRI scanner and "victims" were outside, we used a system of real-time video recording to display the hand of
 30 the agent receiving the painful shocks on the screen of the agent inside the MRI [Caspar, Ioumpa *et al.* 2020]. The procedure with the electric shocks is highly relevant here, because when the electrical stimulation is delivered it produces a visible muscular contraction on the hand. This visible element is what triggers the neural empathic response in the brain of the observer. Due
 35 to the high noise produced by the MRI scanner, agents were receiving orders through headphones, but the experimenter was present in the MRI scanner room to increase the belief that she was giving real-time orders.

Neuroimaging results showed that empathy-related regions, notably in the anterior cingulate cortex and the anterior insula, were less active when
 40 participants were obeying orders as opposed to when they were acting freely, despite the intensity of the electric shocks being exactly the same in both experimental conditions. These findings have been replicated with electroencephalography (EEG), where brain potentials over the centro-parietal cortex (i.e., P3, LPP, [Coll 2018]) were lower when participants obeyed orders,
 45 compared to when they acted on their own accord [Pech & Caspar 2023].

Additionally, in a study involving the first generation born in Rwanda after the Genocide Against Tutsis, those with the highest amplitude of the P3 and LPP while obeying an order to administer a shock were more likely to refuse the experimenter's orders [Caspar, Gishoma *et al.* 2022], thus highlighting also the potential role of empathy in the context of obedience.

The finding that obeying orders reduces empathy is particularly interesting because our participants were aware that the pain intensity of the shocks was the same in both experimental conditions. They had tested the machine before starting the experiment and were explicitly informed that the shock threshold would remain constant. Nevertheless, the electric shock, identical in both scenarios, was perceived as less painful when delivered under orders. Participants even reported subjectively that the shocks felt less painful when they obeyed orders compared to when they acted freely [Caspar, Ioumpa *et al.* 2020]. These results overall underscore how compliance with orders can diminish our natural aversion to causing harm to others.

5.3 Interpersonal feeling of guilt

Guilt is a powerful emotion which usually arises when we violate social norms [Haidt 2003]. It usually motivates transgressors or perpetrators to make amend, to restore damaged social relationships and to perform good deeds [Tangney, Stuewig *et al.* 2014].

All the studies I have conducted with the above-mentioned paradigm were systematically accompanied by a debriefing at the end of the experimental session. In that debriefing, among several open questions, I asked the participants how sorry they felt for the “victim” and how bad they felt for sending those painful shocks in exchange for money. An interesting observation is that participants systematically reported that they felt less sorry for the victim and less bad for sending the shocks when they obeyed orders compared to when they acted freely and sent the same shocks [Caspar, Christensen *et al.* 2016, Caspar, Vuillaume *et al.* 2017, Caspar, Cleeremans *et al.* 2018]. Such observation would indicate that participants experienced less guilt when they were obeying orders, even though the same actions were conducted when they were free to decide.

However, it is difficult to know if they minimize their feelings of sorriiness and badness to provide a justification to their actions *a posteriori*, or if it reflects a process that was already ongoing when they were performing the actions. We thus again used a brain approach in order to understand if obeying orders would reduce brain activity in regions associated with the feeling of guilt and this is exactly what we observed [Caspar, Ioumpa *et al.* 2020]. When they obeyed the order to send shocks to the victim compared to freely choosing, activity in guilt-related brain regions was reduced suggesting that the experience of guilt may be blurred in obedience situations. Further, using also fMRI, a team of researchers used a virtual obedience paradigm

to simulate antisocial behavior under coercion [Cheng, Chou *et al.* 2021]. They investigated how psychopathic traits would alter neural activations of guilt under coercion. They found that guilt feelings induced by harming under obedience predicted higher activity in anterior middle cingulate cortex (aMCC), a key brain region associated with the feeling of guilt [Yu, Koban *et al.* 2020], but that psychopathic traits significantly mediated this guilt-brain association. Finally, another study [Caspar, Ioumpa *et al.* 2022] showed that the activations in guilt-related brain regions were also equally low for commanders giving orders, for transmitters, and for agents executing orders. Taken together, these results suggest that obeying orders may impact how the brain processes guilt.

6 Novel insights into obedience through interdisciplinary approaches

The main argument I aimed to defend in this article was that it is crucial to engage in more multi-approach, multi-population, and multi-disciplinary perspectives to better understand the phenomenon of obedience. By examining obedience through these diverse lenses, researchers can aim to construct a more nuanced and comprehensive model that accounts for the complexities of human behavior under authoritative pressure. So far, multi-approach, multi-population, and multi-disciplinary perspectives on obedience remain too limited in the scientific literature. However, I will conclude this article by providing some lines of research that would combine these multiple approaches to show how they can broaden our understanding of obedience to authority.

An interesting first observation is that the reduction of agency and responsibility observed using implicit and neuroimaging methods when people obey orders are somehow related to the agentic state theory of Milgram. Milgram explained that when people follow the orders of an experimenter, they transfer their own agency and responsibility to the experimenter [Milgram 1974]. They become “thoughtless agents of action,” they enter in an “agentic state” [1974, 132–134]. However, while some academics agreed with his theory [Blass 1999b], other were concerned about its validity [Haslam, Reicher *et al.* 2015a]. Despite the initial lack of reliable evidence in favor of the agentic state theory, experimental research combining explicit, implicit, electrophysiological and neuroimaging methods seem to indicate that Milgram was perhaps not entirely wrong. Of course, people can have different reasons to obey or disobey orders. But it appears that once they have accepted to comply, their brain starts processing agency and responsibility differently as in a context of freedom of choice.

Further, it was claimed that the agentic state theory of Milgram can hardly explain diverse degrees of obedience across the different variants of the studies of Milgram [Haslam, Reicher *et al.* 2015a]. The argument suggested that if

agents displace their agency and responsibility towards the experimenter when they obey orders, this should be the case regardless of the different contexts. However, this argumentation forgets the crucial interplay between individual and contextual factors to explain decision-making processes. As an example,

5 Milgram showed in one of his variants that when the learner is present in the same room as the real participant, obedience decreases [Milgram 1974] compared to when the learner sits in another room. Literature in neuroscience has shown that visualizing a person in pain triggers an empathic reaction in the brain of the observer [Jauniaux, Khatibi *et al.* 2019], but additional

10 research has further shown that auditory information can also trigger similar brain activations [Hoenen, Lübke *et al.* 2018]. Empathy is commonly viewed as relying on integrating information from multiple sensory modalities [Morrison, Lloyd *et al.* 2004]. Therefore, the combination of both visual and auditory modalities when the learner is in the same room as the agent could have

15 enhanced empathy, which could have thwarted the effect of receiving orders, regardless of the experience of agency. It has indeed been shown that a higher neural response to the pain of others is associated with a resistance to immoral orders [Caspar, Gishoma *et al.* 2022], [Tricoche, Rovai *et al.* 2024]. The importance of empathy for resisting pressure to others would also be

20 consistent with studies on rescuers that showed that acts of rescue frequently emerged when an individual asks for help in person [Varese & Yaish 1998]. According to this example, a critical future line of research for the study of obedience would therefore conciliate experimental approaches that combine an individual perspective together with the manipulation of social and contextual

25 variables.

From the interviews with the former genocide perpetrators in Rwanda and Cambodia [Anderson 2017b], [Caspar *In press*], [Hinton 1998], [Straus 2007], it has been shown that many reported that obedience to authority

30 was a highly frequent justification provided to explain their participation in the genocide. The second justification reported is the influence of the group [Caspar *In press*], [Mironko 2004]. There is an interesting parallel to make with neuroscience research, as the literature indicates that authority influence impacts neural processing more than group influence. In a study using magnetoencephalography (MEG) and conducted in Europe [El Zein,

35 Dolan *et al.* 2022], participants undertook a gambling task either alone, with others (i.e., conformity), or with someone else deciding for them (i.e., obedience). Results indicated that the sense of responsibility felt by participants diminished as more players were involved, being least in the obedience condition. Further, the authors observed that at 200–300 ms after

40 an outcome, MEG activity of bilateral frontoparietal brain regions decreased linearly, being the highest when participants were alone, less in groups, and least when they followed an order. This suggests group settings diminish one’s sense of responsibility and related brain activity, with the impact being strongest when decisions are made by someone else. Another study conducted

45 in China confirmed these findings [Xie, Chen *et al.* 2016]. Participants had to

quickly decide whether to buy a book based on limited information. Either they were triggered by the impact of the majority, reflected in positive and negative feedback (i.e., conformity), or were directed to buy books which had a majority of negative reviews (i.e., obedience). EEG results showed that obedience decisions induced greater cognitive conflict, as reflected by the amplitude of the N2 component, compared to conformity decisions. These studies have underscored the brain reactions associated with conformity within a group and compliance with authority. It seems that adhering to authoritative commands has a more substantial effect on human actions compared to mere group conformity. This conclusion aligns with the outcomes of interviews, in which perpetrators identified obedience as having a greater effect on their behavior than the influence exerted by peer pressure. Nevertheless, research in this area remains quite restricted and requires more in-depth investigation.

To conclude, approaching the question of how people can commit atrocities when they follow orders implies multi-disciplinary research. Interviews conducted with former genocide perpetrators seem to largely indicate that they did not experience responsibility over their actions as they simply followed orders. Neuroscience research tend to indicate that obeying the orders of an authority impacts the experience of agency, the feeling of responsibility, empathy for the pain caused to the victim and the interpersonal feeling of guilt. Together, such results deepen our understanding of how obedience alters our natural aversion to hurt others.

But the studies of obedience still have a long way to go. A critical research line in the future would be to open such line of research to other populations. Testing WEIRD and non-WEIRD populations may favor the generalization of results in times where the validation of theories on “all human brains” is a crucial next step. I strongly believe that exiting the comfort of our research labs and talking to individuals who have a specific relationship to authority, based on their own experience in specific situations or because of cultural elements, would critically advance the research of obedience.

Bibliography

- AKYUZ, Nil, MARIEN, Hans, *et al.* [2023], Revisiting the agentic shift: Obedience increases the perceived time between own action and results, Preprint, doi: [10.21203/rs.3.rs-3705074/v1](https://doi.org/10.21203/rs.3.rs-3705074/v1).
- ANDERSON, Kjell [2017a], *Perpetrating Genocide: A Criminological Account*, London: Routledge, doi: [10.4324/9781315626239](https://doi.org/10.4324/9781315626239).
- [2017b], “Who was I to stop the killing?” Moral neutralization among Rwandan Genocide perpetrators, *Journal of Perpetrator Research*, 1(1), doi: [10.21039/jpr.v1i1.49](https://doi.org/10.21039/jpr.v1i1.49).

BALCONI, Michela [2010], *Neuropsychology of the Sense of Agency*, Milan: Springer, chap. The neuropsychology of senses of agency: Theoretical and empirical contributions, 47–67, URL https://doi.org/10.1007/978-88-470-1587-6_3.

- 5 BARLAS, Zeynep [2019], When robots tell you what to do: Sense of agency in human- and robot-guided actions, *Consciousness and Cognition*, 75, 102819, doi: [10.1016/j.concog.2019.102819](https://doi.org/10.1016/j.concog.2019.102819).

BAUMRIND, Diana [1964], Some thoughts on ethics of research: After reading Milgram’s “Behavioral Study of Obedience”, *American Psychologist*, 19(6),
10 421–423, doi: [10.1037/h0040128](https://doi.org/10.1037/h0040128).

BEYER, Frederike, SIDARUS, Nura, *et al.* [2017], Beyond self-serving bias: Diffusion of responsibility reduces sense of agency and outcome monitoring, *Social Cognitive and Affective Neuroscience*, 12(1), 138–145, doi: [10.1093/scan/nsw160](https://doi.org/10.1093/scan/nsw160).

- 15 BLASS, Thomas [1999a], *Obedience to Authority: Current Perspectives on the Milgram Paradigm*, New York; London: Psychology Press.

— [1999b], The Milgram Paradigm after 35 years: Some things we now know about obedience to authority, *Journal of Applied Social Psychology*, 29(5), 955–978, doi: [10.1111/j.1559-1816.1999.tb00134.x](https://doi.org/10.1111/j.1559-1816.1999.tb00134.x).

- 20 BUCCHIONI, Giulia, FOSSATARO, Carlotta, *et al.* [2016], Empathy or ownership? Evidence from corticospinal excitability modulation during pain observation, *Journal of Cognitive Neuroscience*, 28(11), 1760–1771, URL https://doi.org/10.1162/jocn_a_01003.

BURGER, Jerry M. [2009], Replicating Milgram: Would people still obey today, 25 *American Psychologist*, 64(1), 1–11, doi: [10.1037/a0010932](https://doi.org/10.1037/a0010932).

CASPAR, Emilie A. [2021], A novel experimental approach to study disobedience to authority, *Scientific Reports*, 11(1), 22927, doi: [10.1038/s41598-021-02334-8](https://doi.org/10.1038/s41598-021-02334-8).

— [2024], *Just Following Orders? Atrocities and the Brain Science of Obedience*, Cambridge: Cambridge University Press.

— [In press], Understanding individual motivations and desistance: Interviews with genocide perpetrators from Rwanda and Cambodia, *Journal of Perpetrator Research*.

- 35 CASPAR, Emilie A., BEYER, Frederike, *et al.* [2021], The obedient mind and the volitional brain: A neural basis for preserved sense of agency and sense of responsibility under coercion, *PLOS ONE*, 16(10), e0258884, doi: [10.1371/journal.pone.0258884](https://doi.org/10.1371/journal.pone.0258884).

- CASPAR, Emilie A., CHRISTENSEN, Julia F., *et al.* [2016], Coercion changes the sense of agency in the human brain, *Current Biology*, 26(5), 585–592, doi: [10.1016/j.cub.2015.12.067](https://doi.org/10.1016/j.cub.2015.12.067).
- 5 CASPAR, Emilie A., CLEEREMANS, Axel, *et al.* [2018], Only giving orders? An experimental study of the sense of agency when giving or receiving commands, *PLOS ONE*, 13(9), e0204027, doi: [10.1371/journal.pone.0204027](https://doi.org/10.1371/journal.pone.0204027).
- CASPAR, Emilie A., GISHOMA, Darius, *et al.* [2022], On the cognitive mechanisms supporting prosocial disobedience in a post-genocidal context, *Scientific Reports*, 12(1), Article 1, doi: [10.1038/s41598-022-26460-z](https://doi.org/10.1038/s41598-022-26460-z).
- 10 CASPAR, Emilie A., IOUMPA, Kalliopi, *et al.* [2020], Obeying orders reduces vicarious brain activation towards victims' pain, *NeuroImage*, 222, 117251, doi: [10.1016/j.neuroimage.2020.117251](https://doi.org/10.1016/j.neuroimage.2020.117251).
- [2022], Commanding or being a simple intermediary: How does it affect moral behavior and related brain mechanisms?, *eNeuro*, 9(5), ENEURO.0508–21.2022, doi: [10.1523/eneuro.0508-21.2022](https://doi.org/10.1523/eneuro.0508-21.2022).
- 15 CASPAR, Emilie A., LO BUE, Salvatore, *et al.* [2020], The effect of military training on the sense of agency and outcome processing, *Nature Communications*, 11, 4366, doi: [10.1038/s41467-020-18152-x](https://doi.org/10.1038/s41467-020-18152-x).
- CASPAR, Emilie A., PECH, Guillaume P., *et al.* [2023], On the impact of the genocide on the intergroup empathy bias between former perpetrators, survivors, and their children in Rwanda, *American Psychologist*, 78(7), 825–841, doi: [10.1037/amp0001066](https://doi.org/10.1037/amp0001066).
- 20 CASPAR, Emilie A., VUILLAUME, Laurène, *et al.* [2017], The influence of (dis)belief in free will on immoral behavior, *Frontiers in Psychology*, 8, doi: [10.3389/fpsyg.2017.00020](https://doi.org/10.3389/fpsyg.2017.00020).
- 25 CHENG, Yawei, CHOU, Judith, *et al.* [2021], Psychopathic traits mediate guilt-related anterior midcingulate activity under authority pressure, *Scientific Reports*, 11(1), Article 1, doi: [10.1038/s41598-021-94372-5](https://doi.org/10.1038/s41598-021-94372-5).
- CIALDINI, Robert B. & GOLDSTEIN, Noah J. [2004], Social influence: Compliance and conformity, *Annual Review of Psychology*, 55(1), 591–621, doi: [10.1146/annurev.psych.55.090902.142015](https://doi.org/10.1146/annurev.psych.55.090902.142015).
- 30 CIKARA, Mina, BRUNEAU, Emile G., *et al.* [2011], Us and them: Intergroup failures of empathy, *Current Directions in Psychological Science*, 20(3), 149–153, doi: [10.1177/0963721411408713](https://doi.org/10.1177/0963721411408713).
- 35 CIORCIARI, John D & HEINDEL, Anne [2014], *Hybrid Justice: The Extraordinary Chambers in the Courts of Cambodia*, Ann Arbor: University of Michigan Press.

CLARK, Phil [2005], When the killers go home: Local justice in Rwanda, *Dissent*, 52(3), 14–21, doi: [10.1353/dss.2005.0052](https://doi.org/10.1353/dss.2005.0052).

— [2010], *The Gacaca Courts, Post-Genocide Justice and Reconciliation in Rwanda: Justice without Lawyers*, Cambridge: Cambridge University Press, doi: [10.1017/cbo9780511761584](https://doi.org/10.1017/cbo9780511761584).

COLL, Michel-Pierre [2018], Meta-analysis of ERP investigations of pain empathy underlines methodological issues in ERP research, *Social Cognitive and Affective Neuroscience*, 13(10), 1003–1017, doi: [10.1093/scan/nsy072](https://doi.org/10.1093/scan/nsy072).

DECETY, Jean [2011], The neuroevolution of empathy, *Annals of the New York Academy of Sciences*, 1231(1), 35–45, doi: [10.1111/j.1749-6632.2011.06027.x](https://doi.org/10.1111/j.1749-6632.2011.06027.x).

EL ZEIN, Marwa, DOLAN, Ray J., et al. [2022], Shared responsibility decreases the sense of agency in the human brain, *Journal of Cognitive Neuroscience*, 34(11), 2065–2081, URL [10.1162/jocn_a_01896](https://doi.org/10.1162/jocn_a_01896).

FOX, Nicole & NYSETH BREHM, Hollie [2018], “I decided to save them”: Factors that shaped participation in rescue efforts during genocide in Rwanda, *Social Forces*, 96(4), 1625–1648, doi: [10.1093/sf/soy018](https://doi.org/10.1093/sf/soy018).

FUJII, Lee Ann [2008], The power of local ties: Popular participation in the Rwandan genocide, *Security Studies*, 17(3), 568–597, doi: [10.1080/09636410802319578](https://doi.org/10.1080/09636410802319578).

— [2010], Shades of truth and lies: Interpreting testimonies of war and violence, *Journal of Peace Research*, 47(2), 231–241, doi: [10.1177/0022343309353097](https://doi.org/10.1177/0022343309353097).

GALLAGHER, Shaun [2000], Philosophical conceptions of the self: Implications for cognitive science, *Trends in Cognitive Sciences*, 4(1), 14–21, doi: [10.1016/s1364-6613\(99\)01417-5](https://doi.org/10.1016/s1364-6613(99)01417-5).

GALLESE, Vittorio & GOLDMAN, Alvin [1998], Mirror neurons and the simulation theory of mind-reading, *Trends in Cognitive Sciences*, 2(12), 493–501, doi: [10.1016/S1364-6613\(98\)01262-5](https://doi.org/10.1016/S1364-6613(98)01262-5).

GALLESE, Vittorio, KEYSERS, Christian, et al. [2004], A unifying view of the basis of social cognition, *Trends in Cognitive Sciences*, 8(9), 396–403, doi: [10.1016/j.tics.2004.07.002](https://doi.org/10.1016/j.tics.2004.07.002).

GERDES, Karen E., SEGAL, Elizabeth A., et al. [2010], Conceptualising and measuring empathy, *British Journal of Social Work*, 40(7), 2326–2343, doi: [10.1093/bjsw/bcq048](https://doi.org/10.1093/bjsw/bcq048).

- GRIGGS, Richard A. & WHITEHEAD, George I. [2015], Coverage of recent criticisms of Milgram's obedience experiments in introductory social psychology textbooks, *Theory & Psychology*, 25(5), 564–580, doi: [10.1177/0959354315601231](https://doi.org/10.1177/0959354315601231).
- 5 HAGGARD, Patrick, CLARK, Sam, *et al.* [2002], Voluntary action and conscious awareness, *Nature Neuroscience*, 5(4), 382–385, doi: [10.1038/nn827](https://doi.org/10.1038/nn827).
- HAGGARD, Patrick & WHITFORD, Ben [2004], Supplementary motor area provides an efferent signal for sensory suppression, *Cognitive Brain Research*, 19(1), 52–58, doi: [10.1016/j.cogbrainres.2003.10.018](https://doi.org/10.1016/j.cogbrainres.2003.10.018).
- 10 HAIDT, Jonathan [2003], The moral emotions, in: *Handbook of Affective Sciences*, edited by R. J. Davidson & K. R. Scherer, Oxford: Oxford University Press, 852–870.
- HAN, Shihui [2018], Neurocognitive basis of racial ingroup bias in empathy, *Trends in Cognitive Sciences*, 22(5), 400–421, doi: [10.1016/j.tics.2018.02.013](https://doi.org/10.1016/j.tics.2018.02.013).
- 15 HASLAM, S. Alexander & REICHER, Stephen D. [2017], 50 years of *Obedience to authority*: From blind conformity to engaged followership, *Annual Review of Law and Social Science*, 13(1), 59–78, doi: [10.1146/annurev-lawsocsci-110316-113710](https://doi.org/10.1146/annurev-lawsocsci-110316-113710).
- 20 HASLAM, S. Alexander, REICHER, Stephen D., *et al.* [2014], Nothing by mere authority: Evidence that in an experimental analogue of the Milgram paradigm participants are motivated not by orders but by appeals to science, *Journal of Social Issues*, 70(3), 473–488, doi: [10.1111/josi.12072](https://doi.org/10.1111/josi.12072).
- [2015a], “Happy to have been of service”: The Yale archive as a window into the engaged followership of participants in Milgram's “obedience” experiments, *British Journal of Social Psychology*, 54(1), 55–83, doi: [10.1111/bjso.12074](https://doi.org/10.1111/bjso.12074).
- 25 — [2015b], Shock treatment: Using immersive digital realism to restage and re-examine Milgram's “obedience to authority” research, *PLoS ONE*, 10(3), e109015, doi: [10.1371/journal.pone.0109015](https://doi.org/10.1371/journal.pone.0109015).
- 30 HEIN, Grit, SILANI, Giorgia, *et al.* [2010], Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping, *Neuron*, 68(1), 149–160, doi: [10.1016/j.neuron.2010.09.003](https://doi.org/10.1016/j.neuron.2010.09.003).
- HILKER, Lyndsay McLean [2012], Rwanda's “Hutsi”: Intersections of ethnicity and violence in the lives of youth of “mixed” heritage, *Identities*, 19(2), 229–247, doi: [10.1080/1070289x.2012.684441](https://doi.org/10.1080/1070289x.2012.684441).
- 35

- HINTON, Alexander Laban [1998], Why did you kill?: The Cambodian genocide and the dark side of face and honor, *The Journal of Asian Studies*, 57(1), 93–122, doi: [10.2307/2659025](https://doi.org/10.2307/2659025).
- 5 HODGES, Sara D., KIEL, Kristi J., *et al.* [2010], Giving birth to empathy: The effects of similar experience on empathic accuracy, empathic concern, and perceived empathy, *Personality and Social Psychology Bulletin*, 36(3), 398–409, doi: [10.1177/0146167209350326](https://doi.org/10.1177/0146167209350326).
- 10 HOENEN, Matthias, LÜBKE, Katrin T., *et al.* [2018], Empathic cognitions affected by undetectable social chemosignals: An EEG study on visually evoked empathy for pain in an auditory and chemosensory context, *Frontiers in Behavioral Neuroscience*, 12, 1–14, doi: [10.3389/fnbeh.2018.00243](https://doi.org/10.3389/fnbeh.2018.00243).
- HOLLEMAN, Gijs A., HOOGE, Ignace T. C., *et al.* [2020], The “Real-world approach” and its problems: A critique of the term ecological validity, *Frontiers in Psychology*, 11, Article 721, doi: [10.3389/fpsyg.2020.00721](https://doi.org/10.3389/fpsyg.2020.00721).
- 15 JAUNIAUX, Josiane, KHATIBI, Ali, *et al.* [2019], A meta-analysis of neuroimaging studies on pain empathy: Investigating the role of visual information and observers’ perspective, *Social Cognitive and Affective Neuroscience*, 14(8), 789–813, doi: [10.1093/scan/nsz055](https://doi.org/10.1093/scan/nsz055).
- 20 JONES, Adam [2002], Gender and genocide in Rwanda, *Journal of Genocide Research*, 4(1), 65–94, doi: [10.1080/14623520120113900](https://doi.org/10.1080/14623520120113900).
- KARCH, Susanne, MULERT, Christoph, *et al.* [2009], The free choice whether or not to respond after stimulus presentation, *Human Brain Mapping*, 30(9), 2971–2985, doi: [10.1002/hbm.20722](https://doi.org/10.1002/hbm.20722).
- 25 KELMAN, Herbert C. [1967], Human use of human subjects: The problem of deception in social psychological experiments, *Psychological Bulletin*, 67(1), 1–11, doi: [10.1037/h0024072](https://doi.org/10.1037/h0024072).
- KEYSERS, Christian & GAZZOLA, Valeria [2014], Dissociating the ability and propensity for empathy, *Trends in Cognitive Sciences*, 18(4), 163–166, doi: [10.1016/j.tics.2013.12.011](https://doi.org/10.1016/j.tics.2013.12.011).
- 30 KLEIN, Kristi J. K. & HODGES, Sara D. [2001], Gender differences, motivation, and empathic accuracy: When it pays to understand, *Personality and Social Psychology Bulletin*, 27(6), 720–730, doi: [10.1177/0146167201276007](https://doi.org/10.1177/0146167201276007).
- KÜHN, Simone, BRASS, Marcel, *et al.* [2013], Feeling in control: Neural correlates of experience of agency, *Cortex*, 49(7), 1935–1942, doi: [10.1016/j.cortex.2012.09.002](https://doi.org/10.1016/j.cortex.2012.09.002).
- 35 LOYLE, Cyanne [2009], Why men participate: A review of perpetrator research on the Rwandan genocide, *Journal of African Conflicts and Peace Studies*, 1(2), 26–42, doi: [10.5038/2325-484x.1.2.2](https://doi.org/10.5038/2325-484x.1.2.2).

- MANDEL, David R. [1998], The obedience alibi: Milgram's account of the Holocaust reconsidered, *Analyse & Kritik*, 20(1), 74–94, doi: [10.1515/auk-1998-0105](https://doi.org/10.1515/auk-1998-0105).
- 5 MCDOOM, Omar Shahabudin [2021], *The Path to Genocide in Rwanda: Security, Opportunity, and Authority in an Ethnocratic State*, Cambridge: Cambridge University Press, doi: [10.1017/9781108868839](https://doi.org/10.1017/9781108868839).
- MECK, Warren H. [2006], Neuroanatomical localization of an internal clock: A functional link between mesolimbic, nigrostriatal, and mesocortical dopaminergic systems, *Brain Research*, 1109(1), 93–107, doi: [10.1016/j.brainres.2006.06.031](https://doi.org/10.1016/j.brainres.2006.06.031).
- 10 MILGRAM, Stanley [1963], Behavioral study of obedience, *The Journal of Abnormal and Social Psychology*, 67(4), 371–378, doi: [10.1037/h0040525](https://doi.org/10.1037/h0040525).
- [1974], *Obedience to Authority: An experimental view*, New York: Harper and Row.
- 15 MILLER, Arthur G. [1986], *The Obedience Experiments: A case study of controversy in social science*, New York: Praeger.
- MIRONKO, Charles [2004], Igitero: means and motive in the Rwandan genocide, *Journal of Genocide Research*, 6(1), 47–60, doi: [10.1080/1462352042000194700](https://doi.org/10.1080/1462352042000194700).
- 20 MORRISON, India, LLOYD, Donna, *et al.* [2004], Vicarious responses to pain in anterior cingulate cortex: Is empathy a multisensory issue?, *Cognitive, Affective, & Behavioral Neuroscience*, 4(2), 270–278, doi: [10.3758/cabn.4.2.270](https://doi.org/10.3758/cabn.4.2.270).
- NACHEV, Parashkev, KENNARD, Christopher, *et al.* [2008], Functional role of the supplementary and pre-supplementary motor areas, *Nature Reviews Neuroscience*, 9(11), 856–869, doi: [10.1038/nrn2478](https://doi.org/10.1038/nrn2478).
- 25 NEDERHOF, Anton J. [1985], Methods of coping with social desirability bias: A review, *European Journal of Social Psychology*, 15(3), 263–280, doi: [10.1002/ejsp.2420150303](https://doi.org/10.1002/ejsp.2420150303).
- 30 PALUCK, Elizabeth Levy & GREEN, Donald P. [2009], Deference, dissent, and dispute resolution: An experimental intervention using mass media to change norms and behavior in Rwanda, *American Political Science Review*, 103(4), 622–644, doi: [10.1017/s0003055409990128](https://doi.org/10.1017/s0003055409990128).
- PECH, Guillaume P. & CASPAR, Emilie A. [2022], Can a video game with a fictional minority group decrease intergroup biases towards non-fictional minorities? A social neuroscience study, *International Journal of Human–Computer Interaction*, 40(2), 482–496, doi: [10.1080/10447318.2022.2121052](https://doi.org/10.1080/10447318.2022.2121052).
- 35

— [2023], Does the cowl make the monk? The effect of military and Red Cross uniforms on empathy for pain, sense of agency and moral behaviors, *Frontiers in Psychology*, 14, 1–12, doi: [10.3389/fpsyg.2023.1255835](https://doi.org/10.3389/fpsyg.2023.1255835).

PERRY, Gina [2013], *Beyond the Shock Machine: The Untold Story of the Milgram Obedience Experiments*, Melbourne: Scribe.

PRUNIER, Gerard [1998], *The Rwanda Crisis: History of a Genocide.*, London: C. Hurst & Co.

RUSHWORTH, Matthew F. S. [2008], Intention, choice, and the medial frontal cortex, *Annals of the New York Academy of Sciences*, 1124(1), 181–207, doi: [10.1196/annals.1440.014](https://doi.org/10.1196/annals.1440.014).

SINGER, Tania, SEYMOUR, Ben, *et al.* [2004], Empathy for pain involves the affective but not sensory components of pain, *Science*, 303(5661), 1157–1162, doi: [10.1126/science.1093535](https://doi.org/10.1126/science.1093535).

— [2006], Empathic neural responses are modulated by the perceived fairness of others, *Nature*, 439(7075), 466–469, doi: [10.1038/nature04271](https://doi.org/10.1038/nature04271).

SLATER, Mel, ANTLEY, Angus, *et al.* [2006], A virtual reprise of the Stanley Milgram obedience experiments, *PLoS ONE*, 1(1), e39, doi: [10.1371/journal.pone.0000039](https://doi.org/10.1371/journal.pone.0000039).

STRAUS, Scott [2007], What is the relationship between hate radio and violence? Rethinking Rwanda’s “Radio Machete”, *Politics & Society*, 35(4), 609–637, doi: [10.1177/0032329207308181](https://doi.org/10.1177/0032329207308181).

SYNOFZIK, Matthis, VOSGERAU, Gottfried, *et al.* [2008], Beyond the comparator model: A multifactorial two-step account of agency, *Consciousness and Cognition*, 17(1), 219–239, doi: [10.1016/j.concog.2007.03.010](https://doi.org/10.1016/j.concog.2007.03.010).

TAJFEL, Henri & TURNER, John [1979], An integrative theory of intergroup conflict, in: *The Social Psychology of Intergroup Relations*, edited by W. G. Austin & S. Worchel, Monterey: Brooks; Cole, 33–37.

TANGNEY, June P., STUEWIG, Jeffrey, *et al.* [2014], Two faces of shame: The roles of shame and guilt in predicting recidivism, *Psychological Science*, 25(3), 799–805, doi: [10.1177/0956797613508790](https://doi.org/10.1177/0956797613508790).

TIMMERS, Inge, PARK, Anna L., *et al.* [2018], Is empathy for pain unique in its neural correlates? A meta-analysis of neuroimaging studies of empathy, *Frontiers in Behavioral Neuroscience*, 12, Article 289, doi: [10.3389/fnbeh.2018.00289](https://doi.org/10.3389/fnbeh.2018.00289).

TRICOCHÉ, Leslie, ROVAI, Antonin, *et al.* [2024], When the brain says “No!”: An MRI study on the neural correlates of resistance to immoral orders, *PsyArXiv*.

- VARESE, Federico & YAISH, Meir [1998], *Altruism: The importance of being asked. The rescue of Jews in Nazi Europe*, Oxford: Sage Publications, URL <https://api.semanticscholar.org/CorpusID:221627026>.
- 5 XIE, Ying, CHEN, Mingliang, *et al.* [2016], Neural basis of two kinds of social influence: Obedience and conformity, *Frontiers in Human Neuroscience*, 10, Article 51, doi: [10.3389/fnhum.2016.00051](https://doi.org/10.3389/fnhum.2016.00051).
- YOSHIE, Michiko & HAGGARD, Patrick [2013], Negative emotional outcomes attenuate sense of agency over voluntary actions, *Current Biology*, 23(20), 2028–2032, doi: [10.1016/j.cub.2013.08.034](https://doi.org/10.1016/j.cub.2013.08.034).
- 10 YU, Hongbo, KOBAN, Leonie, *et al.* [2020], A generalizable multivariate brain pattern for interpersonal guilt, *Cerebral Cortex*, 30(6), 3558–3572, doi: [10.1093/cercor/bhz326](https://doi.org/10.1093/cercor/bhz326).
- ZAKI, Jamil [2014], Empathy: A motivated account, *Psychological Bulletin*, 140(6), 1608–1647, doi: [10.1037/a0037679](https://doi.org/10.1037/a0037679).
- 15 ZAKI, Jamil & MITCHELL, Jason P. [2013], Intuitive prosociality, *Current Directions in Psychological Science*, 22(6), 466–470, doi: [10.1177/0963721413492764](https://doi.org/10.1177/0963721413492764).
- ZIMBARDO, Philip [2007], *The Lucifer Effect: Understanding how good people turn evil*, New York: Random House.