The effects of acute stress on social decision-making

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der mathematisch-naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Adam Schweda

aus Rosenberg, Polen

Düsseldorf, 9. September 2021

aus dem Institut für Experimentelle Psychologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Tobias Kalenscher
2. Prof. Dr. Christian Bellebaum

Tag der mündlichen Prüfung: 13.12.2021

Content

Summary (in English)

Although social behavior is widespread in the animal kingdom, humans are equipped with an unprecedentedly large capacity to navigate effectively through their social realities. Our ability to cooperate with and rely on each other, share resources, and forego personal benefits for abstract social purposes has been crucial to our success in building large-scale communities. Yet, we also possess the ability to derogate from cooperative norms and harm other people, which is, it its extreme form, illustrated by a look at wars and altercations throughout human history. Stress can serve as a rapid and powerful driver of behavior. In fact, recent evidence indicates that human social decision making is altered in situations of acute stress. Yet, this evidence is far from being conclusive: on the one side, some portion of studies supports the "fight-or-flight" hypothesis, which presumes that stress increases the propensity to aggress against an attacker or flee, and hence, predicts that stressed decision makers are less concerned with the well-being of others and act more offensively. On the other side, some studies found evidence for the "tend-and-befriend" hypothesis, which postulates that, in times of hardship, individuals act more prosocially and invest into their social networks in order to, then, receive support from others. The aim of this dissertation is to shed light on the experimental circumstances under which stress actually increases or decreases prosocial tendencies. In study 1, my co-authors and I examined whether both, fight-or-flight and tend-and-befriend can occur simultaneously. Here, stressed participants (vs non-stressed controls) played an intergroup cooperation vs. competition game where both, benefitting the ingroup and harming the outgroup were viable choices. Although stress had no measurable direct effect on the allocation patterns, the results suggest that ingroup-friendly and outgroup-hostile investments are differentially modulated by hypothalamic-pituitary adrenal and sympathetic-adrenal-medullary activity, a finding that tentatively supports the idea of bidirectional effects of stress on social behavior. In study 2, we demonstrated that a small change in a decision's wording could boost prosocial choice towards socially remote individuals by taking advantage of the reluctance to enrich oneself at the expense of others. Specifically, social discounting curves were much shallower when the decision to share was framed as averting loss of the other's endowment ("take-frame") compared to generating a gain for the other at the cost of the player's own endowment ("give-frame"). Further analyses of functional imaging data show that these two types of decisions possess different neural signatures, which points toward distinct underlying processes. In study 3, we showed that stress is able to mitigate these framing effects. Here, a hybrid psychological and physical stressor left prosocial decisions unaffected under the give-frame, but stressed participants exhibited less restraint to benefit by withdrawing money

from socially remote others under the take-frame. These results support the notion that stress can reduce the willingness to comply with social norms that prohibit causing harm to others. All in all, we could further illustrate how manifold stress effects on social behavior can be. The manifest behavior of a stressed individual can, indeed, be a function of the predominantly active physiological process, properties within the decision maker, as well as the type of decision itself. Future scientific endeavors should take into account such intricacies when willing to explore when stress makes us nicer or meaner to each other. Our results make important contributions to a growing and exciting field of decision-making research that yields insights for areas wherever social interaction is essential, such as businesses, hospitals and law.

Summary (in German)

Obwohl soziales Verhalten im Tierreich weit verbreitet ist, sind wir Menschen mit besonderen Fähigkeiten ausgestattet, uns effektiv in unseren sozialen Realitäten zurechtzufinden. Unsere Fähigkeit zu kooperieren, uns zu vertrauen, Ressourcen zu teilen, sowie diese in abstrakte soziale Ziele zu investieren, hat zur Errichtung und zum Erhalt der großen Gemeinschaften beigetragen, in welchen wir leben. Allerdings besitzen wir auch die Fähigkeit, von sozialen Normen abzuweichen und unseren Mitmenschen Schaden zuzufügen. Die Extremen dieses Verhaltens werden durch einen Blick auf historische Auseinandersetzungen und Kriege verdeutlicht. Stress wiederrum ist eine mächtige Determinante unseres Verhaltens. Die jüngste Forschung zeigt, dass auch die soziale Entscheidungsfindung durch Stress verändert werden kann, jedoch ist man sich bezüglich der Wirkrichtung noch uneinig. Einerseits sprechen einige Ergebnisse für die „Fight-or-Flight“-Hypothese, welche besagt, dass ein Entscheider unter dem Einfluss von Stress weniger mit dem Wohlergehen anderer Menschen beschäftigt ist und deswegen mehr zu offensiven Handlungen tendiert. Andererseits sprechen verschiedene Befunde für die „Tend-and-Befriend“-Hypothese, welche davon ausgeht, dass wir unter Belastung dazu tendieren, uns prosozial zu verhalten, um in unsere sozialen Netzwerke zu investieren, von welchen wir wiederrum Hilfe und Unterstützung erhalten. Die dargelegte Dissertationsschrift befasst sich mit der Frage, unter welchen Umständen Stress prosoziales Verhalten wahrscheinlicher oder weniger wahrscheinlich macht. Im Rahmen der ersten Studie ergründeten meine Koautoren und ich, ob „Fight-or-Flight“ und „Tend-and-Befriend“ zur selben Zeit auftreten können. Hierbei konnten einzelne gestresste (vs. nicht gestresste) Spieler in einer Gruppe ihr Geld auf mehrere Weisen investieren, wobei sie sowohl ihrer eigenen Gruppe helfen als auch der anderen Gruppe schaden konnten. Obwohl es keinen direkten Effekt unserer Stressmanipulation auf das Gesamtinvestitionsmuster gab, weisen die Ergebnisse darauf hin, dass die Hypothalamus-Hypophysen-Nebennierenrinden-Achse und das sympathico-adreno-medulläre System unterschiedlich zu prosozialen und feindseligen Investitionsmustern beitragen. Diese Befunde sprechen für die Hypothese, dass Stress einen bidirektionalen Effekt auf prosoziales Verhalten haben kann. In der zweiten präsentierten Studie konnten wir demonstrieren, dass eine kleine Änderung in der Formulierung einer Entscheidung dazu führen kann, dass sich prosoziales Verhalten gegenüber sozial weit entfernten Menschen stark erhöht. Hierbei machten wir uns das Phänomen zunutze, dass Menschen mit starker Zurückhaltung reagieren, wenn sie sich auf Kosten anderer bereichern sollen. Im Einzelnen zeigte sich, dass sich der hyperbolische Abfall im Social Discounting stark verringerte, wenn die Entscheidung, einen bestimmten Betrag mit einer anderen Person zu teilen, als Verhinderung

eines Verlustes für ebendiese Person dargestellt wurde („take-frame“) als wenn die Teilentscheidung zu einer Erhöhung des Gewinns der anderen Person geführt hat („give-frame“). Weiterhin konnten Analysen funktioneller Hirndaten zeigen, dass diese Entscheidungen mit anderen neuronalen Signaturen einhergehen, was die Beteiligung unterschiedlicher Mechanismen vermuten lässt. In einer dritten Studie konnten wir zeigen, dass akuter Stress den oben beschriebenen „Framing-Effekt“ abschwächen konnte. Obwohl sich das soziale Entscheidungsverhalten im „give-frame“ durch den vorherigen Einsatz eines hybriden physiologisch-psychologischen Laborstressors unbeeinflusst zeigte, zeigten sich die gestressten Teilnehmenden weniger zurückhaltend mit Abzügen bei sozial distanzierten Menschen im „take-frame“, welche wiederrum zur Erhöhung des eigenen Gewinns geführt haben. Diese Ergebnisse unterstützen die Annahme, dass akuter Stress die Bereitschaft zu sozial-normgemäßem Verhalten verringert. Insgesamt konnte diese Arbeit erneut unterstreichen, auf welch vielseitige Weise sich akuter Stress auf soziales Verhalten auswirken kann. Das beobachtbare Verhalten unter Stress kann hierbei vom jeweilig dominanten physiologischen Prozess, den Eigenschaften des Entscheiders, sowie Eigenschaften der Entscheidung selbst beeinflusst werden. Künftige wissenschaftliche Bestrebungen, die sich der Ergründung der sozialen Dimension der Stressreaktion zum Ziel setzen, sollten diese Vielschichtigkeit in Betracht ziehen. Unsere Ergebnisse tragen zu einem spannenden und wachsenden Feld der Entscheidungsforschung bei, welches viele Tätigkeitsgebiete bereichern könnte, in welchen soziale Interaktionen zentral sind (z.B. in Unternehmen, in Krankenhäusern oder im Rechtsbereich).

Preface

Large-scale cooperation has made our species flourish. We share our possessions with family and friends because we feel emotionally attached, interested in their well-being and sometimes because we feel morally obliged to do so. We cooperate with acquaintances, strangers, and people we do not even particularly like, partly because strong interpersonal norms scaffold our everyday behavior. Beyond that, we undertake great investments for our societies' common good: although sometimes reluctantly, we pay taxes, insurance fees and social security contributions that pool the risks for the individual and offer support for people living through less prosperous times. We participate in weaving social, infrastructural and economic networks that make our communities and economies resilient.

Why do we do that? Why don't we simply cut ourselves off and live a life in solitude without any burdens and responsibilities? And how come we are doing so well in navigating through our social realities? After all, they require us to remain incessantly vigilant about the consequences of our and other peoples' actions, social signaling, conflicting ambitions and strategies. In fact, being social can cost us a great deal of energy. Evolution has equipped us with a large skillset to handle everyday social interactions, arguably because it increases our personal and inclusive fitness. Design features favoring kin selection – the propensity to share resources and incur costs for relatives - are widespread across different species. Elementary cooperative reciprocity – an exchange rule by which I share resources with individuals that also share resources with me - is not only exclusive to humans and primates. But hardly any other species than humankind has mastered the management and navigation of cooperation at such a large scale, often involving more than just a handful of individuals, social boundaries, considerations about multi-layered interests, and concerns about reputation and norms. With our long history of living in small-scale societies, our evolution has likely shaped our cognitive apparatus to make social exchange as efficient as it is.

Yet, in the everyday world's reality, social interactions do not only rely on concessions; especially in times of hardship, harm, punishment and free-riding start to govern social landscapes. Our history is full of demonstrations of our ability to cause harm on others. Intergroup conflicts are a prime example: individuals grant favor to members of the ingroup, but they harm the outgroup. Yet, creating disadvantages to other individuals can be more subtle: while the act of deliberately punishing someone is considered morally reprehensible, simply excluding her or him from the possibility to benefit is considered less so. In its worst form, systematically depriving certain groups form the opportunity to participate in common goods might lead to collateral damage.

But in the end, what makes us give or not give a hoot about others? And what makes us go berserk? What are signals that indicate the necessity of relentlessly protecting ourselves and our loved ones, even to the detriment of others? Acute stress has been discussed as one such behavioral driver; it signals the need to act, in the best case, quickly; it puts our body into a state of preparedness, mobilizing energy and making us more resistant. We feel our hair standing on end, our heart jumping out of our chest, the sweaty palms, the quavering voice. But that's not it: the stress response comes along with fast-paced *cognitive and emotional* adaptations to handle threats. In the social domain, scientists long assumed the "fight-or-flight"-response to be the dominant social response under stress. Fight-or-flight was hypothesized to diminish empathy, and increase feelings of anger, irritability and hostility. It rapidly prepares us to either escape or to attack the assailant in order to safeguard our physical integrity. Yet, and although fight-or-flight is somewhat intuitive for most of us, a growing body of theoretical and empirical evidence actually suggests the opposite. When under acute stress, we can also behave increasingly prosocial and generous. By doing so, we foster mutuality, support, and protection in our communities, and thus, create a social buffer against potential dangers. Although this so-called tend-and-befriend reaction is less intuitive, it has gained a fair bit of support in the recent years from numerous empirical observations that people shared more resources and exhibited more affiliation to others after stress.

"Fight-or-flight" and "tend-and-befriend" are concepts that appear mutually exclusive. Yet, we could be left inflexible and rigid if only one of these tendencies emerged at the same time: Merely seeking social support under threat would most likely result in demise. Indeed, the fact that fight-or-flight is dominant in non-human animals suggests that it could be the more evolutionarily stable coping strategy. On the other hand, falling into blind fury or mindless getaway might let inclusive fitness drop quickly, and even in the animal world, peers and potential mates can be a sparse commodity. This thesis aims at shedding light on when these two reactions occur, and whether they can even occur simultaneously, so that an increase in prosocial, tending attitudes towards family and friends, and antagonistic tendencies towards remote or potentially threatening individuals could drive behavior at the same time. If the latter held true, it would synergize a large body of empirical literature on social behavior under stress, as well as reports from the field, such as soldiers' increased sense of comradery after bloodcurling times on the battlefield, or violent criminals suffering from mental ill-health, leading to grueling hopelessness and chronic stress.

Thesis outline

So, under which circumstances do fight-or-flight and tend-and-befriend occur? And can they even act in symphony? The current literature does not yet answer the *when's and how's* about us behaving in a friendly, antagonistic or even mixed fashion; research is dispersed across disciplines, and no unifying framework exists. Indeed, research on stress and social behavior in humans is, compared to other domains, still in its early stages, and the presented studies attempt to complement existing concepts to open up new avenues for future investigations.

The first two sections comprise a very basic introduction into social behavior, the stress response, and social decision making under stress. I will directly transition to my very first published research study which is presented in section 3. Here, my co-authors and I exploited a possible bi-directional effect of psycho-evaluative stress on parochial altruism to directly test the hypothesis of co-occurrence of fight-or-flight and tend-and-befriend. We applied a popular intergroup competition game where contributions to different investment options operationalized the willingness to help team members or engage in group conflict. Participants formed a team of three players and were allowed to either invest in a prosocial, ingroup-oriented fashion, or choose to diminish the outgroups' payout at the same time. To make ingroup-outgroup boundaries more distinct, players were informed that members of the ingroup were partisans of the same political party, while outgroup members were assumedly recruited from the supporters' base of the *Alternative fuer Deutschland*, a right-wing populist party that gained considerable negative attention at that time in Germany. Contrary to our expectations, we found evidence for the absence - or at least a negligibility - of an effect of stress on social intergroup preferences, either toward prosociality or antagonism. Yet, a few exploratory analyses draw support for the idea of co-occurrence of fight-or-flight and tend-and-befriend under stress: While saliva cortisol was linked to increased ingroup-friendly investments, heart rate – a proxy for sympathetic nervous activity – was correlated with outgroup-hostility. In section 4, I will discuss these findings thoroughly. One of the most reasonable conclusions was that the overall reluctance to cause harm to others might have masked possible stress effects. This was most likely due to the adherence to social norms – a key determinant of everyday social behavior - which will serve as a basis for the transition to the next two projects I have worked on. In section 5, I proceed by introducing a study in which my co-authors and I developed and tested an economic game where direct harm is tested against discrimination. The novel, modified version of the Social Discounting paradigm was designed to test social choice under a so called give- and a take-frame. In intermittent trials, players could choose to act prosocially by either sharing money from their own initial endowment with another person, or by leaving money with the other person when given the opportunity to avail themselves of the endowment of the other person. Indeed, our data suggest a strong reluctance to take money from other peoples' initial endowment, resulting in a boost in generosity under the take-frame compared to the give-frame, even if the payoff matrix remains the same under both frame conditions. In section 6, I return to the main question of how stress elicits shifts in social preference. I present data from an experiment that investigates how a hybrid stressor actually changes other-regarding investment behavior generated by the above-mentioned framing effect. Participants were subjected to a socio-evaluative and physical stressor before completing the framing version of the Social Discounting task. We found that stress diminishes the framing effect as such, but only for prosocial choices toward socially remote others. In other words, participants in the stress condition showed a diminished framing-related augmentation of sharing behavior at large social distances in the take-frame compared to non-stressed controls. This suggests that stress might, indeed, offset harm aversion to a certain extent, a finding that rather hints toward a fight-or-flight response. Yet, when considered in conjunction with prior literature, the data yield at least some support for the co-occurrence of fight-or-flight and tend-and-befriend: while stress triggered a prosocial reaction toward socially close individuals in former studies, it can also elicit antagonism toward socially distant people by decreasing the adherence to the “do-not-harm"-principle. In section 7, I will discuss the implications of our results within a broader context, and suggest further avenues to efficiently accelerate research on stress and social behavior.

1. On love, sticks, and stones

Our predisposition for fine-tuned, flexible and inclusive cooperation is central in the evolution of humankind. From the early beginnings, hominins have engaged in coordinated social endeavors to secure their basic needs for security and mutuality. From rudimentary and rather pragmatic, yet oftentimes hierarchical groups of our primate ancestors, we have developed to form multifariously organized communities of scavengers, hunters and gatherers. To do so, we learned to distribute our resources, to build and foster cohesive groups, but also to spot, punish and excommunicate free-riders if necessary. In the following subsections, I will introduce the reader to a number of fundamental concepts from within the evolution of altruism. Then, I will present an introduction into theoretical and empirical literature on why we actually form supportive groups, and why the tendency to do so might have co-developed with our propensity to also participate in intergroup conflicts.

1.1. Causes vs. realizations

Cooperation is widespread in the animal kingdom. Ranging from insects (Choe et al., 1997) to sea creatures (e.g. Kuczaj et al., 2015), from avians (Ligon, 1983) to vertebrates (Dublin, 1983; Rand & Nowak, 2013; Schneeberger et al., 2012); even microorganisms in the rhizosphere (e.g. Kiers & Denison, 2008) show behaviors that appear to create mutual benefit for the agents involved, whether directly or indirectly. Not uncommonly, such behaviors even pass species borders (e.g. Eckardt & Zuberbühler, 2004). Examples such as insect workers' celibacy for the sake of nurturing their queens' offspring (Oster & Wilson, 1978; Wenseleers & Ratnieks, 2006), communal breeding in birds (Stacey & Koenig, 1990) or grooming alliances in capuchin monkeys (O'Brien, 1993) clearly show that altruism and cooperation can manifest in a variety of ways. Yet, a different realization of altruistic behaviors in different species does not negate the existence of one core functional design feature. In other words, the biological system which enables the respective behavior – the proximate causation - does not equal the reason why such behavior exists – the ultimate causation - in the first place (Laland et al., 2011; Mayr, 1961; Scott-Phillips et al., 2011). Apparently different traits can be the result of one set of ultimate causes. Then again, a similar trait might be realized by many different proximate mechanisms (Wilson, 2015).

1.2. A gene-based approach: inclusive fitness

Natural selection favors those traits that enable an organism higher reproductive success, which is equal to evolutionary fitness (Darwin, 1859; Dawkins, 1982; Fisher, 1930). Yet, the imperative that an individual must maximize their own offspring in any case stumbles upon scenarios where genetically related individuals interact, and where an individual might also increase the likelihood of passing on a part of their own genetic setup to the next generation by supporting related others, who, in turn, succeed at reproducing (Scott-Phillips et al., 2011). In other words, besides pursuing one's own reproduction, one would help a more or less genetically related individual to increase its reproduction success in order to maximize one's *inclusive fitness*. By doing so, one's genes increase in frequency directly and indirectly. This idea was first put forth by Hamilton (Hamilton, 1964, 1970) as an ultimate core for the evolution of altruistic behaviors across a wide range of species. Hamilton made this explicit by defining a feature to be passed on if the product of the relatedness r and the benefit for the recipient of a behavior b exceeds the cost in the fitness of the individual c (and thus, it proliferates when $rb < c$, which is termed Hamilton's Rule). Prosocial behaviors toward genetically related individuals could realize either via (i) kin discrimination (i. e. perceiving characteristics that signal direct relatedness), (ii) a limited local dispersal of genes (e. g. neighbors will carry a similar set of genes if partner choice is limited by spatial location) or (iii) by identification of an observable trait in non-relatives that is a manifestation of an allele, which, in turn, propagates in case of mutual prosociality (the so-called green-beard effect,Hamilton, 1964; West et al., 2007b). Accordingly, inclusive fitness maximization is not limited to individuals with similar ancestry. Notice that the terms *inclusive fitness* and *kin-selection* are very closely related in that kin-selection encompasses cooperation with close relatives (Smith, 1964). Yet, depending on the definition of *kin*, it includes instances when individuals with genetic overlap interact, even without any co-ancestry (West et al., 2007a). Importantly, inclusive fitness is a fundamental, ultimate design principle that is used to explain prosociality in general, irrespective of species and biological mechanism.

1.3. Scratching backs: reciprocity and its generalization

Another account of prosocial behavior considers individual altruistic behavior as costly on the short term, but beneficial on the long term given that the initial recipient reciprocates at a later occasion (Trivers, 1971). Trivers' example (1971) of a drowning man being rescued by a savior, who in turn is saved by the former one later, clearly illustrates the direct fitness benefits:

reciprocal support leads to the mitigation of individual risks by creating a social buffer, which then increases the likelihood to survive and reproduce on both sides. Such trait does not need any form of relatedness to be beneficial (and therefore selective), at least if the long-term benefit is greater than the initial cost (Lehmann & Keller, 2006). Hence, helping the other is more likely when the probability of meeting the cooperation partner at a later timepoint - and therefore benefiting from previous helpful encounters - is high. But what if the recipient does not reciprocate? The problem of cheating and free riding becomes greater with increasing cost primarily incurred by the cooperating individual. The maintenance of reciprocal altruism is tightly linked to the implementation of enforcement mechanisms (Trivers, 1971; West et al., 2007a, 2007b). Such enforcement mechanisms include additionally rewarding cooperation (e.g. via reputation, see e.g. Gintis et al., 2001; Lehmann & Keller, 2006; Trivers, 1971), conditionally sanctioning and punishing free riding behavior, and ostracism. These enforcement mechanisms increase the likelihood of creating an environment of collaborative individuals, either by excluding free riders from the pool of cooperators, or perhaps even turning them into such (West et al., 2007b). In the laboratory, conditional enforcement mechanisms have, indeed, shown to bring derailing cooperation among dyads and groups of individuals back on track (Fehr & Gächter, 2002; Leibbrandt & López-Pérez, 2012; Lohse & Waichman, 2020).

Enforced reciprocity appears to play a very particular role in human societies (André, 2014; Stevens & Hauser, 2004). Yet, these behaviors are not enough to explain a widespread prosocial human tendency. Although human societies are thoroughly organized these days, not every life domain is governed by binding contracts. Many so-called non-binding agreements exist, and a mere reciprocator would see all sorts of opportunities to free-ride or cheat. For instance, when taking a taxicab in a big city, the likelihood of cooperating with the taxi driver in the future is low (Basu, 1984; example from Fehr et al., 2002). This is a typical one-shot Prisoner's dilemma situation, where the driver initially offers to cooperate and expects the passenger to go along with his offer. Yet, for the passenger, free-riding is easy. She or he could just escape and, if uncaught, just rejoice about a free ride, while for the driver, chasing the passenger would oftentimes be energy-inefficient, at least. A defection like this is what an outcome-maximizing individual would prefer. However, taxi companies would most likely not exist if passengers would regularly refuse to reciprocate the initial cooperation offer from the cab driver. Also, the cab driver would likely flip out and chase the bill-dodger, even though the cost-benefit analysis is subpar. This ubiquitous tendency to be nice to reciprocators while punishing non-reciprocators even though it does not yield any immediate or future benefit is termed strong reciprocity (Fehr et al., 2002; Gintis, 2000). Compared to weak reciprocity – where an actor only maximizes long term outcomes – strong reciprocity can actually explain long-term maintenance of large groups of individuals where most of our encounters are one-off. The evidence for strong reciprocity is compelling: for instance, actors are often deviating from equilibria (e.g. Mengel, 2018), as well as they routinely punish other individuals although no immediate or long-term reward is recompensating the initial cost for doing so (e.g. Bernhard et al., 2006; Buckholtz et al., 2008; Jordan et al., 2016; Zhou et al., 2017). Yet, to explain how such traits might have been selective across evolution, a larger perspective must be adopted.

1.4. Playing the orchestra: on the selection of groups

From the early days on, humans have lived, and were socialized to live in communities (Gintis et al., 2015). This is not unique to humans, but the symphony of social and political mechanisms and the formation of oftentimes egalitarian groups compared to rather hierarchical co-habitation in the animal kingdom definitely is (Gintis et al., 2015). Living in a group is advantageous: it offers protection from predators, increases foraging success and offers a social buffer in terms of individual risk. In fact, there could be a lot at stake for an individual that becomes ostracized. When relegated, a weak and more vulnerable individual might even face death. Groupthink might be deeply engrained in us: from the so-called “refugee crisis” in Europe over the political split of the Trumpian age in the US, to the deep-rooted COVID-19-conspiracies, modern history shows us that it is still a part of our everyday lives. Ingroup favoritism, where a member of a sometimes arbitrarily formed group is showing preferences for the benefit for members of the ingroup over members of the outgroup is easily producible and reproducible in the lab (Böhm et al., 2020; Tajfel et al., 1971).

Living in small societies makes up a large part of human phylogeny. This is why evolutionary theorists have come up with the notion that selection might not just take place at the individual level, but also at the group level (Wynne-Edwards, 1964). Accordingly, individuals from successful and well-coordinated groups are simply more likely to reproduce. Selection especially favors groups of altruistic individuals. This kind of pressure would, in turn, favor the development of traits that support and maintain the collective. If, for instance, a group consisted of selfish outcome maximizers, a common good could be overexploited. In this case, the evolution of reticence as a trait would avoid such overexploitation (West et al., 2007a). Although this idea sounds comforting, it is not too early to say that in its classical form, group selection has been subject to extensive criticisms. The key arguments include that for such group selection to work, individuals would need to heavily self-constraint in terms of mating individuals of the outgroup, and that this is just not what is empirically observed (West et al., 2007a). Still, the basic notion that groups or communities can drive evolution served as a basis for theories that followed. Contrary to the “old group selection” (West et al., 2007a), multilevel selection has loosened the assumption that selection invariably takes place on the level of groups (Wilson & Sober, 1994). Rather, group level selection and individual level selection might occur at the same time, and the result depends on how these two evolutionary drives are orchestrated (Wilson, 2015). As a very simple example one could imagine two groups, one which is full of hyper-competitive, anti-social psychopaths among which co-habitation is almost impossible, and the other one which is cooperative, altruistic and collective-oriented. An altruist would most likely vanish quickly in the former, while a vicious individual would profit from plentiful free riding opportunities in the latter. Selfishness outperforms altruism on an individual level. Yet, and although the altruistic group is more prone to be infiltrated and ripped off by free riders, it performs better at foraging and grants its members security via its coordinated effort. Hence, on the group level, altruistic groups outperform selfish groups (Wilson, 2015). Competition takes place at two levels: individuals compete within a group, but groups compete between each other. This type of multi-directional pressure explains the heterogeneity of traits and time-inconsistency and variability of human social behaviors (Wilson, 2015). An individual can be loving, caring, loyal, friendly, helpful, and sometimes even overly submissive although obviously mistreated. The same person can appear totally egomaniac, manipulative and up for his own good a few moments later. Multilevel theories give an intuitive indication for why we follow stringent cultural norms while feeling less connected to other groups.

Despite the broad application frame of multilevel selection approaches, they suffer from a number of drawbacks on the theoretical level (for a discussion, see Gardner, 2015; Kramer & Meunier, 2016). At the moment, no undisputed unifying formalism for the theory exists. Also, studies differ in how they specify groups, as well as the selection and adaptation thereof (Kramer & Meunier, 2016; Okasha, 2016). Over time, it was oftentimes argued that multi-level selection offers a subpar alternative to the more established inclusive fitness or kin-selection theory, particularly in the light of fundamental extensions thereof (Lehmann & Rousset, 2014, 2020; Marshall, 2015). As mentioned in section 1.2., kin-theorists nowadays define relatedness in a wider fashion. Inclusive fitness theory has been generalized and requires less assumptions in more modern forms (Gardner et al., 2011). It is able to predict a large variety of empirical observations in terms of the cost-benefit ratio of altruism in relationship with relatedness (Kramer & Meunier, 2016). Some theorists even consider multi-level selection and inclusive fitness formally equivalent if between-group selection pressure is higher than within-group selection pressure in multilevel selection (Kramer & Meunier, 2016; Marshall, 2011; Okasha, 2016; West et al., 2007a). At the very least, both can be suitable for describing different scenarios and complement each other.

For the purpose to give the reader a comprehensible introduction to the phylogeny of human altruism, these explanations are sufficient for the moment. Whether groups form because locally concentrated individuals maximize their inclusive fitness, or because they were circumscribed entities during human evolution is part of an ongoing discussion. In fact, human groupthink and community behavior still enjoys great research interest. This is not only because of the positive side of cooperation and collective endeavors. We can, indeed, be very mean to each other. This happens particularly if we don't know each other and if our vis-a-vis is member of an outgroup, particularly one that we dislike.

1.5. Intergroup conflicts

At its extreme, a group conflict can end up in war. In such situation, anger toward the outgroup prevails up to a point where some conflict parties lose a large part of their resources and suffer casualties. Intergroup conflicts were observed in other animals, too. During the Gombe chimpanzee war (Goodall, 2010; see also Glowacki et al., 2017), a tiny collective of chimpanzees splintered from a larger group and relocated to a new territory nearby, until members of the initial group started to invade the new habitat. Over the course of four years, they killed all male chimpanzees from the splinter group and acquired their territory, only to be, ironically, driven away by another chimpanzee community later. Human raids and wars have already taken place between early human hunter-gatherer communities (Gintis et al., 2015). A recent archeological finding of skeletons bears witness to a violent confrontation in Kenya about 10.000 years ago (Lahr et al., 2016). Resentments can be very persistent: Some of the world's historic conflicts, such as the Reconquista in Europe, have had wreaked havoc for hundreds of years.

Speaking from an evolutionary perspective, intergroup conflict is intuitively inconsistent with the idea of individual fitness maximization. On the one hand, an actor incurs fitness costs for the sake of alliance-building, augmenting the own group's resources and fostering the group's fitness in terms of defensive and offensive capabilities. On the other hand, engaging in a conflict can bear tremendous costs to the group, its resources and, in the end, the actor itself. Conflicts are oftentimes not even zero-sum, but negative-sum games.

Classic evolutionary theories of social behavior do hardly account for persistent engagement in larger-scale conflicts and have difficulties to explain why humans regularly end up in these loose-loose situations. Multi-level selection models presume conflict, but do not explain its cause and dynamics (Choi & Bowles, 2007). This is why, recently, evolutionary theorists have come up with the idea of “parochial altruism” to describe a general human tendency to unconditionally act in an ingroup-friendly and outgroup-hostile manner. The concept originates from the observation of surprisingly frequent occurrences of war-like altercations in human history (Bowles, 2006), an empirically firm fundament of literature on ingroup-biases, as well as the above-mentioned lack of regard thereof from the side of classical evolutionary theories on human cooperation (Bernhard et al., 2006; Choi & Bowles, 2007). But could a so apparently unfavorable disposition even proliferate throughout evolutionary history? Choi and Bowles (2007) suggested favorable phylogenic conditions for parochial altruism in late pleistocenic and early holocenic times, where environmentally-caused scarcity of resources coerced human communities into fierce competition. These circumstances might have reduced within-group selection in favor of within-group maintenance and between-group selection, and hence, increased the pressure toward the development of altruistic and parochial traits. As a proof-of-concept, Choi and Bowles (2007) ran a set of agent-based models with four types of actors: outgroup-tolerant altruists, parochial (outgroup-intolerant) altruists, outgroup-tolerant non-altruists and parochial non-altruists. Selection pressure was two-levelled. On the first level, actors competed within groups. In accordance with other multi-level accounts, non-altruists beat altruists in a public goods game because non-altruists profited from the public good without incurring the cost of investing themselves. This, in turn, increases their overall fitness and reproductive success. On the second level, group encounters could either lead to a benevolent exchange of resources, or an outbreak of a conflict. The outcome was determined by the number of parochial altruists within both groups. In case of a peaceful encounter, non-parochial individuals profit as a function of the number of non-parochial individuals of the other group and vice versa. Parochial individuals receive nothing. However, a group engaged in conflict if the number of parochial altruists in their own group exceeded the number of parochial altruists in the other group. Whether a group wins or loses also depends on how many parochial altruists it has in their rows. In case of a victorious battle, parochial altruists gained a fitness benefit. In case of a loss, a set of parochial altruists and civilians (non-parochial altruists) vanished from the group.

The analysis of the agents’ behavior across tens of thousands of generations revealed that there were two critical attractors, one in which outgroup-tolerant non-altruists prevailed, and one in which altruists, parochial non-altruists and parochial altruists made up the largest part of the population. The increase of the proportion of outgroup-tolerant non-altruists was characterized by an overall low incidence of intergroup conflicts. Outgroup-tolerant non-altruists could benefit from intergroup-encounters, while still harvesting the benefits from defecting in the ingroup public goods games. Groups with these types of compositions could, however, not stand up to bellicose conditions when conflicts were more common. Under these scenarios, the parochial altruists, altruists and parochial non-altruists gained relative fitness benefits and their number swiftly increased. The more conflicts there were, the more parochial altruists and non-altruists, as well as pure altruists there were. This was due to the gain in fitness in case of a conflict, as well as the possibility to avoid conflict in case of equal group strength. Choi and Bowles (2007) interpret their results as a proof of feasibility for the proliferation of a parochial altruist trait. Parochial altruists could have been crucial actors in periods of sharp conflicts. They increase their groups' competitiveness during altercations, prevent its eradication, or they prevent getting into a conflict at all. Importantly, besides a hostile stance toward outgroups, parochial altruism equally requires a certain level of ingroup-oriented altruism to enable enough willingness to self-sacrifice to actually act in an outgroup-hostile manner. Hence, the coevolution of these traits as a general parochial module could have been fostered in times of war. Whether this compound trait exists remains to be answered empirically.

1.6. Parochial altruism in the lab

The idea of parochial altruism proposes an ultimately causal explanation for persistence of conflict among humans. It furthermore serves as an evolutionary foundation for modern psychological theories of intergroup behavior in humans. These theories usually describe proximate mechanisms on a psychological level instead of making phylogenic assumptions about the roots of intergroup behaviors. For instance, one account suggests that individuals with high levels of social dominance orientation strive for superiority of their ingroup, which predisposes them to hostile attitudes toward outgroups (Sidanius & Pratto, 2001). The realistic group conflict theory (Bornstein, 2003; Campbell, 1965), on the other side, proposes that when resources are scarce, individuals in groups hastily form a zero-sum perception, which is why even loosely composed groups adopt animosity. It was based on the famous Robber's Cave study (Sherif et al., 1961), where two groups of adolescent boys were exposed to regular competitive scenarios that, in the end, fostered active aggression between the two groups. Then again, for the social identity theory, no real competition is necessary to create intergroup biases (Böhm et al., 2020; Tajfel et al., 1979).

Individuals form a social identity based on the most salient categorization, i. e. as an individual or as a part of a group. Then, they naturally attempt to strengthen a positive self-image by increasing their or their group's distinctiveness relative to others or outgroups. If the positive image of an ingroup is threatened and the individual cannot leave the group, she or he strives for competition. Tajfel and colleagues' (1971) famous experiments showed that even an arbitrary criterion for group assignment could produce ingroup favoritism and outgroup discrimination. Further extensions and integrations of these frameworks exist (e.g. the self-categorization theory, Turner et al., 1987; see Böhm et al., 2020 for a review). At last, the generalized bounded reciprocity theory (Yamagishi & Kiyonari, 2000) is somewhat located at the intersect of proximate and ultimate explanations. This theory – in its simplicity – assumes that actors expect ingroup members to reciprocate more than non-ingroup members due to previous positive experiences, which is why they show ingroup favoritism as a default heuristic. It opposes the idea that social identification is the key determinant, which is supported by studies showing that ingroup favoritism is diminished if group interdependence is offset e.g. by letting the outgroup decide on the player's outcomes (Rabbie et al., 1989).

What makes parochial altruism so special, though, is that it assumes an unconditional tendency toward ingroup love *and* outgroup hate (at least in its classical definition, cf. Böhm, 2016), and does not require the individual to be directly or indirectly recompensated for its costs to benefit the ingroup, e.g. via the value of reputation as a warrior. Much experimental and field research exists that indicates robust intergroup biases. They occur in many sorts of economic games when cooperation partners are assorted to different groups, such as the prisoners' dilemma (see e.g. Dion, 1973; Yamagishi et al., 2008), the dictator game (e.g. Whitt & Wilson, 2007), the trust game (e.g. Smith, 2011), and the ultimatum game (e.g. Gillespie et al., 2013). They occur in minimal groups (e.g. Güth et al., 2009; Turner et al., 1979) and real groups (e.g. Whitt & Wilson, 2007; see also Balliet et al., 2014 for a meta-analysis). Yet oftentimes, the evidence predominantly points toward mere ingroup favoritism, but not outgroup-directed hostility. In many cases, the designs do not include any option to actively harm other players. For instance, Ber-Ner and colleagues (2009) found that participants shared more of their initial endowment in a dictator game with players from the ingroup than players from different outgroups. In a study by Smith (2011), players of a trust game sent less money to outgroup members than ingroup members. In both cases, there is a favoritism toward the ingroup that leaves the outgroup discriminated, but not actually harmed. Similar observations were made using psychometric assessments of attitudes toward in- and outgroup members, collective allocation decisions, as well as the measurement of overt helping behavior (Böhm et al., 2020). Yet, the emergence of the concept of parochial altruism has inspired much research on actual intergroup conflict, shifting the focus from mere ingroup favoritism toward measurable forms of harm to the outgroup.

One of the first empirical research studies on parochial altruism was conducted by Bernhard and colleagues (2006) in a realistic group setting with members of two groups in Papua New Guinea. The two groups were not in a conflict during the time of the experiment. Participants played a third-party punishment game, in which a dictator decided to send a receiver a specific amount of money that she or he received as an endowment, just as in a dictator game. A third player (the third party) could observe the transaction and decide whether she or he would give up on a part of her initial endowment to punish the dictator for not sharing enough with the receiver. Bernhard and colleagues manipulated the group memberships of all three players and found that the punishment of the dictator was most pronounced when the receiver and the third party were belonging to the same group. In other words, third party players who observed one of their own group members to be unfairly treated by an outgroup member tended to engage most in costly punishment. At the same time, dictators sent more to the receivers if they belonged to the same group. These results capture both core aspects proposed by parochial altruism: the tendency to favor ingroup members while also being ready to act at the outgroup's detriment, although not unconditionally.

Further experimental research operationalizes ingroup love and outgroup hate as allocations within and between teams of players, which is closest to the intergroup conflict model by Choi and Bowles (2007) in terms that there is space to capture multiple opposing motivations, either to maximize own payoff, the own group's payoff, and to harm the outgroup. Team games can, hence, comprise a direct measure of parochial altruism (Böhm et al., 2020). Abbink and colleagues (2012), for instance, modelled intergroup conflict in a game with two teams in which each player must decide how much of an initial endowment she or he invests into a pool of war expenditures. These war expenditures then determined whether the team wins or loses the team contest. In case of a win, the team received bonus monetary units. Indeed, more prosocial individuals invested more into the conflict expenditure pool, which indicates a relationship between ingroup love and outgroup hate, as put forth by parochial altruism.

In a very similar fashion, the original Intergroup Prisoner's Dilemma by Bornstein (IPD, Bornstein, 1992; Bornstein et al., 1994) modelled intergroup conflict in a way that contributing to a between-group pool instead of keeping an endowment for oneself increases the likelihood of a bonus for the team, but this depended on the number of outgroup-members investing into the between-group pool. If, for instance, all members of a team consisting of three players decided not to keep the endowment and contribute, while players from the other team kept their endowments, each contributing team player received 18 monetary units. This bonus decreased as a function of contributions from the other team: If one player from the other team decided to invest into the between-group pool, the amount each player from the first team receives dropped to 15, and so on. Bornstein and colleagues (1992; 1994) found average between-group pool investment rates of about 20% to 40% depending on whether the IPD is administered as a repeated or one-shot game. Yet, the difficulty here is that the motif to increase the own group's payoff and to prevent the other group from earning gains is hard to distinguish. This is why Halevy and colleagues (Halevy et al., 2008) extended the original IPD game to better operationalize these motifs.

As in the IPD, players of the Prisoner's Dilemma – Maximizing Differences game (IPD-MD) received an initial endowment, which they could keep or distribute. Yet, the IPD-MD contained a pure within-group pool into which players invest if they strive to maximize their group's payoffs, and, thus, act altruistically toward the ingroup. The amount invested into the within-group pool was multiplied by three and allocated across all three team members, including the player. Here, a player that invests 5€ generates a 2.50€ payoff (50% of 5€) for each team member. An investment into the between-group pool had the same effect, but the amount that is granted to the team members was deduced from the outgroup. Imagine that, for example, a group member invests 5€ into the between-group pool. 2.50€ (50% of 5€) are deduced from every outgroup member, while each ingroup member receives 2.50€. Investments into the between-group pool should directly operationalize parochial altruism. It is a costly investment to the benefit of the ingroup that also harms the outgroup. Although the results from Halevy and colleagues (2008) eventually indicated low contribution rates to intergroup conflict, the paradigm has established as a benchmark for the measurement of parochial altruism as it could assess both, ingroup-love and outgroup-hate simultaneously. In the following years, the paradigm could successfully be used to measure increases in intergroup conflict after, for instance, the deployment of signals of high competitiveness from the outgroup (Halevy et al., 2010), in realistic group settings (at least in a modified take-version, Weisel & Böhm, 2015), as well as when between-group pool investments could pre-emptively decapacitate the outgroup (Böhm et al., 2016). The intranasal administration of oxytocin, on the other hand, increased within-group pool investments (De Dreu et al., 2010). Yet, and all in all, studies show that overall intergroup aggression is low compared to ingroup-beneficial contributions.

These results are somewhat sobering in terms of finding overt and unconditional outgroup hate, which is, at the first glance, not conductive for the concept of parochial altruism. Generally, aggression directed toward outgroups is rarely observed in the laboratory (Balliet et al., 2014; Böhm et al., 2020). Also, a putative correlation between ingroup love and outgroup hate was not found consistently (Thielmann & Böhm, 2016). Yet, outgroup hate might simply be covered by strong everyday norms. Humans appear to be generally harm-averse (Decety & Cowell, 2018), which is why under normal circumstances, a “do-not-harm" principle (Baron, 1995) might prevail. Amidst a highly structured and interdependent society, this reluctance might be highly advantageous: aggression might backfire in form of retaliations, ostracism, sanctions or punishments. An individual might lose valuable reputation which leads to future dismissals from potential cooperation partners. Still, intergroup conflict is recurring, again and again. So what, in the end, could evoke outgroup hate?

A very central protective mechanism to the integrity of the organism is the stress response. Stress physiologically prepares the agent to appropriately behave in presence or in anticipation of a threat. It makes us run faster, act faster, and it shifts the global functional state of our brain toward a rather habit-based mode (Hermans et al., 2014). Accordingly, stress has been found to modulate a large number of affective and cognitive domains. In the last 10 years, behavioral and neuroscientific research has provided fundamental insights in the area of stress and social decision making. Hence, stress could be an appropriate candidate to look at when deciding on what triggers aggressive intergroup behavior. But first, what do we know about stress, and why does it even affect cognition?

2. Stress and social behavior

In the next sections, I will provide the reader a comprehensive but short introduction on what stress is, how it manifests biologically and how it interacts with our brain and behavior. Then, I will put a particular focus on very recent literature that aims at the central question of this thesis, namely what stress does to our propensity to act socially, both in positive and negative ways.

2.1 The physiology of the stress response

The stress response has served us very well in our evolutionary history. It is a wide-ranging response that includes a multitude of organic processes to save us from threats, and mobilize the necessary energy responses to fight, escape or simply do the things that make us more resilient under threat. Stress is old, adaptive, and very well conserved: relatives of the human stress system can be found in other vertebrates like rats (Jameel et al., 2014), fish (Iwama, 1998) and reptiles (Tokarz & Summers, 2011). Stress usually kicks in quickly. It is designed to help us in situations of acute danger, such as an attack, a fight or a non-animate natural threat (Cannon, 1932). Yet, stressors can be versatile: they might include basic irritation like hypoxia or loud noises, but also overwhelming social situations and frustration (Selye, 1950). Classically, two primary systems are assumed to play a key role in the stress response: the Sympatho-Adreno-Medullary Axis (SAM) and the Hypothalamic-Pituitary-Adrenal Axis (HPA).

The former one primarily acts via the sympathetic branches of the autonomous nervous system. Here, the organism enters a catabolic and ergotrophic state. As a result, the heartbeat accelerates, the bronchi widen, the glucose and lipid production is stimulated, and the blood and nutrient supply of the muscles increases, which allows for an overall higher energy expenditure (McCorry, 2007). The pathway starts at the level of preganglionic neurons in the central segments of the spine (T1 to L2). The response is then mostly transmitted to a larger number of post ganglionic neurons, which then transmit signals to the respective target organs. Other preganglionic neurons directly travel to the adrenal medulla where the secretion of adrenaline (80%) and noradrenaline (20%) is elicited. These transmitters are distributed via the blood flow (McCorry, 2007). The sympathetic nervous response oftentimes discharges the entire system at once. A shutdown of the sympathetic response is usually accompanied by the antagonistic parasympathetic system (McCorry, 2007).

Noradrenaline also plays an important role in the brain's ascending arousal system (Richerson et al., 2012). Noradrenergic neurons primarily originate in the locus coeruleus and travel toward cortical and limbic areas, where they modulate wakefulness and attention. The locus coeruleus, however, also projects to the spine (Richerson et al., 2012). While there are some up- and downstream efferences mostly from the ventral medulla, adrenaline plays a subordinate role in the brain. The brain's arousal system also includes acetylcholinergic, serotonergic, dopaminergic and histaminergic pathways (Richerson et al., 2012).

The HPA responds somewhat slower, but the response is maintained for a longer time (de Kloet, 2014). The cascade starts with the hypothalamic medial paraventricular nucleus releasing the corticotropin-release hormone (CRH), which is then transmitted to the anterior pituitary via a network of small capillaries. There, the CRH docks on G-protein coupled receptors, which, in turn, release the adenocorticotropic hormone (ACTH) into the blood. ACTH then elicits the secretion of gluco- and mineralocorticoids in the adrenal glands (primarily in the zona fasciculata and the zona glomerulosa, respectively). Corticosteroids provoke gluconeogenesis and proteolysis to enhance energy supply (Simmons et al., 1984) and are effective throughout a large proportion of organic systems (Timmermans et al., 2019). They also assist with replenishment of energy after acute stress (Joëls et al., 2011). Corticosteroids pass the blood-brain barrier (Sudheimer et al., 2014). As opposed to catecholamines, corticosteriods can supply the entire brain, but they are most effective in areas where corticoid-receptors concentrate. Two main types of receptors exist: the mineralocorticoid (MR) and the glucocorticoid receptor (GR). Both of them concentrate in the hippocampus, the paraventricular nucleus, and the septum. MRs aggregate more in the amygdala and the locus coeruleus. GRs show higher disaggregation. MRs have a higher affinity than GRs (Joëls & Baram, 2009).

Both stress systems have a particular, yet not fully distinct temporal profile. The sympathetic nervous system acts within milliseconds, and noradrenaline is quickly available in the brain, but its activity also reduces relatively quickly. On the other hand, corticoids start to slowly increase and reach a peak after about 30 minutes. Then, they decrease again. There is a time window where both substances – catecholamines and corticoids - take effect simultaneously (Hermans et al., 2011; Joëls et al., 2011). They might, indeed, act in a synergistic fashion (Roozendaal et al., 2006).

The stress response can even be more versatile. Non-genomic effects of corticosteroids are assumed to last for about an hour (Quaedflieg & Schwabe, 2018). Whereas short-term effects of corticosteroids are assumed to promote the cognitive stress response, including increased levels of memory and attention, late genomic effects appear to support energy replenishment, recovery and homeostasis (de Kloet et al., 2008; Henckens et al., 2010; Riis-Vestergaard et al., 2018). Various other signaling pathways are involved in the stress response. Among others, neurotransmitters like serotonine and dopamine play an important role (Richerson et al., 2012). CRH has its own modulating properties, and there are non-negligible interactions between the stress-related systems and cannabinoid receptors, oxytocin, vasopressin, ghrelin, as well as sex hormones (Cota et al., 2007; Joëls & Baram, 2009; Stephens et al., 2016). Hence, we do certainly not possess the ability to control for each sub-process of the stress response, but this does not restrain us in making assumptions about its functional and systemic effects on our thoughts, beliefs, emotions and behaviors.

2.2. The brain on stress

The way the stress response affects the brain already suggests wide-ranging modulations. By its design, the nature of the stress response serves to increase the likelihood of survival. Indeed, our cognition and affect undergo a massive shift in order to allow for situation- and purpose-specific and appropriate behavior, most likely accompanying us long after cessation of the stressor.

The effects of stress are measurable straightaway, and it reaches a large number of processing domains. On the level of attention, there is evidence that stress causes an attentional bias toward threat-related stimuli in high CORT-responders (Roelofs et al., 2007; Tsumura & Shimada, 2012), and that it enhances visual processing of negatively valent stimuli (Wirz & Schwabe, 2020). In the aftermath of a stressor, the amygdala-centered limbic network shows a prolonged activation (van Marle et al., 2010), which could be a correlate of the increased vigilance toward biologically relevant cues (van Marle et al., 2009). Overall, an acute stressor upregulates a salience processing network - consisting of limbic, insular, temporoparietal, inferotemporal and cingulate regions - at the expense of activation in the rather frontally and frontolaterally located executive network (Hermans et al., 2014). It boosts memory encoding and consolidation (Quaedflieg & Schwabe, 2018, although memory generalizes less under stress, see Dandolo & Schwabe, 2016). Stressful events are usually biologically relevant, and an organism should remain cautious when engaging in similar situations in the future. Hence, stress is assumed to be a ‘teaching signal’ that drives the formation of new memories under unpredictability (Trapp et al., 2018). Apart from just establishing and encoding signals of danger, it specifically facilitates the formation of habit-based memories at the cost of cognitive and goal-oriented learning (Quaedflieg & Schwabe, 2018; Wirz et al., 2018), and makes these memories more resistant to updating (Raio et al., 2017). Furthermore, some studies suggest that stress suppresses retrieval of previous – and potentially interfering - memories (Atsak et al., 2016; Quaedflieg & Schwabe, 2018).

Stress also modifies reward processing (Burani et al., 2020; Kinner et al., 2016; Porcelli et al., 2012). There is evidence that some types of acute stress and hydrocortisone administration increase the preference for smaller, sooner rewards in temporal discounting paradigms (Delaney et al., 2014; Haushofer et al., 2018, 2021; Riis-Vestergaard et al., 2018), although this is not found consistently (Haushofer et al., 2013). Things appear a bit more complicated with risk-related decision making. Some studies have found more risky decision making under stress (or after administration of cortisol, e.g. Putman et al., 2010; see Starcke & Brand, 2016). Yet, others observed less, time- or frame-dependent (Pabst et al., 2013a, 2013b; Porcelli & Delgado, 2009) or even unchanged levels of risk taking (Sokol-Hessner et al., 2016; Von Dawans et al., 2012, 2019). Margittai, Nave and colleagues (2018) found decreased loss aversion after administration of hydrocortisone and the $\alpha 2$-adrenoceptor antagonist yohimbine, but such effects appear to be dose-dependent (Metz et al., 2020). Hence, and with regard of all these profound effects of stress on human cognition, it is of little surprise that stress can change how we interact with others.

2.3. The social brain on stress

We all know how it feels to be stressed, and perhaps we have a precise image of what our reaction pattern becomes like. Intuitively, most of us would say that we tend to become more dismissive, grumpy, and angered. However, recent research has shown that it is not that simple. As with the specificity of cognitive domains that are modified, stress appears to have idiosyncratic effects on social behavior.

As a matter of fact, the idea that stress produces a reaction that is interactional by nature lies at the very core of the early 20th century stress research. Walter Cannon (1932) viewed the stress response as preparatory to either attack an enemy or escape the threatening situation. This “fight-or-flight" reaction was assumed to be primarily triggered by the sympathetic nervous system, as well as adrenaline and noradrenaline in the adrenal medulla (Cannon, 1914; as a side note: corticosteroids were discovered later by Reichstein, Hench and Kendall, which they received the 1950’s Nobel Prize for). The notion that stress mobilizes resources in order to guarantee dominance or survival has proven extremely fruitful. In rodents, acute stressors like omission of reinforcers, encounters with an intruding animal, immobilization via restraint of movement, as well as chronic stressors like regular exposure to mild stress and social isolation increase aggressive behavior (Sandi & Haller, 2015). Similar findings exist for dogs (Sandi & Haller, 2015) and primates (Virgin Jr & Sapolsky, 1997). In humans, early studies were able to capture relationships between economic hardship and lynching of people of color in the southern U.S. states in the late 19th to early 20th century (Berkowitz & Harmon-Jones, 2004; Hovland & Sears, 1940). Acutely stressed women tend to react in a more punishing manner toward their children

(Berkowitz & Harmon-Jones, 2004), and the effect of stress on criminal behavior is now widely recognized in criminology (Agnew, 2017). Also, results from recent longitudinal studies indicate increased martial conflict (Timmons et al., 2017) and fan violence (Shadmanfaat et al., 2021) after the occurrence of hardship and daily stressors. Hence, the fight-or-flight reaction is solidly backed up by empirical findings in animals and by studies from non-experimental fields.

Only recently, the development of reliable laboratory stress induction methods has enabled researchers to study the social effects of stress under more controlled conditions. The picture generated here is, however, less consistent. Although some research corroborated the fight-or-flight hypothesis, other studies do not conform with it, even showing effects that are directed to the opposite. In support of fight-or-flight, Vinkers and colleagues (2013) found more rejection rates in a sample of participants that performed an ultimatum game directly after being exposed to a common laboratory stressor. Furthermore, in the same study, stressed participants donated less to a non-governmental organization in a dictator game. FeldmanHall and colleagues (2015) found that stress decreases trust in a repeated trust-game. Boehnke and colleagues (2010; see also von Dawans et al., 2021) showed evidence for more aggression after administration of oral hydrocortisone in female participants, and stress was shown to reduce the consideration for others in a social risk game (Bendahan et al., 2017).

Other studies adopt a very different tone. They, for example, found evidence that stress modifies basic perceptual and recognition thresholds of social stimuli, such as certain facial expressions (Domes & Zimmer, 2019; von Dawans et al., 2020, 2021). Self-rated aversiveness of presented pain in other participants, self-rated emotional contagion, as well as BOLD-signals encoding vicarious pain perception were shown to increase in stressed individuals compared to non-stressed controls (yet, note that some of these effects were gender-dependent; Gonzalez-Liencres et al., 2016; Tomova et al., 2014, 2017; von Dawans et al., 2021; Wolf et al., 2015, see also Buruck et al., 2014). Salivary cortisol was revealed to be associated with feelings of closeness toward an interaction partner (Berger et al., 2016).

Accordingly, research exploring overt social interaction behavior after stress was also able to show an increase in prosocial tendencies. Takahashi and colleagues (2007) found significantly increased proportions of sharing in a dictator game. Von Dawans and colleagues (2012) could replicate this finding, and additionally found that participants were more trustful and trust-reciprocating in a trust game after stress. This was later replicated in a purely female sample (Von Dawans et al., 2019, although here, trustfulness was not significant). Such results are, generally, not explainable with the fight-or-flight theory, but rather conform with the very opposite. Tend-and-befriend (Taylor, 2006; Taylor et al., 2000) describes a behavioral pattern that, contrary to fight-or-flight, promotes the relationship with others with the aim to foster a supportive social buffer. It is assumed that an individual in a threatening situation profits from affiliations, especially if the group she or he is attached to actually offers protection and, hence, reduces the individual's risk to incur harm during perilous encounters. It is originally based on the observation that mothers would face an enormous cost if they just escaped a stressful situation leaving their offspring alone, which is why in its first postulation, tend-and-befriend was only referred to as a reaction that mostly female individuals adopt due to their special role in parenthood.

In general, fight-or-flight and tend-and-befriend are two competing frameworks. They describe very different ultimate mechanisms for opposing social reactions under stress. Although tend-and-befriend was rather conceptualized as an inherently female reaction, the experimental literature suggests that it also pertains to males. In fact, both these theories are supported by a solid amount of evidence: although it appears that tend-and-befriend is found more often in experimental paradigms, fight-or-flight is more established in non-experimental and the animal literature. But given the harsh opposition of these behaviors, how can we reconciliate them?

2.4. Stress as a (potential) mechanism that drives parochial altruism

Fight-or-flight and tend-and-befriend both modulate social behavior. They either downregulate the social preference for the wellbeing of the others or increase it. But if stress should support us to make adaptive, survival-promoting decisions, why couldn't it boost both? If fact, as Taylor (2006) proposes, a mere fight-or-flight reaction would lead to a mindless getaway at a huge cost and unexploited social resources. Especially in the context of inclusive fitness and the reproductive value of survival of genetic relatives, the individual would forgo a lot if fellows were forgotten about. On the other side, an actor that only reacts with tend-and-befriend could lose her or his life without actually being able to protect the others. An adaptive stress reaction should allow for both: the increase in prosocial tendencies toward individuals that are valuable in terms of inclusive fitness and reciprocation, and the ability to react adequately if a conflict is inevitable. These are, indeed, opposing motifs, but given the body of evidence for the very specific and adaptive nature of the stress response (see e.g. Quaedflieg & Schwabe, 2018), a bi-directional response would not be unreasonable.

That said, the empirical literature could perhaps give us first indications for bidirectionalities like these: the finding that stress increases fan aggression (Shadmanfaat et al., 2021) somewhat supports the idea of stress-induced parochialism because it includes ingroup cohesion, as well as outgroup hate. Also, a recent study on chimpanzees shows that engagement in intergroup competition appears to decrease the spread of group members, which is interpreted as better intragroup cohesion (Samuni et al., 2020). Laboratory studies give us some first clues, too. Margittai and colleagues (2015) applied a psychosocial stressor before participants performed a Social Discounting Task (Jones & Rachlin, 2006). Participants played several runs of the dictator game, but receivers varied in terms of their social distance, i.e. the perceived social closeness between the participant and the recipient. After undergoing a psychoevaluative laboratory stressor, stressed participants, indeed, showed a tend-and-befriend response. Yet this time, an increase in generosity was only observable toward socially close, but not distant receivers. In other words, stressed participants shared more money with people they perceived as socially close – usually their partners, parents, siblings or best friends – than socially remote people like nodding acquaintances or complete strangers. Furthermore, there was a negative association between salivary alpha-amylase (which is a marker for sympathetic activity) and generosity toward socially remote people in a separate condition that assessed social discounting 90 minutes after cessation of the social stressor. These results could be conceptually replicated in a further psychopharmacological intervention study by Margittai, van Wingerden and colleagues in 2018. Yet here, while the single administration of hydrocortisone alone increased generosity toward socially close people, the simultaneous administration of yohimbine – an α2-adrenoceptor antagonist which increases noradrenaline in the brain – offsets this effect. Hence, cortisol appears to play a key role in eliciting the tend-and-befriend response, while noradrenaline shuts it down. This is somewhat in accordance with Cannon (1932), who considered catecholamines central to the fight-or-flight response. Yet of course, both studies did not include any option to actively harm other players.

Here, the circle might close. Parochial altruism involves a friendly and prosocial attitude toward ingroup members, and a hostile stance toward outgroup members. A look at the ambiguities in stress research might, indeed, indicate that stress has the potential to do both, increase and decrease prosocial behavior. Hence, stress might induce and drive parochial altruism. This prediction is valuable from two perspectives: on the one side, if proven to hold, it would unify numerous divergent findings from the entire landscape of studies on stress and social behavior in humans. On the other side, it would provide a missing condition for outgroup spite in the lab, which would open up new opportunities to study social behavioral dynamics, in general. In fact, human participants are very reluctant to harm each other (see e.g. Halevy et al., 2008), and stress could be an evolutionarily salient trigger for the preparedness to actually aggress.

In the next section, I will summarize my very first study that directly examined stress-effects on parochial altruism in the laboratory. Here, my co-authors and I directly tested whether a classical laboratory stressor increased parochial altruism. Groups of three participants performed an Intergroup Prisoners' Dilemma – Maximizing Differences game (Halevy et al., 2008). In order to increase group cohesion and create a minimum, malleable level of readiness to aggress against the outgroup, natural political group constellations were exploited: participants were instructed that the group they were playing with consisted of dominant supporters of their own preferred party, and that they played against a previously invited group of supporters of the German right-wing party *Alternative für Deutschland (AfD)*. Such manipulations to increase intergroup salience were successfully used before (see e.g. Hackel et al., 2014; Weisel & Böhm, 2015). Before playing the IPD-MD, participants in the stress group underwent a so-called Group Trier Social Stress Task (Von Dawans et al., 2011) - a nowadays widespread and reliable procedure to induce psychoevaluative stress. Participants in the control group, on the other hand, underwent a very similar procedure which was matched in terms of the chronological sequence and verbal activity of the experimental group, but without the actual psychoevaluative component that causes stress. Most research that studies human behavior and cognition under stress considers male participants for the lack of fluctuations in their hormonal milieu. In order to generalize across sexes, we also included female participants, but additionally collected saliva samples of the sex hormones testosterone, estradiol and progesterone to control for potential moderators of stress effects.

Stressed participants did not differ from the control group in their within- and between-group allocations. However, further exploratory results suggest that cortisol increase was linked to ingroup-friendly investments when conditioned on testosterone. On contrary, heart rate increase was linked to a shift from ingroup-friendly to outgroup-hostile contributions. Although these results do not suggest direct effects of stress on intergroup preferences, they somewhat support the notion that different components of the stress response drive divergent social responses. We discuss the results in light of the existing literature on parochial altruism, outgroup-hate, and the neurochemical environment required to possibly evoke such effects.

3. Summary of study 1

The effects of psychosocial stress on intergroup resource allocation

Adam Schweda[1], Nadira Sophie Faber[2,3], Molly Crockett[4] & Tobias Kalenscher[1]

1: Comparative Psychology, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany.

2: College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom.

3: Oxford Uehiro Centre for Practical Ethics, University of Oxford, Oxford, United Kingdom.

4: Department of Psychology, Yale University, New Haven, Connecticut, USA

Published in: Scientific Reports (2019)

(Please note that this section presents a summary of my own research, so that various parts of the text are equal to the original publication. The original article can be found in the appendix.)

3.1. Introduction

Stress is known to alter social behavior. Fight-or-flight responses (Cannon, 1932) prepare an organism for antagonistic situations (Goldstein, 2010), thus increasing the individual's propensity to aggress and flee. For example, in humans, stress has been shown to reduce empathy (Negd et al., 2011) and generosity (Vinkers et al., 2013), and increase violence (Klaw et al., 2016; Silver & Teasdale, 2005) and criminal behavior (Agnew, 2001). Fight-or-flight responses go along with activation of the sympathetic nervous system, and they are linked to rapid-acting catecholaminergic, mainly noradrenergic (NA) components of the stress response (Allen et al., 2018; Cannon, 1932; Geen & O'Neal, 1969; Miczek et al., 2017; Nelson & Trainor, 2007).

However, recent evidence suggests that stress can also induce prosocial behavior (Taylor, 2006). For example, von Dawans and colleagues (2012) found that psychosocial stress increased males' trust and sharing behavior. This was interpreted as support for the "tend-and-befriend" (Taylor, 2006) hypothesis, which is a proclaimed coping strategy that involves investing into social networks after stress, thus offering help to a group of people in order to seek and offer mutual protection (Buchanan & Preston, 2014; Taylor, 2006; Taylor et al., 2000; Von Dawans et al., 2012). The tend-and-befriend hypothesis has received much empirical support in recent years. For instance, stress has been shown, across sexes, to increase donation rates among participants with pro-environmental attitudes (Sollberger et al., 2016), to increase trust and sharing behavior in male participants (Von Dawans et al., 2012), and generosity in males (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018). Consistent with the tend-and-befriend hypothesis, stressed individuals report more social closeness (Berger et al., 2016), and socio-evaluative stress has been shown to enhance emotional empathy (Wolf et al., 2015). Furthermore, induction of psychosocial stress leads to increased generosity towards socially close recipients of help but not towards socially distant others (Margittai et al., 2015).

There is suggestive evidence that prosocial tend-and-befriend responses to stress are linked to relatively slow-acting cortisol (CORT). For example, exogenous manipulation of CORT activity has been shown to foster financial altruism towards socially close others (Margittai, van Wingerden, et al., 2018), and stress-related CORT-levels covary with greater trust (Steinbeis et al., 2015) and social affiliation (Sollberger et al., 2016). The idea that the separate components of the physiological stress response have dissociable effects on social behavior has been supported by the recent observation that CORT-related financial altruism could be counteracted by additional administration of yohimbine (Margittai, van Wingerden, et al., 2018). This discovery is in line with the finding that noradrenergic activity correlates negatively with overall financial generosity (Margittai et al., 2015) and even implicit intergroup bias (Terbeck et al., 2012, 2015, 2016).

Thus, existing evidence suggests that stress can promote fight-or-flight as well as tend-and-befriend, and that these two tendencies are tentatively related to distinct components of the physiological stress response. However, it is currently unclear when and why stressed individuals show tend-and-befriend or fight-or-flight behavior. Here, we propose that stress boosts both tendencies at the same time by supporting prosocial behavior towards socially close others (tend-and-befriend), who, unlike strangers, can potentially provide comfort and support in stressful times (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018). Simultaneously, stress could foster aggression against socially distant outgroup members who are more likely to present a threat than ingroup members (fight-or-flight). Because of recent evidence for a role of CORT in promoting generosity towards others (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018; Steinbeis et al., 2015; Tomova et al., 2017), and the classic association of fight-or-flight tendencies with sympathetic activation, we further hypothesize that tend-and-befriend and fight-or-flight tendencies are modulated by the dissociable actions of the stress-neuromarkers CORT and NA, with CORT promoting prosociality towards ingroup members, and NA fostering aggressive behavior against outgroup members.

To test these hypotheses, we induced psychosocial stress in 100 male and 102 female participants (total n = 202), using the group version of the Trier Social Stress Test (gTSST, Von Dawans et al., 2011). After performing the gTSST or control procedures, participants played an adapted version of the intergroup prisoner's dilemma maximizing differences game (IPD-MD, Halevy et al., 2008). In this game, three participants were assigned to one group, and they were told they would play against another group that participated on the previous day. To manipulate group affiliation, participants were instructed that the members of their own group held similar political views, and that the members of the other group held radically opposing political views. At the beginning of the game, each participant received an initial economic endowment, which they could distribute across three pools. Contributions to the *keep-pool* would be kept by the participants; 50% of the total sum of contributions to the *within-group pool* would be paid out to each in-group member, including the participant; contributions to the *between-group* pool had the same effect to the ingroup, but each outgroup member would lose the amount each ingroup member received. Hence, contributions to the keep-pool can be interpreted as the motivation to maximize own profit (own-utility maximizing), and contributions to the within-group pool can be interpreted as costly motivation to maximize ingroup profit (called "ingroup love" in the relevant literature, e.g. Halevy et al., 2008). Finally, contributions to the between-group pool can be interpreted as motivation to maximize ingroup profit and, at the same time, harm the outgroup (so called "outgroup-hate"). Moreover, since male participants are known to respond to stress differently than females (Soldin & Mattison, 2009; Ter Horst et al., 2012) and often reveal different, gender-dependent social preferences (Rand et al., 2016; Soutschek et al., 2017) we additionally considered gender in our analyses, as well as the sex hormones testosterone, estradiol and progesterone (Casto & Edwards, 2016; Geniole et al., 2017; Losecaat Vermeer et al., 2016), and a range of other state and trait variables.

We predicted that stressed participants would contribute more money to the between-group pool than non-stressed participants, reflecting the predicted combination of ingroup-love and outgroup-hate, and that contributions to this pool would be correlated with measures of the sympathetic stress response (salivary marker of NA and heart-rate). Furthermore, we expected that the motivation to contribute to either the within-group or the between-group pool over keep-pool investments would be correlated with the amplitude of the salivary CORT response, reflecting the predicted CORT effects on prosociality.

3.2. Methods

One hundred and three male, and 105 female participants were recruited within the Düsseldorf (Germany) area. Three male and three female subjects were excluded. With a final sample size of n = 202, our experiment is comparably well-powered. Participants required to hold at least moderate sympathy for one of Germany's five political parties with seats in the German parliament, except the right-wing populist party (*Alternative für Deutschland*, AfD). We applied several further eligibility criteria (see supplemental online material). To exclude potential confounds and ensure similarity between participants, we collected a number of trait measures before the laboratory experiment (see supplemental online material).

We tested all participants in groups of three. Socio-evaluative stress was induced using the Trier Social Stress Test for groups (gTSST, Von Dawans et al., 2011). In the stress condition, participants were videotaped while exposed to a fictional job interview and, subsequently, to a mental arithmetic task in presence of other participants and in front of an evaluation panel of experimenters. In the control condition, participants prepared a short talk about their friends, and they also performed an arithmetic task, but the evaluation panel paid ostensibly no attention to the participants. Participants spoke simultaneously and they were not videotaped.

We collected multiple saliva samples to determine baseline measures of the sex hormones progesterone, estradiol and testosterone, and to quantify the impact of our stress manipulation based on CORT and α-amylase. Three separate sex-hormone samples were collected in the first half of the experiment before subjection to the stress or control procedure. In total, five CORT/α-amylase samples were collected throughout the entire experiment. As an additional measure of sympathetic activity, we recorded heart-rate (HR) at several time points during the experiment using a commercially available HR-monitor. Participants indicated positive and negative mood by completing a Positive and Negative Affect scale (PANAS; Watson et al., 1988) before, during and after the gTSST/control procedure, and, they also indicated current feelings of shame, insecurity, stress and confidence on visual analogue scales before, during and after the experimental procedures (VAS, 1–100).

We used political preferences to induce intergroup rivalry. The three participants in each testing session were instructed that they would form a group and play against another group of three other participants who performed the game one day before. The instructions explicitly stated that all members of the participants' own group held similar political views and that the members of the other group were supporters of the political party "*Alternative für Deutschland*". Outgroup decisions were not real and shammed by the experimenters. We used an adapted version of the IPD-MD to simulate intergroup behavior in the laboratory (Halevy et al., 2008; Weisel & Böhm, 2015). As mentioned above, in this game, two groups of three participants play against each other. Each participant receives the initial monetary endowment of 10 EUR, which they can freely distribute between three pools. Money contributed to the first pool (the "keep" pool), is kept by the player. Fifty percent of the total sum of contributions to the second pool (the "within-group" pool) are paid out to each ingroup member, including the participant. If a participant contributes 5 EUR to the within-group pool, each ingroup member, including the participant, would receive 2.50 EUR payback. Thus, contributions to the within-group pool are potentially costly to the participant because she only receives a back-payment of 50% of the invested sum if no one else contributes, but the overall sum of all payoffs to all group members is higher than individual contributions to the keep pool. Contributions to the third pool (the "between-group"-pool) have the same effect to the ingroup members as contributions to the within-group pool, but each outgroup member loses the amount each ingroup member receives. For example, if a participant invests 5 EUR to the between-group pool, each ingroup member, including the participant, would receive 2.50 EUR, and each outgroup member would lose 2.50 EUR. Thus, contributions to the between-group pool represent the same social dilemma as contributions to the within-group pool, but additionally entail the possibility to harm the outgroup. The experiment was incentive-compatible.

To analyze IPD-MD distribution patterns, we applied a 2 × 2 × 2 mixed-factorial analysis of variance (ANOVA) with pool as a repeated-measures (within-group vs. between group) as well as condition (stress vs control) and gender (male vs female) as between-subject factors. To complement this analysis, we ran additional Bayes-Factor analyses on a mixed linear model with pool and condition as fixed factors and subject as a random intercept (using the software JASP; The JASP Team, 2018). Furthermore, full Bayesian parameter estimation of the same model was used to estimate posterior parameter distributions and to yield information on credible parameter values (R-Package *brms;* Bürkner, 2017).

3.3. Results and discussion

To rule out systematic stress-unrelated differences between participants of the stress and control groups, we collected a range of individual trait measures. None of the trait measures differed significantly between groups, with the exception of chronotype and a marginally significant group difference in chronic stress. Moderation analyses with these two factors as potential moderators revealed no significantly influencing role on any of the outcome variables.

Compared to controls, participants in the gTSST group had significantly elevated CORT and salivary α-amylase levels. In addition, the subjective stress measures also revealed higher levels of psychological stress, such as negative affect, feelings of shame, and insecurity. Hence, the stress manipulation was successful.

To assess our main hypotheses, we computed mixed ANOVAs to test for effects of condition (stress vs. control) and gender on investments into the within-group and between-group pools. Consistent with earlier findings (De Dreu et al., 2015; Halevy et al., 2008), subjects invested significantly more into the within-group pool than the between-group pool. Male participants contributed more to both pools than female participants, i.e., female participants kept more money for themselves. However, and contrary to our predictions, stress did neither significantly increase within-group, nor between-group pool investments (main effect of condition: $F(1, 197) = 0.04$, $p = 0.844$, $\eta_p^2 < 0.001$, 95% CI = [0, 0.019]; interaction effect condition x pool: $F(1, 197) = 0.13$, $p = 0. 0.720$, $\eta_p^2 = 0.001$, 95% CI = [0, 0.025]). Please note that the latter effect sizes' 95% confidence interval upper bound can still be considered a small effect.

To further receive a realistic distribution of plausible stress-related effects and, hence, further elucidate our null finding, we additionally computed Bayesian credibility intervals, which have been shown to produce high coverage of true parameters (Nalborczyk et al., 2018). The results indicate comparably narrow credibility intervals around zero for stress-related effects, except for the slight increase of investments into the between-group pool (condition: $\beta = 0.02$, 95%-CrI [−0.16, 0.20]; pool x condition: $\beta = 0.08$, 95%-CrI [−0.28, 0.43]). Further equivalence tests ("region of practical equivalence", ROPE; Kruschke, 2018) based on the stress-related posteriors' 95%-highest density intervals (HDI) accept the null hypothesis if defined as HDIs congruent with an interval of ~20% (for the main effect of stress) or ~45% (for the interaction pool x stress) of the grand standard deviation around zero (Kruschke, 2018). To corroborate these results, we applied Bayesian hypothesis testing to obtain a quantitative estimate of the evidence for the null hypothesis. Taking into account the reverse of the Bayes factor in favor of the alternative hypothesis - the BF01 - we find evidence for our null hypothesis (only condition: BF01 = 8.523, "moderate" evidence; condition and the interaction term condition x pool: BF01 = 51.476, "very strong" evidence, Kass & Raftery, 1995), and this result remains stable for a wider array of prior definitions (see supplemental online material). Thus, given the centeredness of the posterior distributions of the stress-related effect-estimates around zero, as well as the Bayes Factors for the null hypothesis, the most reasonable conclusion is that stress does either have no or only a small effect on IPD-MD investments.

Hence, and regarding our main hypothesis, there was no significant effect of psychosocial stress on ingroup-love, outgroup-hate, or selfish choice despite the relatively high power of the experiment. The Bayesian credibility interval of the main effect of stress on investments into the within- and between-pool is considerably narrow; if the heavier tailed 95%-bound of the stress-related posterior distributions of the β-estimates is considered an upper limit of a standardized measure of difference, it only yields small plausible effects (Cohen, 1992), if any. The posterior distribution of the interaction term is tailed towards higher credibility. Here, plausible effect sizes based on the 95% bound of the longer tail of the posterior still range in the medium category (Cohen, 1992). Further calculations of Bayes factors show reasonable evidence in favor of our null hypotheses. This, as well as the narrow frequentist confidence intervals around the effect sizes, point to no – or a non-detectable - effect of our stress manipulation on IPD-MD decisions.

Of course, this analysis still leaves room for doubt of a true null effect. We could accept the HDI-based null hypothesis using criteria of 20% (for the main effect of stress) or 45% (for the interaction between pool and stress) of the standard deviation around 0 in an equivalence test, but these criteria are still very liberal (Kruschke, 2018). The variability in individual contributions was large, so that we cannot exclude the possibility with certainty that we simply failed to detect small or noisy stress effects on choice. However, our sample size is large, and given our initial power, the centeredness of the posterior distributions, and the Bayes Factor analyses the most likely interpretation of our results is that psychosocial stress had very small or non-existent effects on investment behavior in the IPD-MD game.

Next, we asked if investment decisions in the IPD-MD were moderated by changes in the levels of stress markers CORT, α-amylase and heart-rate, independent of a main effect of stress. To this end, we regressed the stress markers on the contributions to the pools. As estimates, we considered the area under the curve with respect to increase (AUCi; Pruessner et al., 2003) for CORT and α-amylase, and the increase in heart-rate from the average of the two baseline recordings to the gTSST/control procedure. A mixed linear model with pool, the stress markers CORT, α-amylase and HR, as well as their interactions with pool and condition as fixed effects was calculated. Our results revealed that changes in heart-rate, but not CORT or α-amylase, modulated decision behavior. We found a significant interaction effect between pool and heart-rate increase.

Simple regressions on the between- and within-group pool separately indicate an association between heart-rate increase and decreasing ingroup-friendly, as well as increasing outgroup-hostile investments.

These secondary analyses, hence, yield at least partial evidence for the hypothesis that catecholaminergic action drives outgroup-hate. Heart-rate increase from baseline to stress, which is heavily influenced by catecholaminergic action (Hall, 2016), predicted a decrease of investment into the within-group pool, and a slight, statistically significant increase into the between-group pool. However, while sympathetic activity is associated with enhanced NA release, our other marker of NA activity, α-amylase, did not significantly correlate with pool investments in the IPD-MD game. Hence, the question remains why heart-rate, but not α-amylase levels, predicted changes in ingroup-love and outgroup-hate. The heart-rate response is a temporally well-resolved measure of sympathetic activity, including, but not restricted to, noradrenergic release, and is also to a degree influenced by the parasympathetic nervous system (Hall, 2016). α-amylase is mainly secreted by the parotid glands, it is directly controlled by sympathetic input, and linked to plasma noradrenaline (Chatterton et al., 1996; Nater & Rohleder, 2009). Correlations between sympathetic indicators and α-amylase are moderate (Nater & Rohleder, 2009) and often noisy. Also in our study, heart-rate and α-amylase correlate significantly, but weakly ($r = 0.19$, $p = 0.018$). Thus, both measures show complex relationships with sympathetic activation, and each have their own caveats (Nagy et al., 2015; Picard et al., 2009) because they measure different subprocesses of arousal. Thus, the relationship between the sympathetic stress response and a shift away from ingroup-love to outgroup-hate is possibly real, but the specific mechanisms are complex and need to be illuminated in future studies.

None of the sex hormones testosterone, progesterone & estradiol, nor their interactions explained variance in IPD-MD contributions. Inspired by the dual-hormone hypothesis (Mehta & Prasad, 2015; Zilioli et al., 2015) that predicts an interaction effect of CORT and testosterone on behavior, we investigated if the inclusion of any of the sex hormones, particularly testosterone, revealed effects of CORT on IPD-MD decisions. For interpretability, we constructed three different mixed linear models in an exploratory analysis in which we regressed pool (within-/between-group), the respective sex hormone, the area under the curve of CORT, and their interaction terms on the investments. In order to condition on participants' gender and control for different effects of the sex hormones for males and females, we additionally entered gender as a factor. We let the intercept vary for each participant. Only the model including testosterone yielded significant findings: when testosterone was considered in the model, CORT increase, as well as testosterone levels predicted pool investments. CORT levels were associated with an increase of allocations into the within-group pool, and testosterone boosted both within- and between-group pool investments. This suggests that CORT and testosterone levels explained the variance in within- and between-group investments that was not accounted for by each hormone alone.

These findings are somewhat in accordance with our hypothesis that the CORT response would be correlated with ingroup-love (Margittai, van Wingerden, et al., 2018; Steinbeis et al., 2015). Neither stress nor CORT were directly associated with ingroup-love, but when controlling for the variance explained by testosterone, CORT indeed positively predicted within-group pool investments. Also, if testosterone levels were conditioned on CORT, testosterone predicted an increase in overall within- and between-group pool contributions. Note that, because the default option was to keep the investment, increased investments into within- or between-group pools reflect a dominance of other-regarding over selfish motives. Hence, this finding suggests that, when considering the variance explained by either hormone, CORT and testosterone indeed predict ingroup-love and other-regarding behavior.

Thus, we found that heart-rate increase predicted a shift from within- to between-group pool investments, and that CORT and testosterone levels explained the variance in within- and between-group investments if considered in a model that conditions on both hormones. This finding is partly consistent with our hypothesis predicting a double-dissociation of sympathetic activity and CORT on ingroup-love and outgroup-hate. However, it has to be interpreted with caution because of the complexity of the results and the inconsistency in redundant stress marker effects, e.g., the lack of correlation of α-amylase with pool investments.

Also, this result seems consistent with the dual-hormone hypothesis (Mehta & Josephs, 2010; Mehta & Prasad, 2015), which postulates that testosterone needs low CORT to predict aggression, while CORT would boost empathy, but only at high testosterone levels (Zilioli et al., 2015). However, despite the fact that, here, CORT levels predicted IPD-MD decisions only when controlling for testosterone (and vice versa), we found no statistical interaction between CORT and testosterone, nor any three-way interaction with pool; our results therefore cannot be readily interpreted as a CORT-testosterone moderation effect on social choice. Thus, our results are, once more, exploratory, complex and call for further investigation.

Overall, it, appears that male participants showed more other-regarding behavior while females were more selfish. This finding is consistent with much of the literature on gender differences in cooperative and competitive behavior, showing that men are often more competitive than women, but can also form cooperative alliances to reach a common goal and protect resources and social status (e.g. McDonald et al., 2012). The ultimate cause of such behaviors is often discussed in terms of the evolutionary importance of forming strong male bonds to enhance chances of success in intergroup conflicts (Van Vugt et al., 2007). However, evidence that testosterone mediated the gender effects on other-regarding behavior in the present study was weak: none of our sampled sex hormones were directly correlated with the male participants' increased ingroup-love and outgroup-hate (except when testosterone and CORT were considered in one exploratory model) and none of our trait measures explained the gender-dependent variance in IPD-MD choices. Hence, the proximal mechanisms underlying the gender differences in other-regarding behavior in the present study remain elusive and might be caused by other factors, such as non-physiological gender differences. For example, a recent meta-analysis on the effects of same- versus mixed-sex group compositions on social behavior found that female participants are slightly less cooperative than men in same-sex settings (Balliet et al., 2011). Moreover, the interpretation of the small size of the gender effect (partial $eta^2 = 0.024$) requires caution, too.

The most robust result of the present study is the null effect of psychosocial stress on pool investment in the IPD-MD game. The absence of any acute stress effect on social decision making is puzzling, given the vast number of studies that found such effects on social choice (e.g. Margittai et al., 2015; Margittai, van Wingerden, et al., 2018; Von Dawans et al., 2012, 2019). It is unlikely that the null-effect of stress on social choices was due to ineffective or insufficient stress induction. There are several other reasons why our manipulation might have failed to work. For instance, it is plausible that our implementation of the IPD-MD was not sufficiently sensitive to the social constructs it was supposed to measure, despite recent claims to the contrary with other implementations (De Dreu et al., 2010, 2015). Our task might have prompted deliberate, strategic thinking, but relied to a much lesser degree on social sentiments (Hofmann et al., 2009; Loewenstein et al., 2015). Social feelings like ingroup-love and outgroup-hate might have been overshadowed by strategic considerations, or they might have been outright irrelevant for task performance. Additional analyses partially support this possibility, suggesting that the IPD-MD might, indeed, be a predominantly strategic decision game, and less an instrument to capture affectively tinted ingroup-love and outgroup-hate (see supplemental online material).

In addition, it is also possible that our ingroup-outgroup manipulation was not salient enough to produce true intergroup rivalry. Although our political framing induced relatively strong outgroup harm (between-subject pool investments) as compared to other studies (15.1% in our study vs.

6% in the original work by Halevy and colleagues, 2008), it was, overall, still low (a recent replication also reached 18.33% on average, see Thomae et al., 2015). Consistent with Weisel and Böhm (2015), we conclude that political rivalries in Germany may not be strong enough to produce a robust outgroup harm response. The expression of any covert tendencies for outgroup harm might, additionally, be shackled by prevailing social norms prohibiting interpersonal aggression. Future studies could replicate the present experiment with a more emotionally salient intergroup manipulation aimed at overcoming harm aversion, such as the recruitment of members of minorities (Hein et al., 2018) or rivalling ethnic tribes (Bernhard et al., 2006).

Overall, our study shows no direct effect of socio-evaluative stress on social decisions in the IPD-MD. However, heart-rate changes in response to stress were associated with a shift from ingroup-love to outgroup-hate. In addition, when considered in one model, CORT and testosterone action was associated with ingroup-love or both, ingroup-love and outgroup-hate – and thus, other regarding behavior. In general, male participants revealed more other-regarding preferences than female participants, and this gender-effect was neither explained directly by sex hormones nor gender-dependent traits. Our study contributes to the literature on stress and decision making by highlighting the boundaries of stress effects on social choice; we conclude that our version of the IPD-MD failed to capture stress-related effects on social sentiments, probably due to its strategic character. Future studies are needed to further elucidate putative stress effects on social behavior.

4. Transition: on virtues and norms

A central conclusion from the results of our first study is that it can be very difficult to bring spontaneous aggression and harm-seeking behavior into the lab. Participants remained reluctant to invest into the between-group pool, even though they were made to believe that the outgroup is composed of sympathizers of the (in these times) highly contentious and potentially unpopular right-wing party AfD. There are two major difficulties when studying aggressive behavior in a standardized laboratory context. First, there are obvious ethical and legal barriers in eliciting substantial levels of aggression, for instance, by using purposeful provocations against participants (McCarthy et al., 2018). Here, the consequences for the researcher designing and running research paradigms are clear and leave little space for dispute. Second, everyday norms are highly prohibitive of overt harmful behavior. Humans possess a wide-ranging reluctance to harm other people; they act according to the “do-not-harm" principle, which is hypothesized to be particularly powerful when the respective harm affects groups of people (Baron, 1995). Harm aversion is considered a key precept in the ontogeny of human morality: it crosses cultural boundaries and is already measurable at a very early age (Decety & Cowell, 2018; Smetana, 2006). Indeed, harm aversion is so powerful that people choose to rather accept harm to themselves than harm to others (Contreras-Huerta et al., 2020; Crockett et al., 2014).

The situation is quite different when the decision is not about directly harming, but refraining from helping others, even though this might equally harm the other (Baron, 1995). Indeed, there is often an asymmetry between the moral acceptance of decisions that actively cause harm to, and decisions that omit to gratify the other person (Baron, 1995; Baron & Ritov, 2004; Spranca et al., 1991). Within the game theoretic social dilemma literature, this effect is also referred to as the social framing effect (Dufwenberg et al., 2011; see also Cartwright, 2016; Goerg et al., 2019) because it involves a simple change of the decision problem’s frame, as well as its conceptual closeness to the literature on loss aversion (Tversky & Kahneman, 1979). In prospect theory, the negative value of losses declines more steeply than the positive value for gains increases (Tversky & Kahneman, 1979). Consider, for instance, the ultimatum game task by Leliveld and colleagues (2008). As usual, two participants are instructed to bargain about a specific share of an amount. Under a classical give-frame, a proposer offers a share of his endowment, for example 3€ out of 10€ endowment. The receiver then decides to either accept or decline the offer. In case of an acceptance, both players receive the respective amounts, i.e. 7€ for the proposer and 3€ for the recipient. In case of a rejection, both players leave empty-handed. Under a take-frame, *the receiver* receives an initial endowment, and the proposer needs to decide on an amount to deduce from the other player’s endowment. Here, the proposer could choose to take 7€ from the receiver’s 10€. In case of an acceptance, the proposer goes home with 7€ and the receiver with 3€. Despite an equal payoff matrix, Leliveld and colleagues (2008) found that proposers tend to claim significantly less in the take-frame than in the give-frame across a series of three experiments. The take-frame increased the magnitude of the initial proposals, which is interpreted as a norm-dependent reluctance to harm the other player. Such framing effects hold for a wide array of social games, although they are not ubiquitous and sometimes might differ in their direction (Cookson, 2000; Cubitt et al., 2011; De Dreu & McCusker, 1997; Dreber et al., 2013; Gerlach & Jaeger, 2016; Goerg et al., 2019; Sonnemans et al., 1998). Omission-related framing effects occur in outgroup hate, too: Weisel and Böhm (2015) showed that outgroup-discriminatory investment patterns towards outgroups at high enmity were more accepted if the IPD-MD included the choice to just refrain from helping the outgroup. Participants simply omitted any outgroup-oriented contributions and, thus, discriminated against the other group in a more pronounced fashion than they chose to actively harm the outgroup in the classical IPD-MD. All in all, it appeared to be easier and less morally reprehensible to just leave the other group out than to actively decide to inflict harm on them.

Neither tend-and-befriend nor fight-or-flight patterns were observable in the previous study. Strategic and normative considerations, as potentially generated by the complex setting, might have restrained decision makers to perform according to their social sentiments. Participants in our study might have just been too preoccupied pondering the effects of their decision on their outcome, possible behaviors of other participants and the risks associated therewith. These properties might have been less optimal to study potentially subtle effects of laboratory stressors with an intensity that is, of course, ethically and legally restricted. The framing effect, however, appears to be a stable phenomenon in a set of behavioral games, and has the potential to produce reliable and large enough effects to serve as a solid basis for manipulation. Fight-or-flight can, for instance, manifest in less sharing under a give-frame (as found in Vinkers et al., 2013), but also more unrestrained taking under a take-frame. Stress was also found to offset the adherence to moral norms (see Starcke et al., 2011). Furthermore - and contrary to the IPD-MD - the social discounting paradigm (Jones & Rachlin, 2006; Strombach et al., 2015) has been repeatedly shown to be sensitive to stress-related modulations (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018). The question why this is the case requires more thorough discussion, as well as further empirical studies. However, to further clarify when and where fight-or-flight and tend-and-befriend patterns occur, this paradigm might prove optimal.

In the next section, I will introduce a series of studies that test a new experimental paradigm introducing the framing effect to social distance dependent social decision making. Then, my coauthors and I will illuminate the neural mechanisms that accompany crucial periods of framed decisions as such in an additional neuroimaging experiment. The attentive reader has, most likely, noticed that I have put a minor focus on neuroscientific introductions into social decision making. This was because my primary activities during my doctoral studies were mostly dedicated to behavioral stress research, which then resulted in the publications in sections 3 and 6. For the studies presented in section 5, I was involved in supporting my colleagues in the design and the acquisition of the behavioral studies, as well as the acquisition of the fMRI-data. However, I was not involved in the conceptualization and implementation of the behavioral and fMRI-analysis, which is why this study deserves overall less focus in my thesis.

In my final research study, I will show how stress modulates the framing effect toward a less norm-oriented decision pattern. The social discounting paradigm (Jones & Rachlin, 2006) describes generosity toward others as a function of the perceived social distance. For instance, someone's willingness to share with their partners or best friends will most likely be higher than with their co-workers. The decision maker might, in turn, be more ready to share with co-workers than with strangers on the street. Jones and Rachlin (2006) discovered that the relationship between generosity and social distance is best described by a hyperbolic function with two parameters. The first parameter – usually expressed by a capital *V* – is the intercept of the hyperbola, while the second parameter – usually a *k* – specifies the steepness of the curve. Such hyperbolic modelling approach can differentiate generosity on two dimensions: the willingness to share money with socially close and socially remote others. One person might, for instance, reveal a (in wider terms) parochial preference for sharing a large amount with family and close friends, but almost nothing with acquaintances or strangers, which is expressed in a high intercept and a very steep decay of the hyperbola. Another person might be generally generous, and not differentiate so much between socially close and remote others: Here, the hyperbola best describing such decision pattern would exhibit a shallower curvature. Studies by Margittai and colleagues (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018) showed that social discounting is, indeed, very sensitive to stress effects. Due to its hyperbolic parametrization, social discounting further carries the advantage that it allows for simultaneous occurrence of fight-or-flight and tend-and-befriend toward close and remote social distances, respectively.

The framing effect is induced by varying who receives the endowment, the player or the receiver (the other person). The player decides to either spend money from his or her endowment for the sake of the other person’s good (give-frame), or to deduce money from the other person’s endowment for her or his own good (take-frame). Importantly, the endowments are equal, and hence, the payoff matrix is equivalent. Across several types of studies in sections 5 and 6, the effect remains the same: The take-frame induces more generous behavior, most likely due to participants’ reluctance to harm the recipient by reducing their ownership for their own sake.

The procedure in the study described in section 6 differs from previous studies by Margittai and colleagues (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018): we used a novel and economic hybrid stressor called the Maastricht Acute Stress Task (MAST, Smeets et al., 2012) that brings together the advantages of different psychoevaluative and physiological stressors. Here, my coauthors and I were able to find a fight-or-flight-type response that manifests in a less pronounced give- and take-frame asymmetry at larger social distances. In other words, stressed participants appear to care less about the fact that they harm distant others and are ready to deduce more money from socially remote recipients for themselves. The results are interpreted within the context of a fight-or-flight-related erosion of moral norms after stress.

5. Summary of study 2

Arbitration between insula and temporoparietal junction subserves framing-induced boosts in generosity during social discounting.

Manuela Sellitto[1], Susanne Neufang[1,2], Adam Schweda[1], Bernd Weber[3] & Tobias Kalenscher[1]

1: Comparative Psychology, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany.

2: Department of Psychiatry and Psychotherapy, Medical Faculty, Henrich-Heine-University Düsseldorf, Düsseldorf, Germany.

3: Institute of Experimental Epileptology and Cognition Research, University Clinic Bonn, Bonn, Germany.

Published in: Neuroimage (2021)

(Please note that this section presents a summary of my own research, so that various parts of the text are equal to the original publication. The original article can be found in the appendix.)

5.1. Introduction

Collaboration offers benefits to their members that they would not be able to achieve individually. However, societies can only function efficiently if their members are willing to contribute to causes whose beneficiaries are socially remote, as it is often the case with social welfare, public health insurance, or state pension systems (see also Kalenscher, 2014). Most people are indeed willing to sacrifice own resources for others (Nowak, 2006; Rilling & Sanfey, 2011), but their generosity typically declines steeply with social distance between them and the recipients of help, a phenomenon called social discounting (Jones & Rachlin, 2006; Strombach et al., 2015).

The social discount function is idiosyncratic (Archambault et al., 2020; Kalenscher, 2017; Vekaria et al., 2017), but it is far from stable within and across individuals. For instance, we and others have shown that participants from individualistic or collectivistic cultures (Strombach et al., 2014) differ in their welfare towards socially close peers, and that stress (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018) can increase generosity towards socially close friends and acquaintances. We further showed that disrupting the temporoparietal junction (TPJ) by means of transcranial magnetic stimulation increases the steepness of social discounting (Soutschek et al., 2016). However, despite this evidence, as well as its paramount theoretical and societal significance, means to *increase* the inclination for costly support of socially remote beneficiaries are elusive.

Here, we provide behavioral and neural evidence for a simple manipulation that aims at significantly increasing individuals' willingness to support socially remote others. We make use of the observation that people are more sensitive to others' losses than gains (Bardsley, 2008; De Dreu & McCusker, 1997; Evans & van Beest, 2017; Everett et al., 2015; R. Li et al., 2017; Liu et al., 2020; Schweda et al., 2020; Sip et al., 2015; Smith et al., 2015; Wang et al., 2017; Xiao et al., 2016; Zheng et al., 2010), and are consequently strongly reluctant to increase their own payoff at the expense of others' welfare (Baumeister et al., 1994; Chang et al., 2011; Chang & Sanfey, 2013; Crockett et al., 2014; List, 2007). We hypothesized that participants would be more altruistic towards others, including socially remote strangers, if a costly generous choice was framed as preventing a monetary loss to others rather than granting them a gain, even if actual economic outcomes were equivalent.

To test this hypothesis, we elicited social preferences in a binary standard version of the social discounting task (gain frame; Strombach et al., 2015) as well as in a loss frame variant. In each trial, participants decided to share money with other individuals on variable social distance levels. They chose between a selfish option, yielding high own-payoff and zero other-payoff, and a generous option, yielding a lower own-payoff and a non-zero other-payoff. The main difference between conditions was the way the decision problem was described: in the gain frame, a generous choice would yield an equivalent gain to the other player, while, in the loss frame, it would imply preventing the loss of a previous endowment to the other player.

To obtain further insights into the psychological and neural mechanisms underlying this framing effect on social discounting, we measured blood oxygen level-dependent (BOLD) responses while participants performed both frame conditions of the social discounting task. We hypothesized that the psychological motives underlying generosity were frame-dependent and dissociable on the neural level. Consistent with our previous work (Strombach et al., 2015), we predicted that generosity in the gain frame was vicariously rewarding and the result of the resolution of the conflict between selfish and altruistic motives. Specifically, generosity in Strombach and colleagues (2015) was associated with activity in TPJ, which suggested facilitation in overcoming the egoism bias via the modulation of value signals in the ventromedial prefrontal cortex (VMPFC), a brain structure known to represent own and vicarious reward value (Bartra et al., 2013; Mobbs et al., 2009). In line with Soutschek and colleagues (2016), and Strombach and colleagues (2015) we therefore expected that generous choices in the gain frame would elicit activation of the VMPFC along with TPJ. Conversely, in the loss frame, we expected that the disinclination to maximize own-gain at the expense of other-loss was motivated by the desire to comply to social norms, such as the do-no-harm principle. We therefore hypothesized increased activity in brain regions that are implicated in the social sentiments that motivate individuals to comply to social norms (Montague & Lohrenz, 2007; Xiang et al., 2013). Such social sentiments have been consistently associated with the insular cortex (Chang et al., 2011; Chang & Sanfey, 2013; Civai et al., 2012; Corradi-Dell'Acqua et al., 2013; Hernandez Lallement et al., 2014; Oldham et al., 2018; Samanez-Larkin et al., 2008; Spitzer et al., 2007; Tomasino et al., 2013; Von Siebenthal et al., 2017; Wang et al., 2017; Yu et al., 2014).

5.2. Methods

First, three separate behavioral studies were carried out to test the validity of our paradigm in different settings and with different compensation procedures. Study 1 was run online (*n* = 61; seven participants later excluded from the analyses due to bad fitting) and participants were paid a fixed allowance of €8.5. Study 2 was run online (*n* = 36) and participants, all psychology students on campus, were reimbursed for their time with a fixed amount of university credits. Study 3 (*n* = 39; eight participants later excluded from the analyses) was run in the laboratory and participants were paid out with the same fully incentive-compatible procedure as in the fMRI study 4.

In accordance with an appropriate power analysis (see manuscript), forty healthy young volunteers were recruited at the Life&Brain Research Center in Bonn for an fMRI study. All participants met MR-compatible inclusion criteria. Due to excessive head motion during measurements (>4 mm, >4° rotation, as computed through Artrepair Toolbox; Stanford Psychiatric Neuroimaging Laboratory, see Cho et al., 2013; Strombach et al., 2015; Wendelken et al., 2011), 10 participants were excluded from all analyses. Thus, the final sample included 30 subjects. As reimbursement, they were paid €20 as participation fee, plus earnings from one randomly picked trial in the social discounting task. Therefore, participants' payoff ranged from €27.5 up to €35.5.

In the modified social discounting task (adapted from Strombach et al., 2015), participants were first asked to imagine people from their social environment represented on a scale ranging from 1 (the person socially closest to them) to 100 (a random stranger). They were instructed to select six real persons located at social distances of 1, 5, 10, 20, 50, and 100. Each trial began with the display of the social distance level of the partner the participant was playing with. Social distance was represented with a ruler scale consisting of 101 icons. The left-most icon, highlighted in purple, depicted the participant. One of the remaining 100 other icons was highlighted in yellow, indicating the social distance of the partner. Furthermore, social distance information was additionally indicated as a number on top of the highlighted yellow icon.

The *gain frame* manipulation was near-identical to the task used in Strombach and colleagues (2015). Briefly, after presenting the social distance information as described above, participants were instructed that, in this trial, the experimenter gave an initial endowment of €0 to the other person. Then, two monetary options appeared, a selfish and a generous option. The selfish option indicated the reward magnitude for the participant, if chosen (e.g., €115 to the participant and no reward to other). The generous option contained a smaller own-reward to the participant (€75) and an other-reward to the other person (€75). Participants indicated their choice of the selfish or the generous alternative by a left or right button press.

In the *loss frame*, participants were informed, after the social distance presentation, that the other person has received an initial endowment of €75. On the next screen, a selfish and a generous alternative appeared. When choosing the selfish alternative, the participant received the own-reward amount indicated in purple (e.g. €115), and the other person lost her initial endowment, as indicated in yellow (-€75), thus leaving her empty-handed. When choosing the generous alternative, the participant received a smaller own-reward indicated in purple (€75), implying that the other person would keep her endowment.

In addition to the framing (gain frame, loss frame) and the social distance levels of the other (1, 5, 10, 20, 50, 100), in each condition, we manipulated the magnitude of the own-reward across trials: we used nine selfish reward amounts per frame condition, ranging from €75 to €155 in steps of €10. The generous alternative's payoff always yielded €75 own-reward and €75 other-reward in all conditions and trials. Both frames were mathematically equivalent. The only difference between conditions was that a €0 other-reward outcome was framed as a loss of the initial endowment in the *loss frame* vs. a null-gain in the *gain frame,* and a €75 other-reward was framed as keep-endowment in the *loss frame* vs. a €75 gain in the *gain frame*. Study 3 and 4 were incentive-compatible (see main manuscript)

The general procedure for studies 1 to 3 was implemented as follows: in studies 1 and 2, participants were instructed about the social discounting task, and then, after answering comprehension questions, they assigned other persons from their social environment to the social distances 1, 5, 10, and 20, completed the task, and finally filled out a questionnaire on social desirability. Monetary payment for study 1 was implemented via Clickworker (GmbH), whereas university credits reimbursement was carried out on campus for study 2. In study 3, after recruitment, participants were invited to the laboratory, they were instructed on the social discounting task along with the comprehension questions. They then completed the task, and filled out a questionnaire on a laptop. Finally, they were reimbursed for participation contingent on their choices

In study 4, participants first received instructions about the social discounting task and then, after applying comprehension questions to check for full understanding of the task, they assigned other persons from their social environment to the social distances 1, 5, 10, and 20 via paper and pencil. Afterwards, participants performed a few sample trials to familiarize with the task structure and they were subsequently cleared for the scanning session. At the end of the scanning session, they answered control questions concerning the social discounting task, and filled out a demographic questionnaire as well as questionnaires assessing social desirability and empathy. Finally, participants were debriefed and received their monetary allowance.

Similar to previous studies, we approximated the participants' decay in generosity across social distance with a hyperbolic function:

$$v = \frac{V}{(1 + k * SD)}$$

where v represents the discounted value of generosity, SD represents social distance, *k* represents the degree of discounting, and V is the intercept at social distance 0 (see e.g. Strombach et al., 2015). While V can be considered an indicator of generosity towards socially close others (e.g. Margittai et al., 2015; Margittai, van Wingerden, et al., 2018), *k* describes the discount rate, i.e., the steepness by which the social discount function decays across social distance. We estimated *k* and V for each participant separately.

To estimate V, we titrated the selfish amount to determine, at each social distance, the point at which the subject was indifferent between the selfish and generous options using logistic regression (see Strombach et al., 2015). To fit the hyperbolic function and estimate *k*, we modeled trial-by-trial choices via a softmax function to compute the probability P of choosing the selected option o_i over the other option o_{ii} on a given trial:

$$Poi = \frac{1}{1 + \exp\left(-1 * m * (voi - voii)\right)}$$

given the subjective values v (based on the current selfish amount and social distance) of the current available options o_1 (vo_1) and o_2 (vo_2) as in the hyperbolic function. The nuisance parameter *m* reflects the stochasticity of individual performance. The larger the *m*, the less noisy the choice pattern. Individual discount rates were defined by the respective *k* value that yielded the best prediction of the observed choice probabilities by applying maximum-likelihood estimation using nonlinear optimization procedures. To this end, we minimized the log-likelihood of the choice probabilities to obtain the best-fitting *k* and *m* parameter estimates, by summing across trials, given a specific set of model parameters *k* and *m*, the logarithm of $P(o_i)$. We additionally performed *parameter recovery simulation* to check that the fitting procedure had generated meaningful parameter values (see Wilson & Collins, 2019). This procedure revealed robust recovery of all parameters, which indicates a good fit of the data. Both estimated variables V and *k* were analyzed using non-parametric statistics as they were, in most of the cases across the four studies, not normally distributed. When participants did not discount at all (i.e., they always chose the generous or the selfish option), *k* was set to 0 and V was set to 80 (i.e., maximum reward amount foregone = maximum selfish amount 155 – generous amount 75) for all generous choices or to 0 for all selfish choices.

The fMRI images were acquired in a 3T scanner and properly preprocessed (see main manuscript). At the first-level analysis, trial-related activity for each participant was modeled by delta functions convolved with a canonical hemodynamic response function to model the effects of interest, as well as six covariates capturing residual motion-related artifacts, and a temporal derivative for each regressor of interest to account for slice timing differences.

For each participant, relevant contrasts were computed for each general linear model (GLM) and entered into second-level random effect analysis. The following variables were considered in the analyses: the loss frame condition; the gain frame condition; generous choices; selfish choices.

Comparisons were run via one-way ANOVAs, within subject, and via one-sample t-tests, where appropriate.

GLM1 searched for differences in BOLD activations between frame conditions during generous choices, where the onset of a generous choice was defined as the participant's button press to choose the generous option after the monetary options had appeared on the screen. It included an unmodulated regressor of all generous choices made in the loss frame condition and an unmodulated regressor of all generous choices made in the gain frame condition. Additionally, the selfish amount magnitude was included as trial-by-trial parametric modulator of all main regressors, separately. We additionally considered the reward foregone as a parametric trial-by-trial regressor. Reaction times (RTs) were used as duration to account for differences between gain and loss frames. Additionally, missed trials were included as regressors of no-interest and modeled with duration = 5 s, i.e., the maximum time allowed to respond.

GLM2 tested for the effect of frame condition, and therefore included an unmodulated regressor of the onsets of the loss frame condition and an unmodulated regressor of the onsets of the gain frame condition. The frame onset was defined as the trial start. The social distance was included as trial-by-trial parametric modulator of the frame onsets, separately for the gain and the loss frames. A stick function was used as duration.

All whole-brain level results, as well as ROI-based results, were initially thresholded at $p < 0.001$ (uncorrected), minimum cluster size = 5 voxels, and then corrected at the cluster level for multiple comparisons ($p < 0.05$, family-wise error rate).

We additionally conducted, where relevant ROI analyses for VMPFC, TPJ, and insular cortex by using anatomical bilateral masks using conventional maps and imaging software (see main manuscript).

The dynamic causal modelling (DCM) analysis focused on the interplay between insula and VMPFC and between TPJ and VMPFC, addressing both (i) regions endogenous connectivity and (ii) condition specific modulation of the regions (driving inputs) and their connections (modulatory inputs). We therefore constructed a hierarchical model with regressors defining both frame conditions activations against the total baseline activation. Thus, we entered in the DCM: a regressor of no-interest for baseline connectivity (‘all trials’, used to correct for global activation) including onsets of the screen presenting the framing information and the social distance, and the onsets of the screen presenting the monetary options, at all trials; a regressor (‘all loss trials’) including onsets of the screen presenting the framing information and the social distance, and the onsets of the screen presenting the monetary options, for the loss frame trials; a regressor ('all gain trials') including onsets of the screen presenting the framing information and the social distance, and the onsets of the screen presenting the monetary options, for the gain frame trials.

Subject-specific coordinates were guided by ROI-based group activation maxima in the three network regions from the univariate group-level results. Volumes of interest (VOI) spheres, with a radius of 6 mm, were built around the posterior insula, rTPJ, and VMPFC. Regional time series were extracted as the first eigenvariate of the three network regions for 'all trials' and mean-corrected for the effect of interest F-contrast at a liberal threshold of $p = 0.1$. This threshold was lowered for some participants until all regions could be detected (Zeidman, Jafarian, Corbin, et al., 2019; Zeidman, Jafarian, Seghier, et al., 2019).

Based on our univariate results, we constructed bilinear models where the endogenous connectivity across the three regions was always assumed. We specified models with nodes reciprocally connected, where the gain and loss frame were allowed to modulate all connections (Li et al., 2015). The resulting 15 models were grouped in two families: A and B. In family A, both condition-specific driving inputs and condition-specific modulatory inputs were assumed. In family B, only condition-specific driving inputs were assumed. *Family A* included eleven models, and *family B* included four models (see the published manuscript for details). All the hypothesized models were entered into Bayesian Model Selection (BMS), to determine the best-fit family and model. The inference method used to compare the models across subjects and session was random effects. Bayesian Model Averaging (BMA) was used for model comparison. Once the optimal model was selected, the participant-specific parameters for the two frame conditions were averaged across the three runs and entered into group analysis with one-sample and paired-sample t-tests, where appropriate. This allowed us to summarize the consistent findings from the subject-specific DCMs using classical statistics (Cho et al., 2013; Z. Li et al., 2015; Neufang et al., 2016; Wiehler et al., 2017; Zhang et al., 2018).

A further mediation analysis aimed at further corroborating the idea that the gain and the loss frame had an effect on generous behavior through the mediating influence of condition-specific neural activations. The frame condition was included as binary independent variable X (dummy variable: 1 = gain; 2 = loss), the proportion of generous choices (gain frame and loss frame) was entered as dependent variable Y, and the neural activations were entered as mediators. Specifically, beta estimates for the posterior insula, the anterior insula, TPJ, and two VMPFC regions were extracted, at the single-subject level, for both frames and included in the model. Neural activations across both frames were treated as parallel mediators. Partially standardized and bootstrapped values are reported, and 95% bias-corrected CIs are adopted. The total achieved power for the mediation analysis was ~0.60.

5.3. Results and discussion

First, in a set of behavioral experiments, we established that our framing manipulation affected generosity towards socially distant others. In all three behavioral studies, we could find that participants had flatter social discounting in the loss than the gain frame that went along with significantly lower k-parameters. No effect was found for the V-parameters, which indicates that the framing effect occurs on large social distances, but not on the close social distances. These results appeared similarly in a model-free approach using AUC-analyses, held when excluding participants with null discounting, and were not explained by social desirability. Hence, while generosity to socially close others was comparable between frame conditions, it decayed significantly less steeply across social distance in the loss than in the gain frame, indicating that participants were considerably more generous towards socially distant others in the loss frame. These results provide evidence for a simple nudge that aims at increasing individuals' willingness to provide costly support to socially remote others. Crucially, between frames, the choice alternatives differed only in the description of the decision problem, but not with regard to their actual economic consequences. Notably, our incentivization procedure made it logically impossible for the other persons to know about their endowment, or the potential loss of it, and participants were explicitly instructed about this; all that mattered was the final positive payoff to self and others. Yet, the fact that our participants were still reluctant to inflict losses to others suggests that they had internalized the social norm of not taking away money from others to such a degree that it prevailed even in the absence of any real economic consequences for others.

To obtain more substantial insights into the psychological and neural mechanisms underlying this framing effect on social discounting, in study 4 we measured BOLD responses while participants performed both frame variants of the social discounting task. We first replicated, once more, the behavioral framing effect on social discounting with pronouncedly flatter discounting curves in the loss than the gain frame.

Our first hypothesis predicted activity in brain structures known to represent vicarious reward value and prosocial behavior in the gain frame (similar to Soutschek et al., 2016; Strombach et al., 2015). Our results from GLM1 indeed revealed clusters located in VMPFC as well as right TPJ to be selectively activated, in addition to other prefrontal regions, when participants made generous choices in the gain frame relative to generous choices in the loss frame. ROI analyses confirmed significant clusters of activation in both VMPFC. Thus, consistent with Hutcherson and colleagues (2015) and Strombach and colleagues (2015) a network comprising VMPFC and rTPJ seems to underlie the motivation for costly generosity in the gain frame. Additionally, the selfish amount magnitude, included as trial-by-trial regressor, did not parametrically modulate activity in VMPFC and rTPJ.

Our second hypothesis predicted that generosity in the loss frame was motivated by social norm compliance rather than other-regarding considerations; generosity should, consequently, go along with a different neural activation pattern in the loss than the gain frame. In a first step, we attempted to isolate frame-dependent neural correlates, independent of participants' choices. To this end, we searched for differential neural activity at trial onset, i.e., when participants learned about the social distance level of the other person and which frame was relevant in the current trial, by contrasting neural activity between the two frames in GLM2. We found significant activation in the right posterior insula in the loss vs. gain frame contrast, which was accompanied by significant activations in frontal regions, including VMPFC, as well as temporal regions. ROI analyses confirmed significant clusters of activation in the right insula as well as in VMPFC. Social distance information, included as trial-by-trial regressor, did not parametrically modulate neural activity in any of these contrasts, suggesting that the activations in insula and VMPFC reflected frame but not social distance information. In support of this conclusion, we found that the right anterior insula, was selectively activated during generous choices in the loss frame relative to generous choices in the gain frame The location within the insula mask was slightly anterior to the peak activation we found at trial onset. Hence, the analysis so far suggests that insula activation reflects the psychological motives underlying generous choice in the loss frame.

We previously provided empirical support for a network model according to which, in a task similar to our gain frame condition, TPJ would facilitate generous decision-making by modulating basic reward signals in the VMPFC, incorporating other-regarding preferences into an otherwise exclusive own-reward value representation, thus computing the vicarious value of a reward to others (Strombach et al., 2015) Here, we expand on this idea and propose that, in addition to the TPJ-VMPFC connectivity in the gain frame, frame-related information in the loss frame would activate insula, which in turn would down-regulate own-value representations in VMPFC, thus promoting generous choices by decreasing the attractiveness of own-rewards. Hence, in brief, we predicted a complex, frame-dependent pattern of connectivity between insula, TPJ, and VMPFC that reflects the different motives underlying generosity in the gain and the loss frame.

To identify the relations between those regions, we estimated their effective connectivity via DCM analysis (Friston et al., 2003). More specifically, we tested the idea that the frame information at the beginning of each trial would drive increased insula activation selectively in the loss frame, and increased TPJ activation selectively in the gain frame. Additionally, we expected increased endogenous connectivity as well as condition-specific modulation between each respective region with VMPFC. Note that we focused our DCM analysis on the posterior insula cluster only, as we were interested in a baseline frame activation; including the anterior insula cluster, specific for generous choice within the loss frame, might have biased the results in favor of our hypotheses. Among the two model families tested, model comparison favored family A, i.e., the family of models that assumed condition-specific effects at the level of both driving input and modulatory input. The winning model was model number *5*, which assumed that the gain frame had an effect on the TPJ and its connectivity with the VMPFC, while the loss frame had an effect on the VMPFC and its connectivity with the insula. Concerning the driving inputs, we compared the average activity in TPJ in the gain frame against 0, and the average activity in VMPFC in the loss frame against 0, but none of the driving inputs was significantly different from 0.

Next, when addressing the modulatory inputs, the only significant difference was found in the loss frame for modulatory activity from the insula to VMPFC against the endogenous connectivity from the insula to VMPFC, reflecting a significant modulation of endogenous connectivity by the loss frame information. In addition, the modulatory input was negative, hinting towards an inhibitory influence of insula on VMPFC in the loss frame.

To provide further support to our idea that the frame effect on social discounting was brought about by condition-specific neural activity patterns, we ran a mediation analysis on the relation between frame information, generous behaviour, and neural activation. Here, frame condition significantly correlated with all neural activations with the exception of one of the two VMPFC clusters included in the analysis. Additionally, while the direct effect of the frame condition on the proportion of generous choices was not significant, the indirect effect of anterior insula on it was significant, indicating that it influenced frame-specific generosity.

Taken together, we indeed found that the anterior insula was significantly more activated when participants made generous choices in the loss frame, relative to the gain frame. Extending these findings, we found that also the posterior part of the insula seemed to be involved in these processes, specifically supporting the representation of the loss frame information even before the decision was made (see also Droutman et al., 2015). Building upon this evidence, we further explored how both activation clusters mediated frame-specific social discounting behavior. We propose and provide empirical support for a network model that predicts that the frame effect on social discounting was associated with a frame-dependent neural connectivity pattern between insula and VMPFC in the loss frame, and TPJ and VMPFC in the gain frame. More specifically, DCM confirmed that posterior insula activation at loss frame onset exerted a negative modulatory effect on VMPFC. It is tempting to speculate that a frame-dependent downregulation of own-reward values in the valuation network during social discounting might lie at the core of the enhanced generosity observed in the loss frame. By contrast, the same analyses confirmed TPJ-VMPFC coupling in the gain frame, consistent with our previous finding (Strombach et al., 2015) that altruism in the gain frame is promoted by increasing the attractiveness of the generous option through TPJ-related upregulation of vicarious reward value signals in the valuation network. Overall, these results call for the idea that the motives behind generosity are likely qualitatively different in the gain and the loss frame, and dissociable on the neural level.

These findings expand on previous evidence that preventing harm to others is a great motivator of prosocial performance (Everett et al., 2015; Wang et al., 2017; Xiao et al., 2016; Zheng et al., 2010; Zhou et al., 2017). However, while others have found that harm prevention was particularly pronounced in a public context (Everett et al., 2015) and dependent on social feedback (Sip et al., 2015; Smith et al., 2015), we show that similar cognitive mechanisms can strongly boost generosity even in a private context and in the absence of social feedback, thus independent of reputational concerns, judgment by social peers, or third-party punishment threats. This suggests that other-harm prevention might be an internalized motive that works unconditionally and universally across contexts, regardless of social consequences. In addition, previous experiments on harm prevention did not manipulate, or provide information on, social distance between donor and recipient (Bardsley, 2008; Crockett et al., 2014; Everett et al., 2015; Z. Li et al., 2015; Liu et al., 2020; Xiao et al., 2016). Hence, while the effects of the resource allocation mode on social discounting were elusive so far, our findings imply that it matters: harm-prevention motives in the loss frame were less dependent on social distance than other-regarding considerations in the gain frame, thus resulting in flatter social discounting.

In particular, our results are consistent with the idea that certain costly altruistic behaviors are not motivated by genuinely other-regarding considerations, but instead by compliance to internalized social norms. But what impels participants to comply to social norms? Here, we propose, along with previous evidence (Chang et al., 2011; Spitzer et al., 2007), that compliance to social norms might be linked to anticipated feelings of guilt, shame, and remorse, and accompanied by insula activation (see also Belfi et al., 2015; Sellitto et al., 2016), which ultimately sustain prosocial behavior. According to this view, insular activation would reflect the negative sentiment associated with social norm transgressions as they occur when being responsible for someone else's loss (i.e. vicarious loss experience). Our data show that this social sentiment and accompanying neural signature can be elicited even when the others' outcomes are merely described as losses, thus, in the absence of real losses to others.

The acceptance and support of the principle of a caring society, and the attitude towards the welfare of socially remote strangers, is central for a civilization to function well. It seems vital for societies to successfully meet current challenges, such as integrating refugees, addressing economic inequality, acceding the trials and promises of a globalized world (Kalenscher, 2014), or managing the public health implications of the current COVID-19 pandemic. Here, we present a simple behavioral framing manipulation that boosts generosity towards socially remote others: framing a selfish choice as a loss to others can motivate prosocial behavior, even if the framing of the choice options is irrelevant for the actual payoff to others. Our neuroimaging data identify the insula as the core component in a network associated with this enhanced generosity in the loss frame. Our results imply that prosocial attitudes towards others are highly malleable and strongly depend on the architecture of the decision problem. The insights gained in this study might, thus, help in designing policies aimed at increasing the acceptance and support of the principle of a caring society, and to change the attitude towards the welfare of socially remote strangers

6. Summary of study 3

Acute stress counteracts framing-induced generosity boosts in social discounting in young healthy men

Adam Schweda[1], Zsofia Margittai[1] & Tobias Kalenscher[1]

1: Comparative Psychology, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany.

Published in: Psychoneuroendocrinology (2020)

(Please note that this section presents a summary of my own research, so that various parts of the text are equal to the original publication. The original article can be found in the appendix.)

6.1. Introduction

Most people are willing to forgo own gains for the benefit of others. Usually, the willingness to do so declines with the social distance to the recipient, a phenomenon called social discounting (Jones & Rachlin, 2006; Strombach et al., 2015). Social discounting has been shown to be robust, but also malleable. Recently, we could demonstrate that participants were much more altruistic towards others, especially strangers, when the decision problem was framed as to obtain a personal financial benefit at the other's expense (take frame) versus to financially benefit the other at an own personal expense (give frame, Sellitto et al., 2019, please note that this citation refers to a preprint of the study presented in the last section). Such framing-induced boosts in generosity towards others are likely to reflect people's reluctance to increase their own payoff at the expense of others due to the compliance to social norms (Bardsley, 2008; Baron, 1995; Baumeister et al., 1994; Chang et al., 2011; Chang & Sanfey, 2013; Crockett et al., 2014; Decety & Cowell, 2018; Sellitto et al., 2021).

Generosity, prosocial behavior and social discounting are strongly influenced by acute stress (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018). Recent evidence suggests that individuals may react to stress with a "tend-and-befriend"-response (Berger et al., 2016; Margittai et al., 2015; Margittai, van Wingerden, et al., 2018; Sollberger et al., 2016; Taylor, 2006; Taylor et al., 2000; Von Dawans et al., 2012, 2019) – an increase in costly generosity towards others, and particularly socially close others (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018) - to mobilize social support in stressful times. Tend-and-befriend implies that stressed individuals would be even less inclined to cause harm to others. Consequently, stress should amplify the frame effect on social discounting (the frame amplification hypothesis).

However, often, the social response to stress is not tend-and-befriend, but fight-or-flight (e.g. Cannon, 1932; McCarty, 2016). Fight-or-flight responses involve antagonistic social behaviors aimed at promoting own survival and well-being, potentially at the opponent's expense. In humans, fight-or-flight-like responses might manifest in more readiness to punish others, and less willingness to trust, to share, and to reciprocate (Starcke et al., 2011; Steinbeis et al., 2015; Vinkers et al., 2013). This line of evidence suggests that stress can lead to less other-regarding thinking combined with an erosion of moral and social norms (Starcke et al., 2011; but see Nickels et al., 2017; Singer et al., 2017). Because the frame effect on social discounting likely depends on other-regarding considerations and social norm compliance (Sellitto et al., 2019), stressed individuals reacting in a fight-or-flight-like manner should show less frame-dependent differences in generosity toward others. Thus, rather than amplifying it, acute stress would be expected to dampen the frame effect (the frame attenuation hypothesis).

Taken together, the frame-amplification hypothesis, inspired by the tend-and-befriend theory, predicts a stress-related amplification of the frame effect on social discounting, while the frame-attenuation hypothesis, inspired by the fight-or-flight model, states that stressed individuals will show a diminished frame effect on social discounting.

In order to decide between these hypotheses, we tested male participants that either underwent a stress procedure (Smeets et al., 2012) or a non-stress control condition, and asked them to complete a social discounting task with give and take frame decisions (Sellitto et al., 2019). Here, the give frame consisted of a variant of the dictator game (Bolton et al., 1998) where participants were endowed with an amount of money and freely decided how much of this endowment they would share with the other person. In the take frame, participants were informed that the other person was endowed with an amount of money, and the participant decided how much money to take away from that person. Importantly, the payoff matrix was equivalent under both frames, and participants were told that, even if the recipients received actual money from the experimenters (i.e. the decisions were incentive compatible), they would not be informed about whether the decision was made under a give or a take frame.

6.2. Methods

We recruited 102 participants at site, of which one participant withdrew from participation due to pain during the stress induction. Before the experiment, participants were screened via telephone for a number of participation criteria typical for experiments within the field of stress research on economic games. All participants completed several trait questionnaires online in order to control for confounding factors that might interfere with the outcome measures (see main manuscript for more details).

Stress was induced using the Maastricht Acute Stress Test (MAST), which has been shown to reliably elicit stress responses measurable via increased activity of the hypothalamic pituitary axis and the sympathetic nervous system (Smeets et al., 2012). The MAST is a hybrid stress task that includes elements of social evaluation, physiological stress and uncontrollability. In the stress condition, participants were instructed to alternate between immersing their hand in 0-2◦C cold water and performing a mental arithmetic task while being socially evaluated and videotaped. The experimenters wore lab coats and behaved in a rigid and non-responsive manner. In the control condition, participants immersed their hand into 36◦C warm water, no camera recordings were made, and they were asked to count loudly and repeatedly from 1 to 25 upwards. The experimenters behaved friendly and wore no lab coats. Interval durations were equal in the stress and control condition.

To quantify the intensity of stress-induction in our participants, we collected saliva samples and heart rate measures over the course of the experiment. Saliva samples were analyzed for the physiological stress-markers cortisol (CORT) and salivary α-amylase (sAA), an indirect marker of sympathetic activity (Nater & Rohleder, 2009). We collected two baseline samples before the MAST and three samples after the MAST. To measure heart rate (HR) – a further marker of sympathetic activity - we used commercial wrist-band photoplethysmographs (Polar A370) to make two 3-minute baseline recordings before MAST onset, as well as a continuous recording during the entire duration of the MAST.

Besides the physiological effects, stress-induction procedures come along with an increase in subjective arousal, tension, and feelings of insecurity. To capture such feelings throughout the experiment, we administered the Positive and Negative Affect Scale (PANAS, Watson et al., 1988) twice before and twice after the stress-induction procedure. Participants additionally indicated feelings of stress, self-confidence, insecurity and shame on visual analogue scales (VAS, see e.g. Hellhammer & Schubert, 2012) every time the PANAS was delivered.

In the social discounting task, we asked participants to imagine 100 people of their social environment on a hypothetical social distance scale, where 1 represents a person they feel closest to and 100 represents a random stranger whom they have never met (Jones & Rachlin, 2006; Margittai et al., 2015; Margittai, van Wingerden, et al., 2018; Strombach et al., 2015). Participants were then asked to indicate the names of people who represent social distance levels 1, 2, 3, 5, 10, and 20. For social distances 50 and 100, participants were asked to imagine somewhat familiar, or completely unknown strangers, respectively. To assess how the framing of the decision problem moderated social discounting, we used an adapted variant of the dictator game (Archambault et al., 2020; Margittai et al., 2015; Margittai, van Wingerden, et al., 2018) with a give and a take frame condition (Sellitto et al., 2019). In both conditions, participants freely decided how to allocate a monetary endowment between themselves and another person. In each trial in the give frame condition, participants received an endowment of either 13EUR, 15EUR or 17EUR, and decided how much to share with the recipient. This was repeated for all three endowment levels and all eight social distance levels (1, 2, 3, 5, 10, 20, 50 or 100) in a repeated measures design. In each trial in the take frame condition, participants were informed that the recipient had received an endowment of 13EUR, 15EUR or 17EUR, and they decided how much money to take away from her or him for themselves. As in Sellitto and colleagues (2019) the payoff matrices were equivalent. Participants were explicitly informed that the other person was unaware of their initial endowment, and would, hence, not learn about the potential loss of it. The task was incentive compatible (Margittai et al., 2015; Sellitto et al., 2019) and free of any deception.

Upon arrival at the laboratory, the first of two 3-minute baseline HR-recordings was made, followed by the collection of the first saliva sample and PANAS. Hereafter, participants received task instructions and provided names of individuals in their social environment representing the social distances. Using several comprehension items, we made sure that participants understood the task. Then, we obtained a second HR baseline recording, and took a PANAS mood questionnaire along with another saliva sample. The MAST was performed, followed by the third saliva sample and PANAS. Participants then carried out the social discounting framing task. The fourth saliva sample was collected after task completion. Then, participants were debriefed and paid out.

The social discount functions are often approximated by a hyperbolic model. Yet, a large number of participants decided not to deduce money from the recipient, which is why a hyperbolic approximation of their choices would be subpar. To circumvent this problem in our main analysis, we adopted a different approach, resembling the one used by Archambault and colleagues (2020): we linearized the social discount function by rank-transforming social distance levels, i.e., the social distance levels 1, 2, 3, 5, 10, 20, 50 and 100 were replaced by social distance ranks 1 through 8, allowing for analyses with a mixed linear model. In order to capture the effects of stress and framing on the discount rates, we regressed the factors frame (give vs. take frame), stress (stress vs. control), and the ranked social distance level (1-8) on trial-by-trial amounts shared (i.e., the monetary amount given to the other in the give frame, or the amount left to the other in the loss frame). We allowed intercepts to vary for each endowment level, and for each participant. We furthermore maximized the random effects structure as suggested by Barr and colleagues (2013) and Matuschek and colleagues (2017). Backward model selection was applied to identify possible relevant, non-redundant regression terms, but none of them was eliminated. Besides that, two-group t-tests or Mann-Whitney-U tests were conducted to compare the baseline trait-measures between groups. Also, condition-specific stress marker elevations were analyzed via classical mixed ANOVAs. In case of a violation of normality, we log-transformed the respective stress marker estimate.

6.3. Results and discussion

To exclude the possibility that stressed and control participants differed in trait and baseline characteristics that could confound our results, first, we compared their a-priori trait measures. Applying a Bonferroni-corrected alpha-level of $\alpha=0.005$, we find no differences between stress and control group.

As a manipulation check, we examined group differences on log-transformed values of salivary CORT and sAA concentrations. Out of 505 samples in total, we lost 28 CORT samples and 54 sAA-samples, mostly due to insufficient saliva. A mixed ANOVA with subsequent simple effects analyses revealed that the MAST provoked an increase in salivary log(CORT) in the stress group compared to the control group, and that this increase already took place directly after the onset of the MAST. By contrast, we found no significant increase in log(sAA) levels between the stress and control groups. Note that the lack of an effect of stress on sAA might have resulted from low statistical power due to the high number of lost sAA samples. If the analysis is run in a mixed linear model, which is – at least to some extent - capable of handling missing data, the stress group showed significantly higher sAA concentration at timepoint 5. Thus, although caution is required, there is some indication that stress increased sAA in our task. The log(HR) – a further indicator of sympathetic activity – showed a stress-related increase in heartbeats per minute for the stress group, but not the control group. Also, the ratings in both the negative and the positive scale of the PANAS were increased after stress-induction, with the former being most likely mediated by subjective arousal. Moreover, mixed ANOVAs and simple effect analyses on the visual analogue scales revealed more stress-related feelings of insecurity, stress, shame and less feelings of self-confidence. Hence, despite the somewhat unclear effects of stress on sAA, all other physiological and psychological measures, including heart rate as a further marker of SAM activity, indicate success of our stress induction.

In order to investigate the effects of stress on the framing effect, we constructed a mixed linear model in which we regressed the amount shared on the rank-transformed social distances, the frame and the stress condition (see also Archambault et al., 2020). Furthermore, the model considers individual and item-specific variation by including varying intercepts for each subject and endowment level (13€, 15€, 17€) in the random effect structure. To maximize the random effect structure (Barr et al., 2013; Matuschek et al., 2017), we entered frame as a random coefficient varying per subject, which yielded better goodness-of-fit than an intercept-only model.

In line with other studies on social discounting, the amount shared with others decreased monotonically across social distance in all treatment and frame conditions. In addition, participants overall shared more money in the take than the give frame, suggesting that the frame manipulation worked. Although main effects should be interpreted cautiously in presence of interaction effects, the inspection and analysis of the data clearly indicates the presence of discounting and framing effects. Hence, we replicated our previous finding (Sellitto et al., 2019) that participants exhibited flatter discounting in the take than the give frame, suggesting higher generosity toward strangers in the take frame.

Further assessment of our data suggests that there was no difference in social discounting between stress and control participants in the give frame. By contrast, in the take frame, the social discount function appeared flatter in control than stress participants, suggesting that non-stressed control participants were more generous in the take frame, and particularly toward strangers, than their stressed counterparts. To be more precise, the three-way interaction between stress x frame x social distance on amount shared reached significance, indicating that stress affected social-distance-dependent generosity differently in the take than the give frame. To break down this three-way interaction effect, we compared the regression slopes, capturing the decline in sharing across social distance, between the give and the take frames; this was done separately for the stress and control participants. We found that the difference in regression slopes between the give and the take frames was more pronounced in non-stressed control participants compared to stressed participants. Hence, while control participants showed flatter social discounting in the take than the give condition, indicating the frame manipulation on social discounting worked, this difference in the steepness of social discounting between frame conditions was less evident in stressed participants. An analogous mixed linear regression model that included social distance as a categorical variable points towards the same interpretation.

Further simple slope analyses on the three-way interaction revealed that stress affected the social discount rates primarily in the take frame, and much less so in the give frame. Taken together, this analysis supports the above mentioned observation that, compared to non-stressed participants, stressed participants were selectively less generous towards strangers, but this effect was found only in the take, not in the give frame condition.

Hence, and overall, our stress manipulation revealed support for the frame-attenuation hypothesis: we found that acute stress diminished the frame effect on social discounting and caused stressed participants to be equivalently generous towards others in the take and the give frames. Furthermore, in the take, but not the give frame, stressed individuals were less generous toward strangers than non-stressed controls. Our finding of a stress-related decrease in generosity towards strangers in the take frame blends with other results demonstrating diminished willingness to share resources under stress (Starcke et al., 2011; Steinbeis et al., 2015; Vinkers et al., 2013), and, in a broader sense, they are consistent with the observation of increased egocentric, antagonistic fight-or-flight-like tendencies under stress (Agnew, 2005; Cannon, 1932; Sandi & Haller, 2015; Silver & Teasdale, 2005).

But what causes the stress-related diminution of the frame effect on social discounting? We recently argued that the frame effect on social discounting is the result of people’s internalized hesitation to transgress the social norm of preventing harm to others, and the associated feelings of guilt and shame (Sellitto et al., 2021). Hence, social norms strongly prohibit taking away money from others in the take frame. Because stress is known to erode social norm compliance (Starcke et al., 2011; Steinbeis et al., 2015; Vinkers et al., 2013), social decision making of stressed individuals will likely be less influenced by social norms than that of non-stressed people, and they would consequently be less hesitant to cause financial harm to others. This explanation can account for the fact that the frame effect on sharing behavior was most pronounced when dealing with unknown strangers: social norms guide social behavior especially towards others at larger social distance levels, but they are less relevant for generous behavior towards friends and family where people are often naturally selfless anyway (Rand et al., 2014; Strombach et al., 2015).

Notably, our finding of reduced prosociality after stress in the take frame stands in contrast with results from other experiments that have shown the opposite pattern of increased pro-sociality after stress (tend-and-befriend; Berger et al., 2016; Margittai et al., 2015; Margittai, van Wingerden, et al., 2018; Singer et al., 2017; Sollberger et al., 2016; Tomova et al., 2014; Von Dawans et al., 2012, 2019; Youssef et al., 2018). It is currently unclear when and why stressed individuals show a stronger or reduced prosocial stress-response. We have recently argued that stress does not provoke either fight-or-flight or tend-and-befriend, but it may boost both tendencies at the same time (Schweda et al., 2019), as outlined in the following. Tend-and-befriend is an alleged coping strategy where stressed individuals invest into their social network in order to receive help in return (Taylor, 2006; Taylor et al., 2000). Accordingly, we have shown that tend-and-befriend behavior is predominantly directed towards socially close others, who, unlike strangers, can potentially provide comfort and support in stressful times (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018). By contrast, fight-or-flight is a strategy primarily aimed at escaping the acute stressor. Acute stress can be expected to originate from socially distant outgroup members who are more likely to present a threat than socially close friends and family, especially at times of conflict. Consequently, antagonistic tendencies should be aimed at socially distant strangers, but less so at socially close others. Even though evidence for this theory is still scarce (Schweda et al., 2019), our current finding of a stress-related and frame-dependent reduction in generosity towards strangers, but not socially close others, is consistent with this hypothesis.

Further questions remain. We could not replicate our previous findings that socio-evaluative or psychopharmacologically-induced stress (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018) increased generosity towards socially close others. One possible explanation for the discrepancy in findings is the employment of the MAST. The type of stressor matters as it has been shown to alter social behavior in several experiments. For example, while physical and psychosocial stressors alone impair prosocial behavior, the two combined actually restore prosociality (von Dawans et al., 2018). Though this finding is not compatible with our results, as we found less prosociality under a combined physical and psychosocial stress induction procedure, this example nonetheless illustrates the complexity of the relationship between social behavior, social norm compliance and stress. Another possibility is the task used to elicit social preferences: The frame version of the social discounting task is more complex and procedurally different to the simple social discounting task used before (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018). Thus, task performance might not be perfectly translatable between tasks, and within-task spill-over effects are to be expected. Whatever the reason for the divergence in results, definitive conclusions about underlying mechanisms of our effects cannot be made with certainty at present. Our results pave the way for future research investigating the frame effect and its interaction with stress in shaping prosocial behaviors. The current study involved male participants only. Following research needs to clarify whether framing and stress interactions on social discounting occur in women, too. Gender differences in social frame effects have been found before (Chowdhury et al., 2017; Ellingsen et al., 2013; Strombach et al., 2016), and we know the stress response is susceptible to variations in sex hormone concentrations (Kirschbaum et al., 1999; Kudielka & Kirschbaum, 2005). Furthermore, stress is multidimensional (Joëls & Baram, 2009), and we still lack knowledge of which mediators of the stress-response affect social behavior. Pharmacological intervention studies using, for instance, corticosteroids and adrenergic drugs, as well as conditions with time lags between stressor and task would be the optimal choice for future studies (e.g. Margittai et al., 2015; Margittai, van Wingerden, et al., 2018; Vinkers et al., 2013).

In summary, our study replicates our previous findings that participants prefer more generous resource allocations to strangers when donations are framed as preventing financial harm to others. We demonstrate that stress mitigates this frame effect on social discounting, so that stressed participants are less generous towards strangers than non-stressed controls. This finding can be tentatively explained as a stress-induced diminished compliance to “do-not-harm” (Baron, 1995). These findings contribute to our understanding of how acute stress alters social norm compliance and interpersonal harm avoidance. Thus, our study broadens our understanding of the impact our psychological state has for our everyday moral and social behavior. We believe that this result is not only relevant for cognitive scientists studying the effects of stress on cognition and behavior, but also important for policy makers and corporate decision makers; knowing under which circumstances stress boosts or corrupts prosociality, especially towards strangers, has practical implications for charity advocacy about the way charity calls or appeals for donations could be worded.

7. General discussion

A stressed fellow can be difficult to handle, and we all know how it feels when our temperaments boil up. In some situations, people can get feisty and belligerent, making scenes and throwing blows at uninvolved bystanders. Other situations might make us seek union with others, feel solidarity, and get soothed by attachments. But when do we react in one way or another? This dissertation committed to exploring when stress makes us more hostile and defensive, or generous and affiliative. It attempts to reconciliate fight-or-flight and tend-and-befriend as the two major frameworks that try to explain how people react in stressful situations. As an overarching hypothesis, my coauthors and I assumed that stress can actually elicit both types of reactions at the same time. This hypothesis is based on totally diverging experimental findings on stress effects in social behavior in prior literature, as well as recent empirical, yet mostly theoretical evidence on the co-evolution of ingroup-directed altruism and outgroup-directed hostility. Furthermore, stress was found to be highly adaptive in other psychological domains: it promotes a shift from slow executive cognitive functions to rather fast and salience-based ones (Hermans et al., 2014) - a functional mode that could have helped us to increase our likelihood of survival in the past. There is evidence that stress-related changes can be very specific: for instance, while it increases memory consolidation and serves as a teaching signal, it also appears to inhibit retrieval, and hence, interference by other memories (Quaedflieg & Schwabe, 2018). Therefore, increasing both tendencies – fight-or-flight and tend-and-befriend – at the same time could primarily constitute an adaptive response, which appears by no means overly specific. Based on previous pharmacological evidence (Margittai, van Wingerden, et al., 2018), we moreover hypothesized that, while tend-and-befriend might result from cortisol action, fight-or-flight might concur with activation of the sympathetic nervous system. Both systems are usually active shortly after facing the stressor.

7.1. Recap

In study 1, my co-authors and I directly tested these hypotheses by applying an intergroup scenario in which groups of three participants were to decide how to allocate an initial endowment: stressed and non-stressed players could choose to maximize their own payoff, increase their ingroup's payoffs, or increase their ingroup's payoff while also decreasing the outgroup's payoff. Group-related social sentiments were augmented by letting each player think that all ingroup members were sympathizers of the same political party, and that they were playing against sympathizers of the German right-wing party AfD. A direct comparison of the pool allocations between the stressed and non-stressed participants yielded no significant results. Further analyses even supported the idea of an actual null result or a difference that was too small to be captured by this sample size. On the other side, we found that overall heart rate increase was associated with outgroup hostile investments, and that cortisol was related to increased ingroup friendly contributions as soon as testosterone was considered in the same linear model. These results tentatively suggest that sympathetic and HPA-activity might differentially modulate social behavior towards either inimical or more benevolent decisions. Nevertheless, a direct effect of stress on allocation patterns in intergroup decisions remained unrevealed. We interpreted these results in terms of the comparably strategic nature of the allocation game, as well as strong interpersonal norms that forbid harming others, even in situations of stark rivalry (see also Weisel & Böhm, 2015).

As opposed to the IPD-MD, generosity in the social discounting task has, indeed, proven to be malleable by stress (Margittai et al., 2015; Margittai, van Wingerden, et al., 2018). To further investigate whether fight-or-flight and tend-and-befriend could co-occur, we assessed how stress modulates prosocial sharing in a scenario where the recipients are either socially close or distant, and where the decision to share is either framed as an act of actively giving money to (give frame), or actively deducing money from a recipient (take frame).

The groundwork was laid by presenting four experiments – all summarized in study 2 - that deliver behavioral and neuroscientific evidence for a reproducible and robust framing effect on social discounting. In experiments 1 to 3, my co-authors and I could show that participants' willingness to share money with socially remote people – where, usually, their readiness to do so is considerably reduced compared to socially close individuals – can be substantially increased by simply framing the decision to share as a prevention of the other's loss. In the fourth experiment, we provide evidence that deciding against own-reward maximization at the expense of the other person in the take frame goes along with activation in the insula and the ventromedial prefrontal cortex, which is in line with previous observations that the insula is recruited when we refrain from moral transgressions in anticipation of aversive feelings such as guilt and shame (e.g. Civai et al., 2012; Corradi-Dell'Acqua et al., 2013; Oldham et al., 2018). Accordingly, further analyses on the effective connectivity between condition-related brain activities suggest that the posterior insula negatively modulates the vmPFC in the take frame, possibly downregulating the value signal of the own-reward choice. On the other side, the TPJ and the vmPFC orchestrate generous decisions in the give frame, likely due to integration of other-regarding reward values, as it was suggested before in previous research (Strombach et al., 2015). These findings point to a qualitatively different neural signature of generosity under the give and the take frame. Activity in the insula could also be shown to have a partial effect on the relationship between the frame condition and overall generous investments. Our data, hence, underline the idea that decisions under a take frame are mechanistically different from decisions under a give frame with the former most likely demanding internally norm-driven considerations about future consequences of potential reprehensions.

Finally, in study 3, my co-authors and I explored how stress affects socially discounted generosity under the give and the take frame. Here, we applied a relatively new hybrid stress induction procedure called the Maastricht Acute Stress Task (MAST, Smeets et al., 2012), including both, socio-evaluative and physiological stress components. After completing the stress induction procedure, participants performed a modified Dictator Game version of the social discounting task with give- and take-framed decisions presented in a pseudorandomized fashion. A mixed linear model showed that stress, indeed, attenuated the framing effect: while non-stressed controls showed a substantially increased level of sharing toward socially remote recipients (as it was found in study 2), this effect was less pronounced in the stress condition. More specifically, stress reduced the overall amount shared at large social distances in the take frame. Given that decisions to prevent remote others' losses are assumed to be largely influenced by the trade-off between own-reward maximization and compliance to social norms, our findings go along with the idea that stress can corrupt the adherence to the “do-not-harm"-principle (Baron, 1995).

Concerning our initial hypothesis of a simultaneous stress-related promotion of tend-and-befriend and fight-or-flight tendencies, the results are, indeed, somewhat mixed. In fact, the divergent association between the neurohormonal markers of stress and the allocation patterns in study 1, i.e. that CORT is correlated to ingroup-friendly decisions while HR related to outgroup hostile contributions, points to distinct roles of the HPA and SAM-axes in the modulation of social behavior. Yet, CORT only became a significant predictor when conditioned upon testosterone, and obviously, stress as such has left the investment patterns unaffected. Furthermore, we could find a stress-related mitigation of the frame effect in study 3, which is interpretable in terms of a fight-or-flight response toward people at larger social distances. Considering the previous discovery (and conceptual replication, Margittai et al., 2015; Margittai, van Wingerden, et al., 2018) of tend-and-befriend responses toward socially close recipients in social discounting, a co-occurrence of these tendencies could be possible. Yet, we could not reproduce the same tend-and-befriend responses as found in Margittai and colleagues, and Margittai, Van Wingerden and colleagues

(2015; 2018). Data from all three studies certainly propound complexity, both in terms of the structural characteristics of the game-theoretic paradigm applied, as well as the different facets of stress.

In the next few sections, I will discuss how these results compare to other studies in the area of stress research, and how they can be comprehended in terms of present and future research. In subsection 7.3, I will then expand the implications and limitations already discussed in the manuscripts and constructively reflect on each single paper. I will then discuss a number of propositions on what future research could do in order to further advance our understanding of stress and social behavior. Every researcher knows that research takes place in the field of tension between the value of a specific study or a set of studies for scientific progress, on the one hand, but also practicability and economic considerations, on the other hand. Therefore, I will also discuss future perspectives from an economic and process-oriented viewpoint which could potentially help to take us forward in creating a solid empirical body to build on.

7.2. Integration

In the recent years, there has been an ongoing debate about the characteristics of different laboratory stressors, and since its introduction over 25 years ago (Kirschbaum et al., 1993), the TSST has clearly been one of the most widely utilized procedures generally (Goodman et al., 2017), and in stress research on social behavior (von Dawans et al., 2021). The advantages are clear: it is an ecologically valid natural stressor that mimics a real-life situation and reliably generates strong and measurable responses across a wide range of stress-related parameters (Allen et al., 2017). In Schweda and colleagues (2019), we applied the group TSST (Von Dawans et al., 2012) with full success, inducing HPA and SAM-reactions, as well as psychological distress, but rather found evidence for the absence of stress-driven modulations of ingroup love and outgroup hate. In the prior literature, the mere utilization of one specific stressor does not alone explain the occurrence of fight-or-flight or tend-and-befriend, although there is a slight tendency that studies applying the TSST produce more tend-and-befriend responses (see von Dawans et al., 2021 for a summary of all stress x social behavior interactions). Due to the rather moderate number of studies, I would like to refrain from making any meta-analytical statements that would require proper weighting of respective designs and experiments. However, there are studies that produce tend-and-befriend responses, and others that find fight-or-flight behavior after applying the TSST: for example, the group TSST was both applied in Von Dawans and colleages (2012) and Vinkers and colleagues (2013). Whereas the former found more prosocial behavior in terms of sharing, trust and trustworthiness, the latter found less generous sharing in a dictator game. Also, the argument that prosocial tendencies after group-wise TSST application are an artifact of increased group cohesion (Steinbeis et al., 2015) doesn’t hold: while Singer and colleagues (2017) observed increased altruism in males making decisions in a set of moral dilemmas using a standard, non-group version of the TSST, Nickels and colleagues (2017) found decreased offers in an ultimatum game and marginally less helping offers in a social risk game in TSST-exposed males (see also Sollberger et al., 2016). Consequently, variations in the stress procedure cannot - prima facie - explain the lack of stress effects in our study, particularly because the TSST-G in Schweda and colleagues (2019) was performed in full accordance with the original protocol by Von Dawans and colleagues (2012). The utilization of one specific stressor does not explain the variability of recent findings, either.

Despite the absence of direct stress effects in the IPD-MD, we could find that CORT is related to an upregulation of ingroup-love investments when conditioned on pre-stress testosterone levels, and that HR is associated with a shift towards outgroup hostile investments. Such pattern, indeed, implies that stress effects might be more fine-grained and complex than postulated by some of the previous studies on stress and human behavior and cognition. This fits very well into what the rather biologically oriented strain of literature already proposed: the stress response comprises a largely heterogenic and complex response pattern in the cross-, as well as the longitudinal section (Hermans et al., 2014; Joëls & Baram, 2009). We are, indeed, not the only ones who find indications for distinct SAM and HPA contributions. In a recent study by Potts and colleagues (2019), stress generally decreased the investment rate in a trust game. However, when regressing the skin conductance level (a measure of SAM activity), CORT, as well as response times on the investment pattern, it turned out that at longer stressor to task latencies, the simultaneous involvement of the SAM and the HPA actually predicted more trustful investments. Although these outcomes somewhat differ from our results, they also suggest a more intricate interplay in the activity of different stress markers (see also Margittai, van Wingerden et al., 2018). Beyond that, in Schweda and colleagues (2019), we furthermore propose that such interactions should be considered within the orchestra of other hormonal markers.

On contrary, we decided to make use of the economic properties of the MAST (Smeets et al., 2012) in Schweda and colleagues (2020). The MAST itself is comparably novel, but proved to be a viable alternative to the TSST in several studies (e.g. Quaedflieg et al., 2013; Schaal et al., 2019; Shilton et al., 2017; Smeets et al., 2012). We could elicit solid measurable stress responses in salivary

CORT and HR, as well as psychological indicators. Yet, the stressor failed to increase sAA levels. Despite its contradiction with the original study by Smeets and colleagues (2012) where the MAST consistently increased sAA, Quaedflieg and colleagues (2017) show that such elevations can be weak just after cessation of the stress procedure. Also, and somewhat similar to Schweda and colleagues (2020), a considerable increase in sAA was only found at a later timepoint. Accordingly, the MAST might be less prone to elicit elevations in sAA, which might hypothetically result from the different temporal profiles of the MAST (total duration ~ 15 Min.) and the TSST (total duration ~ 20 Min), as well as the possibly less intense nature of the MAST (e.g. social evaluation by one experimenter in the MAST vs. a panel of at least two confederates in the TSST). Still, we believe that the MAST is a robust and economic procedure that outperforms other physiological single person stressors (Smeets et al., 2012), and that the stress response in Schweda and colleagues (2020) is, most likely, real.

Stressors, however, might perhaps differ in terms of *what* they actually elicit. Contrary to Margittai and colleagues (2015), who applied a group TSST, we did not find an increase in generosity toward socially close, but rather a drop in generosity towards socially remote individuals after application of the MAST. The occasional lack of the MAST's capacity to evoke increases in sAA might already indicate that the neurohormonal changes the task elicits differ from the TSST, which, in turn, could result in qualitatively and quantitatively distinct modulations of behavior and cognition. In fact, the TSST is somewhat unique with respect to its demands on verbal activity, the cognitive performance, as well as the intensity of social evaluation: participants need to spontaneously generate a narrative of themselves and convey it to the committee whose members are not socially reactive, and hence, behave opposite to what someone expects from its usually cooperative environment. In comparison, the MAST only includes the arithmetic task which is presented to only one experimenter instead of multiple ones. The group TSST creates further uncontrollability and drives social comparison by including multiple participants (see also Vors et al., 2018, for a qualitative assessment of participants' experiences during the TSST). A few findings in the prior literature reflect such differences reasonably well: Rosenbaum and colleagues (2018) report wide-ranging and pertinent activation differences in dorsolateral and inferior frontal, as well as parietal brain regions during the TSST arithmetic task (vs the control condition). On contrary, Schaal and colleagues (2019), who contrasted activations during the arithmetic task during the MAST vs. the control condition, find similar, but more ventrally located, orbitofrontally-extending activations in dorsolateral, and no differences in parietal brain areas. The cold-water immersion task, on the other side, produced a downregulation in left frontopolar regions. Although these results do not perfectly fit for comparison because they do not directly contrast the MAST and the TSST, they tentatively suggest that both tasks differ in the neural activity they evoke. Differences in neural signatures of psychosocial and physical stressors are also found in a meta-analysis by Kogler and colleagues (2015). As it is known that stress effects can be dose-dependent (compare e.g. Margittai, Nave, et al., 2018; Metz et al., 2020), such differences could, indeed, be decisive. Hence, it also remains an empirical question whether the MAST would have produced similar results in the IPD-MD (see also von Dawans et al., 2021). In fact, the only study that compares the effect of different (physical) stressors on social behavior, and finds differences, is the one from Von Dawans and colleagues (2018). Here, the mere use of the cold pressor task reduced prosocial sharing, but adding a socially evaluative component restored it again. On a descriptive level, the combined social evaluative and physiological stressor even increased the sharing behavior, tentatively indicating that that both components must be included to produce tend-and-befriend patterns, whereas a primarily physical component rather drives a fight-or-flight response. This does, however, not explain the emergence of fight-or-flight-type patterns in Schweda and colleagues (2020) as the socially evaluative component is even more prominent in the MAST than in the socially evaluative cold pressor test used in Von Dawans and colleagues (2018).

In order to explain the occurrence or non-occurrence of tend-and-befriend and fight-or-flight effects, it could be equally important to incorporate the game theoretical properties of the applied paradigm. Studies that actually use economic games after stress find a number of paradigms being reactive to stress effects (see table 1). Multiple aspects become obvious here: first, the trust game, the ultimatum and the dictator game (or variations of it, including the social discounting paradigm) are the most prominent ones. No single one of these is a consistent subject to either tend-and-befriend or fight-or-flight responses. At large, it also seems that tend-and-befriend and fight-or-flight might manifest differently under different externalities. That is, tend-and-befriend can produce more sharing (e.g. Von Dawans et al., 2012) but also less rejection of offers in ultimatum games (Steinbeis et al., 2015). Fight-or-flight might manifest in less trust in trust games (e.g. Potts et al., 2019), as well as less sensitivity to harm-based norms when deciding to take away resources from others (Schweda et al., 2020). Also, it appears even more that the TSST protocols tend to produce more tend-and-befriend responses. Yet, the scope here would automatically discard studies that also use moral dilemmas (e.g. Starcke et al., 2011). Also, the number of studies is still moderate, and a mere addition of studies would disregard differences in quality and strength of evidence. Hence, again, I do not intend to make any meta-analytic statements that would require a more thorough review and weighting of design-features and statistical results.

Study	Game	Stressor	Response
Von Dawans et al., 2012	Dictator Game Trust Game	G-TSST	Increased sharing Increased trust and reciprocity to trust
Von Dawans et al., 2019	Dictator Game Trust Game	G-TSST	Increased sharing Increased reciprocity to trust
Vinkers et al., 2013	Dictator Game (C) Ultimatum Game (A&R)	TSST	Decreased in early and late stress Decreased acceptance rates in early vs. late stress
Sollberger et al., 2016	Dictator Game (C)	TSST	Increased sharing in participants with low proenvironmental orientation. Reduced amount shared in participants who decided to share.
Youssef et al., 2018	Ultimatum Game (A&R)	TSST	Proposal rejection rate reduced in female participants
Margittai et al., 2015	Social Discounting	G-TSST	Increased generosity toward socially close people
Nickels et al., 2017	Ultimatum Game (P) Prisoners' Dilemma	TSST	Marginally significant Stress x Gender interaction Increased cooperation in female participants
FeldmanHall et al., 2015	Trust Game	CPT	Stress induced decrease in trust toward humans, but increased trust in non-social lottery
Potts et al., 2019	Trust Game	CPT + SECPT	Decreased trust in collapsed stress group
von Dawans et al., 2018	Dictator Game Trust Game	CPT + SECPT + SE*	Pattern indicated reduced sharing by CPT, restored by SECPT Pattern indicated reduced reciprocity to trust by CPT, restored by SECPT
Steinbeis et al., 2015	Trust game Ultimatum Game (A&R)	Anticipatory stress	Less trust Increased acceptance rates
Margittai, van Wingerden, et al., 2018	Social Discounting	Yohimbine + hydrocortisone administration	Increased generosity toward socially close people when hydrocortisone is administered, but the effect is offset when yohimbine is administered additionally
Bendahan et al., 2017	Social** Risk Taking	TSST-G	Decreased other-regarding considerations

(C) = To charity

(A&R) : Acceptance or Rejection

(P) : Proposal

* SECPT: Socially Evaluated Cold Pressor Task; SE: Social Evaluation Only

** vs. Non-Social

Table 1. Summary of studies that assess stress effects in economic games. This content is partly inspired by von Dawans and colleagues (2021)

On a more global level, all above-mentioned studies that could find fight-or-flight and tend-and-befriend behaviors apply simple, two-partner interaction games. Although Steinbeis and colleagues (2015) implemented a minimal group scenario and compared stress-related social behavior toward in- and outgroup members, decisions were never causing gains or losses toward other group members than the two players involved. The IPD-MD paradigm used in our experiment (Schweda et al., 2019) clearly required a more thorough representation of interdependence from the own group members and the outgroup members, which distinguishes this approach from other studies and might enact different motifs than in standard two-player interaction games. Indeed, one strain of literature suggests that ingroup favoritism is mostly explainable via reciprocity expectation, and that this incorporates a rather strategic drive to maximize future resources (Durrheim et al., 2016; Rabbie et al., 1989). Even though the IPD-MD was a one-shot game in our first study, the players' focus and evaluation of reciprocity expectation might be internalized and generalized (see also Yamagishi & Kiyonari, 2000). Of course, reciprocity expectation plays a huge role in two-player interaction games, too: in the trust game, the player expects the trustee to reciprocate, and in the prisoner's dilemma, a cooperative proposer expects non-defection. Yet, and although they are largely strategic in nature, two-player games could leave more space for social sentiments: individuals involved in dyadic interactions might construct a more socio-affectively thorough representation of the other person, which can then more quickly produce, for instance, feelings of warm glow, vicarious reward (e.g. Andreoni, 1989; Hartmann et al., 2017; Morelli et al., 2015) or perhaps even spite. Groups, however, are more abstract, and players in the IPD-MD in Schweda and colleagues (2019) were not supposed to interact prior to the IPD-MD, which could have caused the representation of the others to be less socio-emotional in nature. In support of this interpretation, we found that ingroup-friendly contributions in the IPD-MD were correlated with performance in the Cognitive Reflection Test (Frederick, 2005), suggesting that participants who grasped the payoff structure of the IPD-MD better invested more into the within-group pool. Our result is furthermore backed up by a prior, unpublished project from our laboratory using minimal groups with 6 attendant participants playing the IPD-MD after undergoing the group TSST vs. a control condition. Here, and in line with Schweda and colleagues (2019), the data similarly yielded evidence for the absence of an effect, or a very small effect. These results are valuable in that they restrict the space in which stress effects occur. If this interpretation proves to be correct, stress only modulates decisions that go along with a more personal and socio-affective representation of others. As a mechanistic foundation for this notion, it might be assumed that stress only modulates activity in brain areas specifically recruited in primarily altruistic vs. primarily strategic social decision making (see also Cutler & Campbell-Meiklejohn, 2019, for a meta-analysis).

This interpretation does not exclude the possibility of a stress effect on parochial altruism, but rather suggests that stress effects could have been masked by an insufficiently social representation of the groups and their members. In addition, research has continued to point out more and more intricacies concerning the nature of intergroup social preferences. In fact, between-group pool investments in the IPD-MD only capture the intention to benefit the ingroup by harming the outgroup, which is termed *strong parochial altruism* (Böhm et al., 2020). However, in an intergroup conflict setting, multiple types of preferences are possible. Böhm and colleagues (2021) recently evaluated the stability of group-related social preferences within the Austrian election period in 2016. Here, social preferences are described by the following dominant tendencies: Universal altruists (willing to share with in- and outgroup), universal egoists (not willing to share with anyone), universal competitors (willing to exhibit spiteful behavior toward in- and outgroup members), weakly parochial altruists (willing to benefit the ingroup without any spiteful attitudes toward the outgroup), strongly parochial altruists (willing to benefit the ingroup and exhibit costly spiteful behavior toward the outgroup) and parochial egoists (*not* willing to benefit the ingroup, but exhibit spite toward the outgroup). These preferences were shown to be relatively stable across the study period. What is, however, more intriguing is the overall distribution of the preferences. The total proportion of strong parochial altruists only amounts to roughly 20% in all three study samples, whereas universal altruists took around 30% of the share. On the other side, weakly parochial altruism was represented by about 11%, and universal competitors and parochial egoists both made up about 20-25%. This large heterogeneity of phenotypes with a relatively low number of people even willing to exhibit hostile behavior toward an outgroup within a realistic intergroup conflict (here: 40-45%) might appropriately answer the lack of identifiable stress effects in Schweda and colleagues (2019). If this proportion of people is generally willing to put their investments against an outgroup at all, and, hence, only their preferences are considered malleable, then the effect size of any manipulation on the overall mean of between-group pool investments is expected to drastically shrink. On the other side, about 60% of participants in Böhm and colleagues (2021) consists of universal altruists, weakly parochial altruists and strong parochial altruists which would exhibit a preference for investments favorable to the ingroup. It must also be noticed that the intergroup conflict produced during the election in Austria might be substantially more powerful than the laboratory situation in our study. Hence, although the sample size in Schweda and colleagues (2019) is considerable in comparison to other studies in the field, it might have still been too small to detect differences produced by only a part of the participants. This line of argument, as well as the dominant use of two-player games in prior stress literature (see table 1) also call into question whether intergroup settings are the best choice for efficient research on stress and social behavior, at least without large samples.

Our finding that the MAST could mitigate the framing effect at large social distances (Schweda et al., 2020) comes into play right here: possibly due to its simple structure and traceability, the social discounting paradigm could, indeed, produce stress effects, and we have potentially found additional evidence for a co-occurrence of fight-or-flight and tend-and-befriend, at least when our results are considered in conjunction with prior research. Our findings have two major implications: first, it again supports the role of perceived social closeness on stress effects (see Margittai et al., 2015; Margittai, van Wingerden, et al., 2018), which is far from trivial, precisely because it gives us further clues that the nature of the social sentiments elicited matter for the stress response. Second, the externality, and hence, whether the money is distributed towards a recipient or away for her or him, appears to matter, as well. The latter point makes the case for a more detailed account of how fight-or-flight actually manifests – as a decrease of the threshold to harm, or a decline in the willingness to help, or both. Although our observations are somewhat supported by Bendahan and colleagues (2017), who found that stress could reduce other-regarding consideration, a direct comparison of a positive and a negative externality has not been made before. Hence, our data suggest that fight-or-flight could rather come about in form of an increased propensity to withdraw goods from others. Note, however, that such disinhibition to harm is not consistently found and could be stressor-dependent (Steinbeis et al., 2015; Von Dawans et al., 2012; von Dawans et al., 2018).

Surely, large social distances differ from badly connoted outgroups, particularly because participants are explicitly instructed to avoid assigning people they resent to any of the social distances, including social distances 50 and 100. This is why it is difficult to interpret our observations in Schweda and colleagues (2020) in terms of the hypothesis that stress increases “parochial altruism”. Yet, large social distances most likely do not just represent “someone” as, for instance, a recipient in most of the dictator games. Potential evidence for this can be derived from a simple comparison of mean investment rates between the dictator game, which often fall between 20% and 30% (see Camerer, 2011; Engel, 2011), and investment rates toward socially remote people in the social discounting paradigm, which often reach about 10% for social distance 100 and 10-15% for social distance 50 (see e.g. Jones & Rachlin, 2006; Margittai et al., 2015; Margittai, van Wingerden, et al., 2018; Schweda et al., 2020; Strombach et al., 2014, 2015). Large social distances are seemingly perceived as more socially distant, or at least not as behaviorally relevant for one’s actions than a stranger in the dictator game. This could be due to the quick establishment of a feeling of solidarity and social closeness to a single mentioned interaction partner that is also involved in a laboratory study, or the anticipation of future interaction opportunities during the same experiment. Whatever it might be, supplemental analyses of our data in Schweda and colleagues (2020) show that the stress x frame interaction takes place exactly at social distances 50 and 100. Our stress manipulation, therefore, facilitates violations of the “do-not-harm"-principle toward individuals which are considered even more remote than anonymous co-players in other two-partner interaction games. This suggests that a certain level of *unconcern with the other person* is needed for fight-or-flight to occur. On the other hand, it might be speculated that the frequently observed tend-and-befriend responses in two-player games (Von Dawans et al., 2012, 2019) might be a result of implicit feelings of closeness created via the experimental context. Interestingly, previous literature on framing suggests that there is no give/take framing effect in the dictator game (see Dreber et al., 2013; Goerg et al., 2019). Yet, instead of invalidating our results from Sellitto and colleagues (2021) and Schweda and colleagues (2020), these observations further emphasize the very special nature of the social discounting paradigm, and that it possibly measures behavior that is affected by other social sentiments.

Anyway, the link between the notion that stress promotes hostile acts - or at least lowers the threshold for infliction of harm - toward socially distant individuals, and the idea that stress promotes hate against outgroups might, hypothetically, lie in that outgroup members are possibly categorically shifted toward higher social distances. For instance, the concept of “infrahumanization” postulates that outgroup members are attributed with less uniquely human characteristics such as secondary emotions like felicity and embarrassment (Demoulin et al., 2009; Leyens et al., 2001). Hence, outgroup members could be just perceived as individuals with whom one should be less concerned. This would also imply that the affectively hostile component is not necessary to produce overtly hostile reactions under stress. Yet, all in all, questioning how groups are perceived differently in comparison to “the others” is quite fundamental for the understanding of the proximate and ultimate mechanisms of social behavior, and such discussion might quickly escalate in a reassessment of the theoretical cornerstones of the evolution of social behavior. Stress research must exactly pick apart which social sentiments are modulated by stress, and which of them simply remain unaffected. Finding the answers might, indeed, require a more granular approach.

In summary, the studies presented clearly pointed out boundary conditions of stress effects on social behavior and extended the scope of possible determinants, including the social sentiment involved, the applied stressor, the social distance, and the framing effect. An integrated look at our findings in conjunction with other literature reveals that no single stressor and no single paradigm could reliably and consistently evoke either the tend-and-befriend or the fight-or-flight responses. Alas, the evidence at hand suggests a complex interaction between the active stress components, the social sentiments elicited within the respective paradigm, including the level, depth or socio-affective nature of representation of other players, the externality of the decision, as well as baseline individual social preferences. What makes the topic even more unwieldy is the probable involvement of sex and sex hormones and the duration of stress exposure (Nickels et al., 2017; Schaal et al., 2019; von Dawans et al., 2021). In the next section, I will expand the strengths and weaknesses of the studies presented in sections 3, 5 and 6. Based on these aspects, I will propose and discuss future outlooks and possible trajectories of research on stress and social behavior.

7.3. Strengths, limitations and outlook

With Schweda and colleagues (2019), we were – to my knowledge - the first ones that applied a realistic intergroup setting to assess social behavior under stress. The IPD-MD was used and evaluated in a considerable number of contexts before and was deemed a robust and replicable paradigm across a variety of studies (see also Thomae et al., 2015). Clearly, the sample was sizeable, which allowed us to make conclusions about the magnitude of a possible effect with relatively small error bounds. The stress reaction induced turned out to be robust with solid increases in salivary cortisol, alpha-amylase, and increased feelings of arousal and stress. Based on this, we could provide evidence for a relatively narrow error span around the effect size of zero, which could be interpreted as evidence for the necessity of a certain representability of social qualities in the respective object of the altruistic or hostile act. Our null result does, by no means, imply that the idea of a simultaneous occurrence of fight-or-flight and tend-and-befriend in an intergroup setting can be fully discarded. Yet, the design must be optimized to make such effects tangible. Multiple alternative approaches could be realized: first, the intergroup bias (including ingroup love and outgroup hate) can be amplified by, for instance, using more sensitive group settings (e.g. using pre-existing resentments). Second, an augmented socio-affective representation of the group members could be experimentally induced by, for example, letting players solve tasks together and against each other. Finally, recent research on social behavior in groups points towards heterogeneity of inter-individual social preferences (see Böhm et al., 2021) that needs to be taken seriously. Future studies should incorporate ways to operationalize the full range of these preferences (e.g. universal altruism or parochial egoism) and explore how the stress response might augment individual baseline tendencies, or even shift these initial preferences. Here, paradigms such as the intergroup parochial and universal cooperation game by Aaldering and Böhm (2020), or the intergroup social value orientation slider measure by Böhm and colleagues (2021) might prove to be fruitful approaches. Of course, such undertakings would require adequate sample sizes.

The study by Sellitto and colleagues (2021) offers valuable and robust evidence for the existence of a powerful give-take framing effect in social discounting. It has been shown that the framing effect is easily reproduced under a multitude of settings, which makes it even more attractive for practical application outside the laboratory. The existence of framing effects in social discounting is far from being trivial since they are not ubiquitous (see Dreber et al., 2013; Goerg et al., 2019). Future research should, however, extend these findings by assessing how social discounting is modulated by a take frame under alternative scenarios. For instance, we could not induce an increase in the V-parameter after the frame manipulation. It is possible that this is due to an attractor at around 50% of the proportion shared at low social distances (i.e. at a "fair split"), which makes the social discount function more “stable” at this point (see also Dreber et al., 2013). Note, however, that the V-parameter has been shown to be malleable before (e.g. Margittai et al., 2015; Margittai, van Wingerden, et al., 2018). Such explanation might, therefore, include the application of different social norms in interactions with socially close individuals where the act of taking is perhaps less sanctioned, particularly when it results in an equal allocation of the respective good. Furthermore, our data indicates that individuals possess a very strong reluctance to reduce the recipients’ outcome, so that numerous participants ended up with a flat discount function. On the one hand, this integrally supports the crucial role of the “do-not-harm"-principle. On the other hand, flat discount functions might complicate the data analysis, particularly when researchers rely on small sample sizes and depend on parametric methods, e.g. because they would like to include further predictors of the framing effect and the respective regression models are biased because of non-normality or heteroskedasticity. Future research could attempt to widen the variability of distributions of investments - and hence, the discounting parameters - by exploiting magnitude effects. Accordingly, Bechler and colleagues (2015) found that the overall investment rate decreases as a function of initial endowments. Additionally, a manipulation of the conversion rate of amount withdrawn from the recipients’ endowment to the gain for the participant could also drag the social discount function down at larger social distances. For instance, whereas in the original version, a withdrawal of 5€ from the recipients’ endowment would lead to an equivalent gain of 5€ for the participant, a different condition could decrease the necessary withdrawal for the same gain to, for example, 2.50€. Such manipulation could perhaps make the k-parameters' distributions approximate normality slightly better because participants would be more willing to withdraw money from the recipient for the sake of a higher gain (vs. the actual loss for the recipient). This way, the paradigm could be more practical in scenarios where a researcher would like to assess which types of manipulations can increase the framing effect, and hence, the participants' moral sensitivity in decision making. Also, the data could be better handled by mixed non-linear models (see Young, 2017).

With regard to the neuroimaging results in Sellitto and colleagues (2021), I must openly admit to my hesitancy to deal out too much critique. Again, I was not involved in the analysis, and I do not feel qualified to criticize the techniques used. Of course, the sample size might have compromised the power in some analyses. The power calculated a priori for the mediation model, which amounted to 60%, is certainly subpar. However, we could observe multiple times that the framing effect is strong, and this could translate into better signal-to-noise ratios in BOLD-signals. With all honesty, I lack specialization to make any statements on power of specific fMRI analysis techniques, and particularly advanced ones like dynamic causal modelling.

In the third study (Schweda et al., 2020) we have again shown that the social discounting paradigm is highly sensitive to stress effects. It remains to be emphasized that this is the third time in a row that this paradigm was shown to be reactive in stress settings. I am not aware of any studies that report null results - published or unpublished. Furthermore, we could apply a highly robust economic game approach to directly show that fight-or-flight can manifest in a corruption of the willingness to act according to social norms, which complements prior research using moral dilemmas (Starcke et al., 2011), as well as social and non-social lottery games (Bendahan et al., 2017).

As the person who has analyzed the data, I again stress that the paradigm should be optimized in terms of statistical approximability with hyperbolic discount functions, for example by applying the suggestions mentioned above. Surely, the models reported in the manuscript remain interpretable as such: assessing their sensitivity, we could not detect any violations in normality in the reported models' residuals, and the application of heteroskedasticity-robust standard errors, as well as outlier-robust mixed linear models with an underlying student-t likelihood function yielded equivalent results. However, the two-parameter hyperbolic function (Jones & Rachlin, 2006) indubitably carries the advantage of being able to directly model and differentiate between generosity toward socially close and distant recipients. This difference is neither arbitrary, nor just an ungrounded top-down assumption of the model: first, prior research – including studies from our own lab - could show that such differences matter in many domains (e.g. Margittai et al., 2015; Margittai, van Wingerden, et al., 2018; Soutschek et al., 2017; Strang et al., 2017). Second, it is possible to recover two separate, but correlated factors when treating the investments towards each social distance as items. I conducted these analyses with the investments in the give frame from Schweda and colleagues (2020), and found that a parallel analysis, as well as Velicer's minimum average partial (Velicer, 1976), equivalently indicate the existence of two components or factors. Here, I could observe that social distances from 1 to 20, and particularly 1 to 5, strongly load on one factor, and social distances 5 to 100, and particularly 20, 50 and 100, load on a second factor. Hence, the divergence of generosity towards small and large social distances is not just of theoretical importance but is also supported by data driven methods. In our regression analyses, we could, thus, very well display differences between the slopes of the discount functions, but we lacked the possibility to extract and quantify differences between the intercepts.

In line with this, I must also appreciate the critique that the classical (give frame) social discounting paradigm might produce floor effects at larger social distances. On the right-hand side of the social distance scale, people are less willing to share, which sometimes results in a right-skewed distribution of investments. In this case, it could become difficult to properly extract k-parameters (i.e. the distribution of possible parameter values is broad and estimates could become imprecise, see also Young, 2017). In our case, stress might have, indeed, also reduced sharing toward socially distant others in the give frame, but this could have remained undetected due to this lower bound. Again, such difficulties could be circumvented by exploiting, for instance, magnitude effects that might draw the function upward.

The studies presented now finally join the ranks of the very recent developments that indicate that stress is way more than an unspecific survival program. Yet, globally speaking, the investigation how stress modifies human social perception, cognition, reasoning and decision making is still in its infancy. In fact, the first experimental evidence for a tend-and-befriend reaction in humans will celebrate its 10th anniversary soon (Von Dawans et al., 2012). A review of the literature brings us to an easy message: there is still much to do. We are still largely ignorant about when stress brings about modulations of social behavior, and we are still not able to predict under what circumstances stress triggers benevolent or belligerent responses. Alas, I believe that we must somewhat depart from the expectation of a simple solution here, respecting some, if not many, boundary conditions. Indeed, a broader look at stress studies suggests that responses might differ as a function of the stressor applied, the time at which social behavior is measured, the game applied, the social sentiment elicited, the externality of the decision, the neurohormonal environment, and the gender and personality of the participant. What would help here is even a basic understanding of what tend-and-befriend and fight-or-flight mean in order to properly classify the findings up to this point. For instance, should an increased reluctance to share money with a recipient in a dictator game (Vinkers et al., 2013) be interpreted as a fight-or-flight response or is the motivation behind this rather parochial, i.e. the player keeps the money to invest into his own network in case of emergency? Why does stress sometimes cause egoistic choices in a positive externality (Vinkers et al., 2013), and sometimes more unconcern in a negative externality (Schweda et al., 2020)? And how important is the relationship between aggression and self-inhibition (Schweda et al., 2020)? Does stress really promote a preconfigured set of social reactions, or does it just amplify pre-existing preferences and sentiments? Is the individual's role within the group important? What happens under chronic stress? And of course, what are the mechanistic underpinnings of such responses or shifts?

I hope the reader gets an intuition for the nature of the problem: research on stress and social behavior is research within multi-layered but interacting complex systems. Such research requires an interdisciplinary perspective, but must remain integrative. Nonetheless, recent trends in science offer promising approaches. First of all, a positive turnaround toward pre-registered multi-laboratory studies with large sample sizes (e.g. Akkermans et al., 2019; Hagger et al., 2016; Kopiske et al., 2016; Maier et al., 2020) could prove fruitful for the dissemination of these complex interactions. Here, participants from both sexes could be allocated across conditions with economic games with varying levels of social representation of others, strategic profiles, as well as their externality of the choice. The focus on personality and social role (i.e. baseline social preferences) should be widened by including simple measurements on social value, empathy and aggression (see e.g. Aaldering & Böhm, 2020). Different stressors could be incorporated, and a complete set of possible mediators, neurohormonal stress-markers, but also sex-hormone levels must be considered to allow for mechanistic approximations. Surely, such projects would require solid amounts of patience, planning, organization, and above all, cooperation. Yet, I personally believe that the behavioral and social sciences are amid a considerable transformational process, and that this is a good moment for such large-scale endeavors that would possibly receive unprecedented support nowadays.

A second development in the recent years has been the larger focus on behavioral modelling. Models can – if successfully applied – open up the black box and infer components or determinants underlying behavior, which could otherwise not be deduced from just observing the behavior alone. In order to disentangle what stress does to our social brain, the researcher could choose from a large set of models that are applied in many niches of behavioral science today. For instance, the drift diffusion model by Yu and colleagues (2021) could help to further differentiate whether a fight-or-flight behavior would go along with a promotion of the goal to maximize the own reward or a reduction in the willingness to avoid discomfort to the other. The application of these models in conjunction with neuroimaging (e.g. Hutcherson & Tusche, 2021) could promote the understanding of how stress modulates social value representations in the brain. New insights on how social behavior is mechanistically realized could, thus, help stress research to identify the relevant pathways.

Finally, the evaluation of stress effects on social behavior outside the laboratory must be promoted. What role do acute or chronic stressors play in the interactions between individuals or groups of individuals? Can stressful events even promote civil unrest or wars? In political science, it has been long discussed whether economic shocks increase the likelihood of conflict (see e.g. Chassang & i Miquel, 2009; Janus & Riera-Crichton, 2015). It is, however, unclear which role acute or potentially chronic stress might have on interpersonal decision making that lead to such conflicts. A further and very obvious area of application is psychiatry. It is, for example, known that patients with complex and chronic post-traumatic stress disorder exhibit aberrant social and emotional patterns (Cloitre et al., 2013; Saraiya et al., 2021). The way cumulative effects of sustained trauma disrupt executive control and valuation circuits, and how the constant arousal-associated stress amplifies to the maintenance of such patterns remains to be a topic of huge interest.

7.4. Conclusion

In this dissertation, I presented evidence from three publications with the aim to assess the impact of stress on social decision making and behavior. Results from the first study show that stress itself has no direct effect on intergroup decision making, but that activity of the HPA and the SAM differentially contribute to participants' tendency to behave in a within-group friendly or between-group hostile fashion. The second study evaluated a novel modification of the social discounting paradigm where an individual decides whether to withdraw money from recipients at various social distances for the sake of its individual gain. This was compared to the classical social distance paradigm where the individual allocates money from an own endowment toward a set of close or distant others. This simple manipulation has boosted investment rates at larger social distances considerably. My co-authors and I could furthermore provide evidence for differential neural signatures for decisions in the give and the take frame. In the third study, we applied this paradigm in conjunction with a hybrid stressor to explore how stress modulates this social framing effect. We found participants in the stress condition to exhibit a decreased framing effect at larger social distances. In other words, decisions of the stressed participants were less affected by the take frame manipulation, and hence, stressed individuals were less reluctant to take away money from socially remote others. When interpreted in the light of prior evidence, it appears that stress corrupts the adherence to interpersonal norms that are recruited to prevent decisions allegedly harmful to others.

In terms of answering which response – fight-or-flight vs. tend-and-befriend – is the predominant one, our studies could clearly add pieces to the puzzle: although exploratory analyses revealed distinct relationships of ingroup-friendly and outgroup-hostile tendencies with cortisol and heart rate response, it generally seems that our intergroup context was less sensitive to overall stress effects. Yet, the social discounting task was even more so. Considering our results in conjunction with prior literature on stress and economic games points to a complex interplay between the stress modality and component, the experimental approach, as well as the social sentiments involved – either by state or by trait. Future research must face the challenge of picking these relationships apart. On the long run, this could enable us to not only understand and find practical use of how and when stress exhilarates or compromises social interactions – for instance in parliaments, hospitals, businesses and courtrooms - but also to reconsider the role and complexity of such hard-wired physiological responses for our evolution as social animals. Perhaps this understanding will reveal traces of how the human organism has evolved as a function its social and cultural environment, and what was expected from us to survive under conditions as they existed in our ancient history.

8. References

Aaldering, H., & Böhm, R. (2020). Parochial versus universal cooperation: Introducing a novel economic game of within-and between-group interaction. *Social Psychological and Personality Science*, *11*(1), 36–45.

Abbink, K., Brandts, J., Herrmann, B., & Orzen, H. (2012). Parochial altruism in inter-group conflicts. *Economics Letters*, *117*(1), 45–48.

Agnew, R. (2001). Building on the foundation of general strain theory: Specifying the types of strain most likely to lead to crime and delinquency. *Journal of research in crime and delinquency*, *38*(4), 319–361.

Agnew, R. (2005). *Why do criminals offend?: A general theory of crime and delinquency*. Roxbury Los Angeles, CA.

Agnew, R. (2017). Building on the foundation of general strain theory: Specifying the types of strain most likely to lead to crime and delinquency. *Recent Developments in Criminological Theory*, 311–354.

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., Busch, V., Lothwesen, K., Elvers, P., & Fischinger, T. (2019). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*, *33*(6), 1099–1118.

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). The Trier Social Stress Test: Principles and practice. *Neurobiology of Stress*, *6*, 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001

Allen, J. J., Anderson, C. A., & Bushman, B. J. (2018). The General Aggression Model. *Current Opinion in Psychology*, *19*, 75–80. https://doi.org/10.1016/j.copsyc.2017.03.034

André, J. (2014). Mechanistic constraints and the unlikely evolution of reciprocal cooperation. *Journal of evolutionary biology*, *27*(4), 784–795.

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy*, *97*(6), 1447–1458. https://doi.org/10.1086/261662

Archambault, C., Kalenscher, T., & de Laat, J. (2020). Generosity and livelihoods: Dictator game evidence on the multidimensional nature of sharing among the Kenyan Maasai. *Journal of Behavioral Decision Making*, *33*(2), 196–207.

Atsak, P., Guenzel, F. M., Kantar-Gok, D., Zalachoras, I., Yargicoglu, P., Meijer, O. C., Quirarte, G. L., Wolf, O. T., Schwabe, L., & Roozendaal, B. (2016). Glucocorticoids mediate stress-induced impairment of retrieval of stimulus-response memory. *Psychoneuroendocrinology*, *67*, 207–215. https://doi.org/10.1016/j.psyneuen.2016.02.006

Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin*, *137*(6), 881–909. https://doi.org/10.1037/a0025354

Balliet, D., Wu, J., & De Dreu, C. K. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological bulletin*, *140*(6), 1556.

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics*, *11*(2), 122–133.

Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making*, *8*(2), 71–83. https://doi.org/10.1002/bdm.3960080202

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, *94*(2), 74–85.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Basu, K. (1984). *The less developed economy: A critique of contemporary theory*. Blackwell.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological bulletin*, *115*(2), 243.

Bechler, C., Green, L., & Myerson, J. (2015). Proportion offered in the Dictator and Ultimatum Games decreases with amount and social distance. *Behavioural processes*, *115*, 149–155.

Belfi, A. M., Koscik, T. R., & Tranel, D. (2015). Damage to the insula is associated with abnormal interpersonal trust. *Neuropsychologia*, *71*, 165–172. https://doi.org/10.1016/j.neuropsychologia.2015.04.003

Bendahan, S., Goette, L., Thoresen, J., Loued-Khenissi, L., Hollis, F., & Sandi, C. (2017). Acute stress alters individual risk taking in a time-dependent manner and leads to anti-social risk. *European journal of neuroscience*, *45*(7), 877–885.

Ben-Ner, A., McCall, B. P., Stephane, M., & Wang, H. (2009). Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. *Journal of Economic Behavior & Organization*, *72*(1), 153–170. https://doi.org/10.1016/j.jebo.2009.05.007

Berger, J., Heinrichs, M., von Dawans, B., Way, B. M., & Chen, F. S. (2016). Cortisol modulates men's affiliative responses to acute social stress. *Psychoneuroendocrinology*, *63*, 1–9. https://doi.org/10.1016/j.psyneuen.2015.09.004

Berkowitz, L., & Harmon-Jones, E. (2004). Toward an understanding of the determinants of anger. *Emotion*, *4*(2), 107.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, *442*(7105), 912–915. https://doi.org/10.1038/nature04981

Böhm, R. (2016). Intuitive Participation in Aggressive Intergroup Conflict: Evidence of Weak Versus Strong Parochial Altruism. *Frontiers in Psychology*, *7*, 1535. https://doi.org/10.3389/fpsyg.2016.01535

Böhm, R., Fleiß, J., & Rybnicek, R. (2021). On the Stability of Social Preferences in Inter-Group Conflict: A Lab-in-the-Field Panel Study. *Journal of Conflict Resolution*, *65*(6), 1215–1248. https://doi.org/10.1177/0022002721994080

Böhm, R., Rusch, H., & Baron, J. (2020). The psychology of intergroup conflict: A review of theories and measures. *Journal of Economic Behavior & Organization*, *178*, 947–962.

Böhm, R., Rusch, H., & Gürerk, Ö. (2016). What makes people go to war? Defensive intentions motivate retaliatory and preemptive intergroup aggression. *Evolution and Human Behavior*, *37*(1), 29–34. https://doi.org/10.1016/j.evolhumbehav.2015.06.005

Böhnke, R., Bertsch, K., Kruk, M. R., Richter, S., & Naumann, E. (2010). Exogenous cortisol enhances aggressive behavior in females, but not in males. *Psychoneuroendocrinology*, *35*(7), 1034–1044.

Bolton, G. E., Katok, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, *27*(2), 269–299. https://doi.org/10.1007/s001820050072

Bornstein, G. (1992). The free-rider problem in intergroup conflicts over step-level and continuous public goods. *Journal of Personality and Social Psychology*, *62*(4), 597.

Bornstein, G. (2003). Intergroup Conflict: Individual, Group, and Collective Interests. *Personality and Social Psychology Review*, *7*(2), 129–145. https://doi.org/10.1207/S15327957PSPR0702_129-145

Bornstein, G., Erev, I., & Goren, H. (1994). The effect of repeated play in the IPG and IPD team games. *Journal of Conflict resolution*, *38*(4), 690–707.

Bowles, S. (2006). Group Competition, Reproductive Leveling, and the Evolution of Human Altruism. *Science*, *314*(5805), 1569–1572. https://doi.org/10.1126/science.1134829

Buchanan, T. W., & Preston, S. D. (2014). Stress leads to prosocial action in immediate need situations. *Frontiers in Behavioral Neuroscience*, *8*, 5. https://doi.org/10.3389/fnbeh.2014.00005

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, *60*(5), 930–940.

Burani, K., Gallyer, A., Ryan, J., Jordan, C., Joiner, T., & Hajcak, G. (2020). Acute Stress Reduces Reward-Related Neural Activity: Evidence from the Reward Positivity. *BioRxiv*, 2020.01.08.898593. https://doi.org/10.1101/2020.01.08.898593

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Buruck, G., Wendsche, J., Melzer, M., Strobel, A., & Dörfel, D. (2014). Acute psychosocial stress and emotion regulation skills modulate empathic reactions to pain in others. *Frontiers in psychology*, *5*, 517.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton university press.

Campbell, D. T. (1965). Ethnocentric and other altruistic motives. *Nebraska symposium on motivation*, *13*, 283–311.

Cannon, W. B. (1914). The emergency function of the adrenal medulla in pain and the major emotions. *American Journal of Physiology-Legacy Content*, *33*(2), 356–372.

Cannon, W. B. (1932). Natural Defenses of the Organism. In *The wisdom of the body* (S. 213–215). W W Norton & Co.

Cartwright, E. (2016). A comment on framing effects in linear public good games. *Journal of the Economic Science Association*, *2*(1), 73–84.

Casto, K. V., & Edwards, D. A. (2016). Testosterone, cortisol, and human competition. *Hormones and Behavior*, *82*, 21–37. https://doi.org/10.1016/j.yhbeh.2016.04.004

Chang, L. J., & Sanfey, A. G. (2013). Great expectations: Neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience*, *8*(3), 277–284. https://doi.org/10.1093/scan/nsr094

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion. *Neuron*, *70*(3), 560–572. https://doi.org/10.1016/j.neuron.2011.02.056

Chassang, S., & i Miquel, G. P. (2009). Economic shocks and civil war. *Quarterly Journal of Political Science*, *4*(3), 211–228.

Chatterton, R. T., Vogelsong, K. M., Lu, Y., Ellman, A. B., & Hudgens, G. A. (1996). Salivary α-amylase as a measure of endogenous adrenergic activity. *Clinical Physiology*, *16*(4), 433–448. https://doi.org/10.1111/j.1475-097X.1996.tb00731.x

Cho, Y. T., Fromm, S., Guyer, A. E., Detloff, A., Pine, D. S., Fudge, J. L., & Ernst, M. (2013). Nucleus accumbens, thalamus and insula connectivity during incentive anticipation in typical adults and adolescents. *NeuroImage*, *66*, 508–521. https://doi.org/10.1016/j.neuroimage.2012.10.013

Choe, J. C., Choe, J. C., & Crespi, B. J. (1997). *The evolution of social behaviour in insects and arachnids*. Cambridge University Press.

Choi, J.-K., & Bowles, S. (2007). The Coevolution of Parochial Altruism and War. *Science*, *318*(5850), 636–640. https://doi.org/10.1126/science.1144237

Chowdhury, S. M., Jeon, J. Y., & Saha, B. (2017). Gender Differences in the Giving and Taking Variants of the Dictator Game. *Southern Economic Journal*, *84*(2), 474–483. https://doi.org/10.1002/soej.12223

Civai, C., Crescentini, C., Rustichini, A., & Rumiati, R. I. (2012). Equality versus self-interest in the brain: Differential roles of anterior insula and medial prefrontal cortex. *NeuroImage*, *62*(1), 102–112. https://doi.org/10.1016/j.neuroimage.2012.04.037

Cloitre, M., Garvert, D. W., Brewin, C. R., Bryant, R. A., & Maercker, A. (2013). Evidence for proposed ICD-11 PTSD and complex PTSD: A latent profile analysis. *European journal of psychotraumatology*, *4*(1), 20706.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Contreras-Huerta, L. S., Lockwood, P. L., Bird, G., Apps, M. A., & Crockett, M. J. (2020). Prosocial behavior is associated with transdiagnostic markers of affective sensitivity in multiple domains. *Emotion*. http://dx.doi.org/10.1037/emo0000813

Cookson, R. (2000). Framing Effects in Public Goods Experiments. *Experimental Economics*, *3*(1), 55–79. https://doi.org/10.1023/A:1009994008166

Corradi-Dell’Acqua, C., Civai, C., Rumiati, R. I., & Fink, G. R. (2013). Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: An fMRI study. *Social Cognitive and Affective Neuroscience*, *8*(4), 424–431. https://doi.org/10.1093/scan/nss014

Cota, D., Steiner, M.-A., Marsicano, G., Cervino, C., Herman, J. P., Grübler, Y., Stalla, J., Pasquali, R., Lutz, B., Stalla, G. K., & Pagotto, U. (2007). Requirement of Cannabinoid Receptor Type 1 for the Basal Modulation of Hypothalamic-Pituitary-Adrenal Axis Function. *Endocrinology*, *148*(4), 1574–1581. https://doi.org/10.1210/en.2005-1649

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences*, *111*(48), 17320–17325. https://doi.org/10.1073/pnas.1408988111

Cubitt, R. P., Drouvelis, M., & Gächter, S. (2011). Framing and free riding: Emotional responses and punishment in social dilemma games. *Experimental Economics*, *14*(2), 254–272. https://doi.org/10.1007/s10683-010-9266-0

Cutler, J., & Campbell-Meiklejohn, D. (2019). A comparative fMRI meta-analysis of altruistic and strategic decisions to give. *Neuroimage*, *184*, 227–241.

Dandolo, L. C., & Schwabe, L. (2016). Stress-induced cortisol hampers memory generalization. *Learning & Memory*, *23*(12), 679–683. https://doi.org/10.1101/lm.042929.116

Darwin, C. (1859). *On the origin of species*. W. Clowes and Sons and Charing Cross, London.

Dawkins, R. (1982). *The extended phenotype* (Bd. 8). Oxford University Press Oxford.

De Dreu, C. K. W., Dussel, D. B., & Velden, F. S. T. (2015). In intergroup conflict, self-sacrifice is stronger among pro-social individuals, and parochial altruism emerges especially among cognitively taxed individuals. *Frontiers in Psychology*, *6*, 572. https://doi.org/10.3389/fpsyg.2015.00572

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., Ten Velden, F. S., Van Dijk, E., & Feith, S. W. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, *328*(5984), 1408–1411. https://doi.org/10.1126/science.1189047

De Dreu, C. K. W., & McCusker, C. (1997). Gain–loss frames and cooperation in two-person social dilemmas: A transformational analysis. *Journal of Personality and Social Psychology*, *72*(5), 1093–1106. https://doi.org/10.1037/0022-3514.72.5.1093

de Kloet, E. R. (2014). From Receptor Balance to Rational Glucocorticoid Therapy. *Endocrinology*, *155*(8), 2754–2769. https://doi.org/10.1210/en.2014-1048

de Kloet, E. R., Karst, H., & Joëls, M. (2008). Corticosteroid hormones in the central stress response: Quick-and-slow. *Frontiers in Neuroendocrinology*, *29*(2), 268–272. https://doi.org/10.1016/j.yfrne.2007.10.002

Decety, J., & Cowell, J. M. (2018). Interpersonal harm aversion as a necessary foundation for morality: A developmental neuroscience perspective. *Development and Psychopathology*, *30*(1), 153–164. https://doi.org/10.1017/S0954579417000530

Delaney, L., Fink, G., & Harmon, C. P. (2014). *Effects of Stress on Economic Decision-Making: Evidence from Laboratory Experiments*. Social Science Research Network. https://papers.ssrn.com/abstract=2420705

Demoulin, S., Cortes, B. P., Viki, T. G., Rodriguez, A. P., Rodriguez, R. T., Paladino, M. P., & Leyens, J. (2009). The role of in-group identification in infra-humanization. *International Journal of Psychology*, *44*(1), 4–11.

Dion, K. L. (1973). Cohesiveness as a determinant of ingroup-outgroup bias. *Journal of Personality and Social Psychology*, *28*(2), 163–171. https://doi.org/10.1037/h0035725

Domes, G., & Zimmer, P. (2019). Acute stress enhances the sensitivity for facial emotions: A signal detection approach. *Stress*, *22*(4), 455–460.

Dreber, A., Ellingsen, T., Johannesson, M., & Rand, D. (2013). Do people care about social context? Framing effects in dictator games. *Experimental Economics*, *16*(3), 349–371.

Droutman, V., Bechara, A., & Read, S. J. (2015). Roles of the Different Sub-Regions of the Insular Cortex in Various Phases of the Decision-Making Process. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00309

Dublin, H. T. (1983). Cooperation and reproductive competition among female African elephants. *Social behavior of female vertebrates*, 291–313.

Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, *73*(2), 459–478. https://doi.org/10.1016/j.geb.2011.02.003

Durrheim, K., Quayle, M., Tredoux, C. G., Titlestad, K., & Tooke, L. (2016). Investigating the evolution of ingroup favoritism using a minimal group interaction paradigm: The effects of inter-and intragroup interdependence. *PloS one*, *11*(11), e0165974.

Eckardt, W., & Zuberbühler, K. (2004). Cooperation and competition in two forest monkeys. *Behavioral Ecology*, *15*(3), 400–411. https://doi.org/10.1093/beheco/arh032

Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2013). Gender differences in social framing effects. *Economics Letters*, *118*(3), 470–472. https://doi.org/10.1016/j.econlet.2012.12.010

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7

Evans, A. M., & van Beest, I. (2017). Gain-loss framing effects in dilemmas of trust and reciprocity. *Journal of Experimental Social Psychology*, *73*, 151–163. https://doi.org/10.1016/j.jesp.2017.06.012

Everett, J. A., Faber, N. S., & Crockett, M. J. (2015). The influence of social preferences and reputational concerns on intergroup prosocial behaviour in gains and losses contexts. *Royal Society Open Science*, *2*(12), 150546.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human nature*, *13*(1), 1–25.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140.

FeldmanHall, O., Raio, C. M., Kubota, J. T., Seiler, M. G., & Phelps, E. A. (2015). The effects of social context and acute stress on decision making under uncertainty. *Psychological science*, *26*(12), 1918–1926.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press. https://www.biodiversitylibrary.org/bibliography/27468

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7

Gardner, A. (2015). The genetical theory of multilevel selection. *Journal of Evolutionary Biology*, *28*(2), 305–319. https://doi.org/10.1111/jeb.12566

Gardner, A., West, S. A., & Wild, G. (2011). The genetical theory of kin selection. *Journal of evolutionary biology*, *24*(5), 1020–1043.

Geen, R. G., & O’Neal, E. C. (1969). Activation of cue-elicited aggression by general arousal. *Journal of Personality and Social Psychology*, *11*(3), 289–292. https://doi.org/10.1037/h0026885

Geniole, S. N., Bird, B. M., Ruddick, E. L., & Carré, J. M. (2017). Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis. *Hormones and Behavior*, *92*, 37–50. https://doi.org/10.1016/j.yhbeh.2016.10.002

Gerlach, P., & Jaeger, B. (2016). *Another frame, another game? Explaining framing effects in economic games*. SocArXiv. https://doi.org/10.31235/osf.io/yf36n

Gillespie, S. M., Mitchell, I. J., Johnson, I., Dawson, E., & Beech, A. R. (2013). Exaggerated intergroup bias in economical decision making games: Differential effects of primary and secondary psychopathic traits. *PloS one*, *8*(8), e69565.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of theoretical biology*, *206*(2), 169–179.

Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of theoretical biology*, *213*(1), 103–119.

Gintis, H., Van Schaik, C., Boehm, C., Chapais, B., Flack, J. C., Pagel, M., Pruetz, J. D., Spikins, P., Whiten, A., & Erdal, D. (2015). Zoon politikon: The evolutionary origins of human political systems. *Current Anthropology*, *56*(3), 340–341.

Glowacki, L., Wilson, M. L., & Wrangham, R. W. (2017). The evolutionary anthropology of war. *Journal of Economic Behavior & Organization*. https://doi.org/10.1016/j.jebo.2017.09.014

Goerg, S. J., Rand, D. G., & Walkowitz, G. (2019). Framing effects in the prisoner’s dilemma but not in the dictator game. *Journal of the Economic Science Association*, 1–12.

Goldstein, D. S. (2010). Adrenal Responses to Stress. *Cellular and Molecular Neurobiology*, *30*(8), 1433–1440. https://doi.org/10.1007/s10571-010-9606-9

Gonzalez-Liencres, C., Breidenstein, A., Wolf, O. T., & Brüne, M. (2016). Sex-dependent effects of stress on brain correlates to empathy for pain. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *105*, 47–56. https://doi.org/10.1016/j.ijpsycho.2016.04.011

Goodall, J. (2010). *Through a Window: My Thirty Years with the Chimpanzees of Gombe*. HMH.

Goodman, W. K., Janson, J., & Wolf, J. M. (2017). Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. *Psychoneuroendocrinology*, *80*, 26–35. https://doi.org/10.1016/j.psyneuen.2017.02.030

Güth, W., Ploner, M., & Regner, T. (2009). Determinants of in-group bias: Is group affiliation mediated by guilt-aversion? *Journal of Economic Psychology*, *30*(5), 814–827.

Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, *52*, 15–23.

Hagger, M. S., Chatzisarantis, N. L., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., & Bruyneel, S. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*(4), 546–573.

Halevy, N., Bornstein, G., & Sagiv, L. (2008). “In-Group Love” and “Out-Group Hate” as Motives for Individual Participation in Intergroup Conflict: A New Game Paradigm. *Psychological Science*, *19*(4), 405–411. https://doi.org/10.1111/j.1467-9280.2008.02100.x

Halevy, N., Chou, E. Y., Cohen, T. R., & Bornstein, G. (2010). Relative deprivation and intergroup competition. *Group Processes & Intergroup Relations*, *13*(6), 685–700. https://doi.org/10.1177/1368430210371639

Hall, J. E. (2016). Rhythmical Excitation of the Heart. In *Guyton and Hall Textbook of Medical Physiology* (S. 123–129). Elsevier Health Sciences.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of theoretical biology*, *7*(1), 17–52.

Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature*, *228*(5277), 1218–1220.

Hartmann, P., Eisend, M., Apaolaza, V., & D'Souza, C. (2017). Warm glow vs. altruistic values: How important is intrinsic emotional reward in proenvironmental behavior? *Journal of Environmental Psychology*, *52*, 43–55. https://doi.org/10.1016/j.jenvp.2017.05.006

Haushofer, J., Cornelisse, S., Seinstra, M., Fehr, E., Joëls, M., & Kalenscher, T. (2013). No Effects of Psychosocial Stress on Intertemporal Choice. *Plos One*, *8*(11), e78597. https://doi.org/10.1371/journal.pone.0078597

Haushofer, J., Jain, P., Musau, A., & Ndetei, D. (2021). Stress may increase choice of sooner outcomes, but not temporal discounting. *Journal of Economic Behavior & Organization*, *183*, 377–396. https://doi.org/10.1016/j.jebo.2020.12.024

Haushofer, J., Jang, C., & Lynham, J. (2018). *Stress and Temporal Discounting: Do Domains Matter? Working paper*. *Retrieved in on June 23rd from: https://haushofer.ne.su.se/*.

Hein, G., Engelmann, J. B., & Tobler, P. N. (2018). Pain relief provided by an outgroup member enhances analgesia. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1887), 20180501. https://doi.org/10.1098/rspb.2018.0501

Hellhammer, J., & Schubert, M. (2012). The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. *Psychoneuroendocrinology*, *37*(1), 119–124. https://doi.org/10.1016/j.psyneuen.2011.05.012

Henckens, M. J. A. G., van Wingen, G. A., Joëls, M., & Fernández, G. (2010). Time-Dependent Effects of Corticosteroids on Human Amygdala Processing. *Journal of Neuroscience*, *30*(38), 12725–12732.

Hermans, E. J., Henckens, M. J. A. G., Joëls, M., & Fernández, G. (2014). Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences*, *37*(6), 304–314. https://doi.org/10.1016/j.tins.2014.03.006

Hermans, E. J., Marle, H. J. F. van, Ossewaarde, L., Henckens, M. J. A. G., Qin, S., Kesteren, M. T. R. van, Schoots, V. C., Cousijn, H., Rijpkema, M., Oostenveld, R., & Fernández, G. (2011). Stress-Related Noradrenergic Activity Prompts Large-Scale Neural Network Reconfiguration. *Science*, *334*(6059), 1151–1153. https://doi.org/10.1126/science.1209603

Hernandez Lallement, J., Kuss, K., Trautner, P., Weber, B., Falk, A., & Fliessbach, K. (2014). Effort increases sensitivity to reward and loss magnitude in the human brain. *Social Cognitive and Affective Neuroscience*, *9*(3), 342–349. https://doi.org/10.1093/scan/nss147

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and Self-Control From a Dual-Systems Perspective. *Perspectives on Psychological Science*, *4*(2), 162–176. https://doi.org/10.1111/j.1745-6924.2009.01116.x

Hovland, C. I., & Sears, R. R. (1940). Minor studies of aggression: VI. Correlation of lynchings with economic indices. *The Journal of Psychology*, *9*(2), 301–310.

Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron*, *87*(2), 451–462. https://doi.org/10.1016/j.neuron.2015.06.031

Hutcherson, C. A., & Tusche, A. (2021). Evidence accumulation, not “self-control,” explains dorsolateral prefrontal activation during normative choice. *BioRxiv*, 2020.10.06.328476. https://doi.org/10.1101/2020.10.06.328476

Iwama, G. K. (1998). Stress in Fish. *Annals of the New York Academy of Sciences*, *851*, 304–310. https://doi.org/10.1111/j.1749-6632.1998.tb09005.x

Jameel, M. K., Joshi, A. R., Dawane, J., Padwal, M., Joshi, A., Pandit, V. A., & Melinkeri, R. (2014). Effect of Various Physical Stress Models on Serum Cortisol Level in Wistar Rats. *Journal of Clinical and Diagnostic Research*, *8*(3), 181–183. https://doi.org/10.7860/JCDR/2014/7210.4116

Janus, T., & Riera-Crichton, D. (2015). Economic shocks, civil war and ethnicity. *Journal of Development Economics*, *115*, 32–44.

Joëls, M., & Baram, T. Z. (2009). The neuro-symphony of stress. *Nature reviews neuroscience*, *10*(6), 459.

Joëls, M., Fernandez, G., & Roozendaal, B. (2011). Stress and emotional memory: A matter of timing. *Trends in Cognitive Sciences*, *15*(6), 280–288. https://doi.org/10.1016/j.tics.2011.04.004

Jones, B., & Rachlin, H. (2006). Social discounting. *Psychological Science*, *17*(4), 283–286. https://doi.org/10.1111/j.1467-9280.2006.01699.x

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, *530*(7591), 473–476.

Kalenscher, T. (2014). Attitude toward health insurance in developing countries from a decision-making perspective. *Journal of Neuroscience, Psychology, and Economics*, *7*(3), 174–193. https://doi.org/10.1037/npe0000024

Kalenscher, T. (2017). Social psychology: Love thy stranger as thyself. *Nature Human Behaviour*, *1*(5), 0108. https://doi.org/10.1038/s41562-017-0108

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kiers, E. T., & Denison, R. F. (2008). Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annual review of ecology, evolution, and systematics*, *39*, 215–236.

Kinner, V. L., Wolf, O. T., & Merz, C. J. (2016). Cortisol alters reward processing in the human brain. *Hormones and Behavior*, *84*, 75–83. https://doi.org/10.1016/j.yhbeh.2016.05.005

Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic Medicine*, *61*(2), 154–162. https://doi.org/10.1097/00006842-199903000-00006

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, *28*(1–2), 76–81. https://doi.org/10.1159/000119004

Klaw, E. L., Demers, A. L., & Da Silva, N. (2016). Predicting Risk Factors for Intimate Partner Violence Among Post-9/11 College Student Veterans. *Journal of Interpersonal Violence*, *31*(4), 572–597. https://doi.org/10.1177/0886260514556102

Kogler, L., Müller, V. I., Chang, A., Eickhoff, S. B., Fox, P. T., Gur, R. C., & Derntl, B. (2015). Psychosocial versus physiological stress—Meta-analyses on deactivations and activations of the neural correlates of stress reactions. *NeuroImage*, *119*, 235–251. https://doi.org/10.1016/j.neuroimage.2015.06.059

Kopiske, K. K., Bruno, N., Hesse, C., Schenk, T., & Franz, V. H. (2016). The functional subdivision of the visual brain: Is there a real illusion effect on action? A multi-lab replication study. *Cortex*, *79*, 130–152.

Kramer, J., & Meunier, J. (2016). Kin and multilevel selection in social evolution: A never-ending controversy? *F1000Research*, *5*(F1000 Faculty Rev-776).

Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280. https://doi.org/10.1177/2515245918771304

Kuczaj, S. A., Winship, K. A., & Eskelinen, H. C. (2015). Can bottlenose dolphins (Tursiops truncatus) cooperate when solving a novel task? *Animal cognition*, *18*(2), 543–550.

Kudielka, B. M., & Kirschbaum, C. (2005). Sex differences in HPA axis responses to stress: A review. *Biological Psychology*, *69*(1), 113–132. https://doi.org/10.1016/j.biopsycho.2004.11.009

Lahr, M. M., Rivera, F., Power, R. K., Mounier, A., Copsey, B., Crivellaro, F., Edung, J. E., Fernandez, J. M. M., Kiarie, C., Lawrence, J., Leakey, A., Mbua, E., Miller, H., Muigai, A., Mukhongo, D. M., Van Baelen, A., Wood, R., Schwenninger, J.-L., Grün, R., … Foley, R. A.

(2016). Inter-group violence among early Holocene hunter-gatherers of West Turkana, Kenya. *Nature*, *529*(7586), 394–398. https://doi.org/10.1038/nature16477

Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and effect in biology revisited: Is Mayr’s proximate-ultimate dichotomy still useful? *Science*, *334*(6062), 1512–1516.

Lehmann, L., & Keller, L. (2006). The evolution of cooperation and altruism – a general framework and a classification of models. *Journal of Evolutionary Biology*, *19*(5), 1365–1376. https://doi.org/10.1111/j.1420-9101.2006.01119.x

Lehmann, L., & Rousset, F. (2014). Fitness, inclusive fitness, and optimization. *Biology & Philosophy*, *29*(2), 181–195. https://doi.org/10.1007/s10539-013-9415-x

Lehmann, L., & Rousset, F. (2020). When do individuals maximize their inclusive fitness? *The American Naturalist*, *195*(4), 717–732.

Leibbrandt, A., & López-Pérez, R. (2012). An exploration of third and second party punishment in ten simple games. *Journal of Economic Behavior & Organization*, *84*(3), 753–766.

Leliveld, M. C., van Dijk, E., & van Beest, I. (2008). Initial Ownership in Bargaining: Introducing the Giving, Splitting, and Taking Ultimatum Bargaining Game. *Personality and Social Psychology Bulletin*, *34*(9), 1214–1225. https://doi.org/10.1177/0146167208318600

Leyens, J., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology*, *31*(4), 395–411.

Li, R., Smith, D. V., Clithero, J. A., Venkatraman, V., Carter, R. M., & Huettel, S. A. (2017). Reason’s Enemy Is Not Emotion: Engagement of Cognitive Control Networks Explains Biases in Gain/Loss Framing. *Journal of Neuroscience*, *37*(13), 3588–3598.

Li, Z., Yan, C., Xie, W.-Z., Li, K., Zeng, Y.-W., Jin, Z., Cheung, E. F. C., & Chan, R. C. K. (2015). Anticipatory pleasure predicts effective connectivity in the mesolimbic system. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00217

Ligon, J. D. (1983). Cooperation and reciprocity in avian social systems. *The American Naturalist*, *121*(3), 366–384.

List, J. A. (2007). On the Interpretation of Giving in Dictator Games. *Journal of Political Economy*, *115*(3), 482–493. https://doi.org/10.1086/519249

Liu, J., Gu, R., Liao, C., Lu, J., Fang, Y., Xu, P., Luo, Y., & Cui, F. (2020). The neural mechanism of the social framing effect: Evidence from fMRI and tDCS studies. *Journal of Neuroscience*, *40*(18), 3646–3656.

Loewenstein, G., O'Donoghue, T., & Bhatia, S. (2015). Modeling the interplay between affect and deliberation. *Decision*, *2*(2), 55–81. https://doi.org/10.1037/dec0000029

Lohse, J., & Waichman, I. (2020). The effects of contemporaneous peer punishment on cooperation with the future. *Nature communications*, *11*(1), 1–8.

Losecaat Vermeer, A. B., Riečanský, I., & Eisenegger, C. (2016). Competition, testosterone, and adult neurobehavioral plasticity. In B. Studer & S. Knecht (Hrsg.), *Progress in Brain Research* (S. 213–238). Elsevier. http://www.sciencedirect.com/science/article/pii/S0079612316300577

Maier, M. A., Buechner, V. L., Dechamps, M. C., Pflitsch, M., Kurzrock, W., Tressoldi, P., Rabeyron, T., Cardeña, E., Marcusson-Clavertz, D., & Martsinkovskaja, T. (2020). A preregistered multi-lab replication of Maier et al. (2014, Exp. 4) testing retroactive avoidance. *Plos One*, *15*(8), e0238373.

Margittai, Z., Nave, G., Van Wingerden, M., Schnitzler, A., Schwabe, L., & Kalenscher, T. (2018). Combined effects of glucocorticoid and noradrenergic activity on loss aversion. *Neuropsychopharmacology*, *43*(2), 334–341.

Margittai, Z., Strombach, T., van Wingerden, M., Joëls, M., Schwabe, L., & Kalenscher, T. (2015). A friend in need: Time-dependent effects of stress on social discounting in men. *Hormones and Behavior*, *73*, 75–82. https://doi.org/10.1016/j.yhbeh.2015.05.019

Margittai, Z., van Wingerden, M., Schnitzler, A., Joëls, M., & Kalenscher, T. (2018). Dissociable roles of glucocorticoid and noradrenergic activation on social discounting. *Psychoneuroendocrinology*, *90*, 22–28. https://doi.org/10.1016/j.psyneuen.2018.01.015

Marshall, J. A. (2015). *Social evolution and inclusive fitness theory: An introduction*. Princeton University Press.

Marshall, J. A. R. (2011). Group selection and kin selection: Formally equivalent approaches. *Trends in Ecology & Evolution*, *26*(7), 325–332. https://doi.org/10.1016/j.tree.2011.04.008

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mayr, E. (1961). Cause and effect in biology. *Science*, *134*(3489), 1501–1506.

McCarthy, R. J., Elson, M., Vazire, S., & Tullett, A. (2018). A conceptual review of lab-based aggression paradigms. *Collabra: Psychology*, *4*(1).

McCarty, R. (2016). The Fight-or-Flight Response: A Cornerstone of Stress Research. In *Stress: Concepts, Cognition, Emotion, and Behavior: Handbook of Stress Series, Edited by George Fink*. Academic Press.

McCorry, L. K. (2007). Physiology of the Autonomic Nervous System. *American Journal of Pharmaceutical Education*, *71*(4), 78.

McDonald, M. M., Navarrete, C. D., & Van Vugt, M. (2012). Evolution and the psychology of intergroup conflict: The male warrior hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 670–679. https://doi.org/10.1098/rstb.2011.0301

Mehta, P. H., & Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Hormones and Behavior*, *58*(5), 898–906. https://doi.org/10.1016/j.yhbeh.2010.08.020

Mehta, P. H., & Prasad, S. (2015). The dual-hormone hypothesis: A brief review and future research agenda. *Current Opinion in Behavioral Sciences*, *3*, 163–168. https://doi.org/10.1016/j.cobeha.2015.04.008

Mengel, F. (2018). Risk and Temptation: A Meta-study on Prisoner's Dilemma Games. *The Economic Journal*, *128*(616), 3182–3209.

Metz, S., Waiblinger-Grigull, T., Schulreich, S., Chae, W. R., Otte, C., Heekeren, H. R., & Wingenfeld, K. (2020). Effects of hydrocortisone and yohimbine on decision-making under risk. *Psychoneuroendocrinology*, *114*, 104589.

Miczek, K. A., DeBold, J. F., Gobrogge, K., Newman, E. L., & Almeida, R. M. M. (2017). The Role of Neurotransmitters in Violence and Aggression. In *The Wiley Handbook of Violence and Aggression* (S. 1–13). Wiley. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119057574.whbva019

Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A. J., Schweizer, S., Frith, C. D., & Dalgleish, T. (2009). A Key Role for Similarity in Vicarious Reward. *Science*, *324*(5929), 900–900. https://doi.org/10.1126/science.1170539

Montague, P. R., & Lohrenz, T. (2007). To Detect and Correct: Norm Violations and Their Enforcement. *Neuron*, *56*(1), 14–18. https://doi.org/10.1016/j.neuron.2007.09.020

Morelli, S. A., Sacchet, M. D., & Zaki, J. (2015). Common and distinct neural correlates of personal and vicarious reward: A quantitative meta-analysis. *NeuroImage*, *112*, 244–253. https://doi.org/10.1016/j.neuroimage.2014.12.056

Nagy, T., van Lien, R., Willemsen, G., Proctor, G., Efting, M., Fülöp, M., Bárdos, G., Veerman, E. C. I., & Bosch, J. A. (2015). A fluid response: Alpha-amylase reactions to acute laboratory stress are related to sample timing and saliva flow rate. *Biological Psychology*, *109*, 111–119. https://doi.org/10.1016/j.biopsycho.2015.04.012

Nalborczyk, L., Bürkner, P.-C., & Williams, D. R. (2018). *Pragmatism should not be a substitute for statistical literacy, a commentary on Albers, Kiers, and van Ravenzwaaij*. https://psyarxiv.com/fmtkj/

Nater, U. M., & Rohleder, N. (2009). Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: Current state of research. *Psychoneuroendocrinology*, *34*(4), 486–496. https://doi.org/10.1016/j.psyneuen.2009.01.014

Negd, M., Mallan, K. M., & Lipp, O. V. (2011). The role of anxiety and perspective-taking strategy on affective empathic responses. *Behaviour Research and Therapy*, *49*(12), 852–857. https://doi.org/10.1016/j.brat.2011.09.008

Nelson, R. J., & Trainor, B. C. (2007). Neural mechanisms of aggression. *Nature Reviews Neuroscience*, *8*(7), 536–546. https://doi.org/10.1038/nrn2174

Neufang, S., Akhrif, A., Herrmann, C. G., Drepper, C., Homola, G. A., Nowak, J., Waider, J., Schmitt, A. G., Lesch, K.-P., & Romanos, M. (2016). Serotonergic modulation of ‘waiting impulsivity’ is mediated by the impulsivity phenotype in humans. *Translational Psychiatry*, *6*(11), e940–e940. https://doi.org/10.1038/tp.2016.210

Nickels, N., Kubicki, K., & Maestripieri, D. (2017). Sex differences in the effects of psychosocial stress on cooperative and prosocial behavior: Evidence for ‘flight or fight’in males and ‘tend and befriend’in females. *Adaptive Human Behavior and Physiology*, *3*(2), 171–183.

Nowak, M. A. (2006). Five Rules for the Evolution of Cooperation. *Science*, *314*(5805), 1560–1563. https://doi.org/10.1126/science.1133755

O’Brien, T. G. (1993). Allogrooming behaviour among adult female wedge-capped capuchin monkeys. *Animal behaviour*, *46*(3), 499–510.

Okasha, S. (2016). The relation between kin and multilevel selection: An approach using causal graphs. *The British Journal for the Philosophy of Science*, *67*(2), 435–470.

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., & Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. *Human Brain Mapping*, *39*(8), 3398–3418. https://doi.org/10.1002/hbm.24184

Oster, G. F., & Wilson, E. O. (1978). *Caste and ecology in the social insects*. Princeton University Press.

Pabst, S., Brand, M., & Wolf, O. T. (2013a). Stress effects on framed decisions: There are differences for gains and losses. *Frontiers in behavioral neuroscience*, *7*, 142.

Pabst, S., Brand, M., & Wolf, O. T. (2013b). Stress and decision making: A few minutes make all the difference. *Behavioural Brain Research*, *250*, 39–45. https://doi.org/10.1016/j.bbr.2013.04.046

Picard, G., Tan, C. O., Zafonte, R., & Taylor, J. A. (2009). Incongruous changes in heart period and heart rate variability with vagotonic atropine: Implications for rehabilitation medicine. *PM & R: The Journal of Injury, Function, and Rehabilitation*, *1*(9), 820–826. https://doi.org/10.1016/j.pmrj.2009.07.017

Porcelli, A. J., & Delgado, M. R. (2009). Acute stress modulates risk taking in financial decision making. *Psychological Science*, *20*(3), 278–283.

Porcelli, A. J., Lewis, A. H., & Delgado, M. R. (2012). Acute Stress Influences Neural Circuits of Reward Processing. *Frontiers in Neuroscience*, *6*, 157. https://doi.org/10.3389/fnins.2012.00157

Potts, S. R., McCuddy, W. T., Jayan, D., & Porcelli, A. J. (2019). To trust, or not to trust? Individual differences in physiological reactivity predict trust under acute stress. *Psychoneuroendocrinology*, *100*, 75–84. https://doi.org/10.1016/j.psyneuen.2018.09.019

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*, *28*(7), 916–931. https://doi.org/10.1016/S0306-4530(02)00108-7

Putman, P., Antypa, N., Crysovergi, P., & van der Does, W. A. (2010). Exogenous cortisol acutely influences motivated decision making in healthy young men. *Psychopharmacology*, *208*(2), 257–263.

Quaedflieg, C. W. E. M., Meyer, T., van Ruitenbeek, P., & Smeets, T. (2017). Examining habituation and sensitization across repetitive laboratory stress inductions using the MAST. *Psychoneuroendocrinology*, *77*, 175–181. https://doi.org/10.1016/j.psyneuen.2016.12.009

Quaedflieg, C. W. E. M., & Schwabe, L. (2018). Memory dynamics under stress. *Memory*, *26*(3), 364–376. https://doi.org/10.1080/09658211.2017.1338299

Quaedflieg, C. W. E. M., Schwabe, L., Meyer, T., & Smeets, T. (2013). Time dependent effects of stress prior to encoding on event-related potentials and 24h delayed retrieval. *Psychoneuroendocrinology*, *38*(12), 3057–3069. https://doi.org/10.1016/j.psyneuen.2013.09.002

Rabbie, J. M., Schot, J. C., & Visser, L. (1989). Social identity theory: A conceptual and empirical critique from the perspective of a behavioural interaction model. *European Journal of Social Psychology*, *19*(3), 171–202. https://doi.org/10.1002/ejsp.2420190302

Raio, C. M., Hartley, C. A., Orederu, T. A., Li, J., & Phelps, E. A. (2017). Stress attenuates the flexible updating of aversive value. *Proceedings of the National Academy of Sciences*, *114*(42), 11241–11246.

Rand, D. G., Brescoll, V. L., Everett, J. A. C., Capraro, V., & Barcelo, H. (2016). Social heuristics and social roles: Intuition favors altruism for women but not for men. *Journal of Experimental Psychology. General*, *145*(4), 389–396. https://doi.org/10.1037/xge0000154

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in cognitive sciences*, *17*(8), 413–425.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677. https://doi.org/10.1038/ncomms4677

Richerson, G. B., Aston-Jones, G., & Saper, C. B. (2012). The Modulatory Functions of the Brain Stem. In *Principles of Neural Science* (Bd. 5, S. 1038–1041). McGraw-Hill.

Riis-Vestergaard, M. I., van Ast, V., Cornelisse, S., Joëls, M., & Haushofer, J. (2018). The effect of hydrocortisone administration on intertemporal choice. *Psychoneuroendocrinology*, *88*, 173–182. https://doi.org/10.1016/j.psyneuen.2017.10.002

Rilling, J. K., & Sanfey, A. G. (2011). The Neuroscience of Social Decision-Making. *Annual Review of Psychology*, *62*(1), 23–48. https://doi.org/10.1146/annurev.psych.121208.131647

Roelofs, K., Bakvis, P., Hermans, E. J., van Pelt, J., & van Honk, J. (2007). The effects of social stress and cortisol responses on the preconscious selective attention to social threat. *Biological Psychology*, *75*(1), 1–7. https://doi.org/10.1016/j.biopsycho.2006.09.002

Roozendaal, B., Okuda, S., Van der Zee, E. A., & McGaugh, J. L. (2006). Glucocorticoid enhancement of memory requires arousal-induced noradrenergic activation in the basolateral amygdala. *Proceedings of the National Academy of Sciences*, *103*(17), 6741–6746.

Rosenbaum, D., Hilsendegen, P., Thomas, M., Haeussinger, F. B., Metzger, F. G., Nuerk, H.-C., Fallgatter, A. J., Nieratschker, V., & Ehlis, A.-C. (2018). Cortical hemodynamic changes during the Trier Social Stress Test: An fNIRS study. *NeuroImage*, *171*, 107–115. https://doi.org/10.1016/j.neuroimage.2017.12.061

Samanez-Larkin, G. R., Hollon, N. G., Carstensen, L. L., & Knutson, B. (2008). Individual Differences in Insular Sensitivity During Loss Anticipation Predict Avoidance Learning.

Psychological Science, *19*(4), 320–323. https://doi.org/10.1111/j.1467-9280.2008.02087.x

Samuni, L., Mielke, A., Preis, A., Crockford, C., & Wittig, R. (2020). Intergroup competition enhances chimpanzee (Pan troglodytes verus) in-group cohesion. *International Journal of Primatology*, *41*, 342–362. https://doi.org/10.1007/s10764-019-00112-y

Sandi, C., & Haller, J. (2015). Stress and the social brain: Behavioural effects and neurobiological mechanisms. *Nature Reviews Neuroscience*, *16*(5), 290–304. https://doi.org/10.1038/nrn3918

Saraiya, T. C., Fitzpatrick, S., Zumberg-Smith, K., López-Castro, T., E. Back, S., & A. Hien, D. (2021). Social–Emotional Profiles of PTSD, Complex PTSD, and Borderline Personality Disorder Among Racially and Ethnically Diverse Young Adults: A Latent Class Analysis. *Journal of Traumatic Stress*, *34*(1), 56–68.

Schaal, N. K., Hepp, P., Schweda, A., Wolf, O. T., & Krampe, C. (2019). A Functional Near-Infrared Spectroscopy Study on the Cortical Haemodynamic Responses During the Maastricht Acute Stress Test. *Scientific Reports*, *9*(1), 13459. https://doi.org/10.1038/s41598-019-49826-2

Schneeberger, K., Dietz, M., & Taborsky, M. (2012). Reciprocal cooperation between unrelated rats depends on cost to donor and benefit to recipient. *BMC evolutionary biology*, *12*(1), 1–7.

Schweda, A., Faber, N. S., Crockett, M. J., & Kalenscher, T. (2019). The effects of psychosocial stress on intergroup resource allocation. *Scientific reports*, *9*(1), 1–12.

Schweda, A., Margittai, Z., & Kalenscher, T. (2020). Acute stress counteracts framing-induced generosity boosts in social discounting in young healthy men. *Psychoneuroendocrinology*, *121*, 104860. https://doi.org/10.1016/j.psyneuen.2020.104860

Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science*, *6*(1), 38–47.

Sellitto, M., Ciaramelli, E., Mattioli, F., & di Pellegrino, G. (2016). Reduced Sensitivity to Sooner Reward During Intertemporal Decision-Making Following Insula Damage in Humans. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00367

Sellitto, M., Neufang, S., Schweda, A., Weber, B., & Kalenscher, T. (2021). Arbitration between insula and temporoparietal junction subserves framing-induced boosts in generosity during social discounting. *NeuroImage*, *238*, 118211. https://doi.org/10.1016/j.neuroimage.2021.118211

Sellitto, M., Schweda, A., Weber, B., & Kalenscher, T. (2019). To give or not to take – insular cortex subserves framing-induced boosts in generosity during social discounting. *BioRxiv*, 841338. https://doi.org/10.1101/841338

Selye, H. (1950). Stress and the General Adaptation Syndrome. *British Medical Journal*, *1*(4667), 1383–1392.

Shadmanfaat, S. M. S., Kabiri, S., Solensten, B., Willits, D. W., & Cochran, J. (2021). A Longitudinal Study of Fan Aggression: A Test of General Strain Theory. *Deviant Behavior*, 1–13. https://doi.org/10.1080/01639625.2021.1877095

Sherif, M., Harvey, O., Hood, W., & Sherif, C. (1961). Intergroup conflict and group relations: The robbers cave experiment. *The Institute of Group Relations, Norman, OK*.

Shilton, A. L., Laycock, R., & Crewther, S. G. (2017). The Maastricht Acute Stress Test (MAST): Physiological and Subjective Responses in Anticipation, and Post-stress. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00567

Sidanius, J., & Pratto, F. (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

Silver, E., & Teasdale, B. (2005). Mental Disorder and Violence: An Examination of Stressful Life Events and Impaired Social Support. *Social Problems*, *52*(1), 62–78. https://doi.org/10.1525/sp.2005.52.1.62

Simmons, P. S., Miles, J. M., Gerich, J. E., & Haymond, M. W. (1984). Increased proteolysis. An effect of increases in plasma cortisol within the physiologic range. *The Journal of Clinical Investigation*, *73*(2), 412–420. https://doi.org/10.1172/JCI111227

Singer, N., Sommer, M., Döhnel, K., Zänkert, S., Wüst, S., & Kudielka, B. M. (2017). Acute psychosocial stress and everyday moral decision-making in young healthy men: The impact of cortisol. *Hormones and Behavior*, *93*, 72–81. https://doi.org/10.1016/j.yhbeh.2017.05.002

Sip, K. E., Smith, D. V., Porcelli, A. J., Kar, K., & Delgado, M. R. (2015). Social closeness and feedback modulate susceptibility to the framing effect. *Social Neuroscience*, *10*(1), 35–45. https://doi.org/10.1080/17470919.2014.944316

Smeets, T., Cornelisse, S., Quaedflieg, C. W. E. M., Meyer, T., Jelicic, M., & Merckelbach, H. (2012). Introducing the Maastricht Acute Stress Test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. *Psychoneuroendocrinology*, *37*(12), 1998–2008. https://doi.org/10.1016/j.psyneuen.2012.04.012

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children’s moral and social judgments. In *Handbook of moral development* (S. 119–153). Lawrence Erlbaum Associates Publishers.

Smith, A. (2011). Identifying in-group and out-group effects in the trust game. *The BE Journal of Economic Analysis & Policy*, *11*(1).

Smith, D., Sip, K. E., & Delgado, M. R. (2015). Functional connectivity with distinct neural networks tracks fluctuations in gain/loss framing susceptibility. *Human brain mapping*, *36*(7), 2743–2755.

Smith, J. M. (1964). Group selection and kin selection. *Nature*, *201*(4924), 1145–1147.

Sokol-Hessner, P., Raio, C. M., Gottesman, S. P., Lackovic, S. F., & Phelps, E. A. (2016). Acute stress does not affect risky monetary decision-making. *Neurobiology of stress*, *5*, 19–25.

Soldin, O., & Mattison, D. (2009). Sex Differences in Pharmacokinetics and Pharmacodynamics. *Clinical pharmacokinetics*, *48*(3), 143–157. https://doi.org/10.2165/00003088-200948030-00001

Sollberger, S., Bernauer, T., & Ehlert, U. (2016). Stress influences environmental donation behavior in men. *Psychoneuroendocrinology*, *63*, 311–319. https://doi.org/10.1016/j.psyneuen.2015.10.017

Sonnemans, J., Schram, A., & Offerman, T. (1998). Public good provision and public bad prevention: The effect of framing. *Journal of Economic Behavior & Organization*, *34*(1), 143–161. https://doi.org/10.1016/S0167-2681(97)00042-5

Soutschek, A., Burke, C. J., Beharelle, A. R., Schreiber, R., Weber, S. C., Karipidis, I. I., Velden, J. ten, Weber, B., Haker, H., Kalenscher, T., & Tobler, P. N. (2017). The dopaminergic reward system underpins gender differences in social preferences. *Nature Human Behaviour*, *1*(11), 819. https://doi.org/10.1038/s41562-017-0226-y

Soutschek, A., Ruff, C. C., Strombach, T., Kalenscher, T., & Tobler, P. N. (2016). Brain stimulation reveals crucial role of overcoming self-centeredness in self-control. *Science Advances*, *2*(10), e1600992. https://doi.org/10.1126/sciadv.1600992

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The Neural Signature of Social Norm Compliance. *Neuron*, *56*(1), 185–196. https://doi.org/10.1016/j.neuron.2007.09.011

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of experimental social psychology*, *27*(1), 76–105.

Stacey, P. B., & Koenig, W. D. (1990). *Cooperative breeding in birds: Long term studies of ecology and behaviour*. Cambridge University Press.

Starcke, K., & Brand, M. (2016). Effects of stress on decisions under uncertainty: A meta-analysis. *Psychological bulletin*, *142*(9), 909.

Starcke, K., Polzer, C., Wolf, O. T., & Brand, M. (2011). Does stress alter everyday moral decision-making? *Psychoneuroendocrinology*, *36*(2), 210–219. https://doi.org/10.1016/j.psyneuen.2010.07.010

Steinbeis, N., Engert, V., Linz, R., & Singer, T. (2015). The effects of stress and affiliation on social decision-making: Investigating the tend-and-befriend pattern. *Psychoneuroendocrinology*, *62*, 138–148. https://doi.org/10.1016/j.psyneuen.2015.08.003

Stephens, M. A. C., Mahon, P. B., McCaul, M. E., & Wand, G. S. (2016). Hypothalamic–pituitary–adrenal axis response to acute psychosocial stress: Effects of biological sex and circulating sex hormones. *Psychoneuroendocrinology*, *66*, 47–55. https://doi.org/10.1016/j.psyneuen.2015.12.021

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in cognitive sciences*, *8*(2), 60–65.

Strang, S., Gerhardt, H., Marsh, N., Artigas, S. O., Hu, Y., Hurlemann, R., & Park, S. Q. (2017). A matter of distance—The effect of oxytocin on social discounting is empathy-dependent. *Psychoneuroendocrinology*, *78*, 229–232.

Strombach, T., Jin, J., Weber, B., Kenning, P., Shen, Q., Ma, Q., & Kalenscher, T. (2014). Charity Begins at Home: Cultural Differences in Social Discounting and Generosity. *Journal of Behavioral Decision Making*, *27*(3), 235–245. https://doi.org/10.1002/bdm.1802

Strombach, T., Margittai, Z., Gorczyca, B., & Kalenscher, T. (2016). Gender-specific effects of cognitive load on social discounting. *PloS one*, *11*(10), e0165289.

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I. I., Tobler, P. N., & Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proceedings of the National Academy of Sciences*, *112*(5), 1619–1624. https://doi.org/10.1073/pnas.1414715112

Sudheimer, K. D., O'Hara, R., Spiegel, D., Powers, B., Kraemer, H. C., Neri, E., Weiner, M., Hardan, A., Hallmayer, J., & Dhabhar, F. S. (2014). Cortisol, cytokines, and hippocampal volume interactions in the elderly. *Frontiers in Aging Neuroscience*, *6*, 153. https://doi.org/10.3389/fnagi.2014.00153

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European journal of social psychology*, *1*(2), 149–178.

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader*, *56*(65), 9780203505984–16.

Takahashi, T., Ikeda, K., & Hasegawa, T. (2007). Social evaluation-induced amylase elevation and economic decision-making in the dictator game in humans. *Neuroendocrinology Letters*, *28*(5), 662–665.

Taylor, S. E. (2006). Tend and Befriend: Biobehavioral Bases of Affiliation Under Stress. *Current Directions in Psychological Science*, *15*(6), 273–277. https://doi.org/10.1111/j.1467-8721.2006.00451.x

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review*, *107*(3), 411–429. https://doi.org/10.1037/0033-295X.107.3.411

Ter Horst, J. P., de Kloet, E. R., Schächinger, H., & Oitzl, M. S. (2012). Relevance of stress and female sex hormones for emotion and cognition. *Cellular and Molecular Neurobiology*, *32*(5), 725–735. https://doi.org/10.1007/s10571-011-9774-2

Terbeck, S., Kahane, G., McTavish, S., McCutcheon, R., Hewstone, M., Savulescu, J., Chesterman, L. P., Cowen, P. J., & Norbury, R. (2015). β-Adrenoceptor blockade modulates fusiform gyrus activity to black versus white faces. *Psychopharmacology*, *232*(16), 2951–2958. https://doi.org/10.1007/s00213-015-3929-7

Terbeck, S., Kahane, G., McTavish, S., Savulescu, J., Cowen, P. J., & Hewstone, M. (2012). Propranolol reduces implicit negative racial bias. *Psychopharmacology*, *222*(3), 419–424. https://doi.org/10.1007/s00213-012-2657-5

Terbeck, S., Savulescu, J., Chesterman, L. P., & Cowen, P. J. (2016). Noradrenaline effects on social behaviour, intergroup relations, and moral decisions. *Neuroscience & Biobehavioral Reviews*, *66*, 54–60. https://doi.org/10.1016/j.neubiorev.2016.03.031

The JASP Team. (2018). *JASP* (0.9.1.0) [Computer software].

Thielmann, I., & Böhm, R. (2016). Who Does (Not) Participate in Intergroup Conflict? *Social Psychological and Personality Science*, *7*(8), 778–787. https://doi.org/10.1177/1948550616660160

Thomae, M., Wood, M., & Immelman, N. (2015). Replication of N Halevy, G Bornstein, L Sagiv (2008) as part of „Estimating the reproducibility of psychological science“ by the open science collaboration. *Science*, *349*(6251), aac4716.

Timmermans, S., Souffriau, J., & Libert, C. (2019). A General Introduction to Glucocorticoid Biology. *Frontiers in Immunology*, *10*. https://doi.org/10.3389/fimmu.2019.01545

Timmons, A. C., Arbel, R., & Margolin, G. (2017). Daily patterns of stress and conflict in couples: Associations with marital aggression and family-of-origin aggression. *Journal of family psychology*, *31*(1), 93.

Tokarz, R. R., & Summers, C. H. (2011). Stress and Reproduction in Reptiles. In D. O. Norris & K. H. Lopez (Hrsg.), *Hormones and Reproduction of Vertebrates* (S. 169–213). Academic Press. https://doi.org/10.1016/B978-0-12-374930-7.10007-X

Tomasino, B., Lotto, L., Sarlo, M., Civai, C., Rumiati, R., & Rumiati, R. I. (2013). Framing the ultimatum game: The contribution of simulation. *Frontiers in human neuroscience*, *7*, 337.

Tomova, L., Majdandžić, J., Hummer, A., Windischberger, C., Heinrichs, M., & Lamm, C. (2017). Increased neural responses to empathy for pain might explain how acute stress increases prosociality. *Social Cognitive and Affective Neuroscience*, *12*(3), 401–408. https://doi.org/10.1093/scan/nsw146

Tomova, L., von Dawans, B., Heinrichs, M., Silani, G., & Lamm, C. (2014). Is stress affecting our ability to tune into others? Evidence for gender differences in the effects of stress on self-other distinction. *Psychoneuroendocrinology*, *43*, 95–104. https://doi.org/10.1016/j.psyneuen.2014.02.006

Trapp, S., O'Doherty, J. P., & Schwabe, L. (2018). Stressful Events as Teaching Signals for the Brain. *Trends in Cognitive Sciences*, *22*(6), 475–478. https://doi.org/10.1016/j.tics.2018.03.007

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology*, *46*(1), 35–57.

Tsumura, H., & Shimada, H. (2012). Acutely Elevated Cortisol in Response to Stressor Is Associated with Attentional Bias Toward Depression-Related Stimuli but Is Not Associated with Attentional Function. *Applied Psychophysiology and Biofeedback*, *37*(1), 19–29. https://doi.org/10.1007/s10484-011-9172-z

Turner, J. C., Brown, R. J., & Tajfel, H. (1979). Social comparison and group interest in ingroup favouritism. *European journal of social psychology*, *9*(2), 187–204.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory.* Basil Blackwell.

Tversky, A., & Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

van Marle, H. J. F., Hermans, E. J., Qin, S., & Fernández, G. (2009). From Specificity to Sensitivity: How Acute Stress Affects Amygdala Processing of Biologically Salient Stimuli. *Biological Psychiatry*, *66*(7), 649–655. https://doi.org/10.1016/j.biopsych.2009.05.014

van Marle, H. J. F., Hermans, E. J., Qin, S., & Fernández, G. (2010). Enhanced resting-state connectivity of amygdala in the immediate aftermath of acute psychological stress. *NeuroImage*, *53*(1), 348–354. https://doi.org/10.1016/j.neuroimage.2010.05.070

Van Vugt, M., De Cremer, D., & Janssen, D. P. (2007). Gender differences in cooperation and competition: The male-warrior hypothesis. *Psychological Science*, *18*(1), 19–23. https://doi.org/10.1111/j.1467-9280.2007.01842.x

Vekaria, K. M., Brethel-Haurwitz, K. M., Cardinale, E. M., Stoycos, S. A., & Marsh, A. A. (2017). Social discounting and distance perceptions in costly altruism. *Nature Human Behaviour*, *1*(5), 0100. https://doi.org/10.1038/s41562-017-0100

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327.

Vinkers, C. H., Zorn, J. V., Cornelisse, S., Koot, S., Houtepen, L. C., Olivier, B., Verster, J. C., Kahn, R. S., Boks, M. P. M., Kalenscher, T., & Joëls, M. (2013). Time-dependent changes in altruistic punishment following stress. *Psychoneuroendocrinology*, *38*(9), 1467–1475. https://doi.org/10.1016/j.psyneuen.2012.12.012

Virgin Jr, C. E., & Sapolsky, R. M. (1997). Styles of male social behavior and their endocrine correlates among low-ranking baboons. *American Journal of Primatology*, *42*(1), 25–39.

Von Dawans, B., Ditzen, B., Trueg, A., Fischbacher, U., & Heinrichs, M. (2019). Effects of acute stress on social behavior in women. *Psychoneuroendocrinology*, *99*, 137–144. https://doi.org/10.1016/j.psyneuen.2018.08.031

Von Dawans, B., Fischbacher, U., Kirschbaum, C., Fehr, E., & Heinrichs, M. (2012). The social dimension of stress reactivity: Acute stress increases prosocial behavior in humans. *Psychological science*, *23*(6), 651–660.

Von Dawans, B., Kirschbaum, C., & Heinrichs, M. (2011). The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology*, *36*(4), 514–522. https://doi.org/10.1016/j.psyneuen.2010.08.004

von Dawans, B., Spenthof, I., Zimmer, P., & Domes, G. (2020). Acute Psychosocial Stress Modulates the Detection Sensitivity for Facial Emotions. *Experimental Psychology*, *67*(2), 140–149. https://doi.org/10.1027/1618-3169/a000473

von Dawans, B., Strojny, J., & Domes, G. (2021). The effects of acute stress and stress hormones on social cognition and behavior: Current state of research and future directions. *Neuroscience & Biobehavioral Reviews*, *121*, 75–88. https://doi.org/10.1016/j.neubiorev.2020.11.026

von Dawans, B., Trueg, A., Kirschbaum, C., Fischbacher, U., & Heinrichs, M. (2018). Acute social and physical stress interact to influence social behavior: The role of social anxiety. *PloS one*, *13*(10), e0204665.

Von Siebenthal, Z., Boucher, O., Rouleau, I., Lassonde, M., Lepore, F., & Nguyen, D. K. (2017). Decision-making impairments following insular and medial temporal lobe resection for drug-resistant epilepsy. *Social Cognitive and Affective Neuroscience*, *12*(1), 128–137. https://doi.org/10.1093/scan/nsw152

Vors, O., Marqueste, T., & Mascret, N. (2018). The Trier social stress test and the Trier social stress test for groups: Qualitative investigations. *PLoS One*, *13*(4), e0195722.

Wang, X. T., Rao, L.-L., & Zheng, H. (2017). Neural substrates of framing effects in social contexts: A meta-analytical approach. *Social Neuroscience*, *12*(3), 268–279. https://doi.org/10.1080/17470919.2016.1165285

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weisel, O., & Böhm, R. (2015). “Ingroup love” and “outgroup hate” in intergroup conflict between natural groups. *Journal of Experimental Social Psychology*, *60*, 110–120. https://doi.org/10.1016/j.jesp.2015.04.008

Wendelken, C., O’Hare, E. D., Whitaker, K. J., Ferrer, E., & Bunge, S. A. (2011). Increased Functional Selectivity over Development in Rostrolateral Prefrontal Cortex. *Journal of Neuroscience*, *31*(47), 17260–17268.

Wenseleers, T., & Ratnieks, F. L. W. (2006). Enforced altruism in insect societies. *Nature*, *444*(7115), 50–50. https://doi.org/10.1038/444050a

West, S. A., Griffin, A. S., & Gardner, A. (2007a). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of evolutionary biology*, *20*(2), 415–432.

West, S. A., Griffin, A. S., & Gardner, A. (2007b). Evolutionary Explanations for Cooperation. *Current Biology*, *17*(16), R661–R672. https://doi.org/10.1016/j.cub.2007.06.004

Whitt, S., & Wilson, R. K. (2007). The Dictator Game, Fairness and Ethnicity in Postwar Bosnia. *American Journal of Political Science*, *51*(3), 655–668. https://doi.org/10.1111/j.1540-5907.2007.00273.x

Wiehler, A., Petzschner, F. H., Stephan, K. E., & Peters, J. (2017). Episodic Tags Enhance Striatal Valuation Signals during Temporal Discounting in pathological Gamblers. *eNeuro*, *4*(3), ENEURO.0159-17.2017. https://doi.org/10.1523/ENEURO.0159-17.2017

Wilson, D. S. (2015). *Does altruism exist?: Culture, genes, and the welfare of others*. Yale University Press.

Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and brain sciences*, *17*(4), 585–607.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, *8*, e49547.

Wirz, L., Bogdanov, M., & Schwabe, L. (2018). Habits under stress: Mechanistic insights across different types of learning. *Current Opinion in Behavioral Sciences*, *20*, 9–16. https://doi.org/10.1016/j.cobeha.2017.08.009

Wirz, L., & Schwabe, L. (2020). Prioritized attentional processing: Acute stress, memory and stimulus emotionality facilitate attentional disengagement. *Neuropsychologia*, *138*, 107334. https://doi.org/10.1016/j.neuropsychologia.2020.107334

Wolf, O. T., Schulte, J. M., Drimalla, H., Hamacher-Dang, T. C., Knoch, D., & Dziobek, I. (2015). Enhanced emotional empathy after psychosocial stress in young healthy men. *Stress*, *18*(6), 631–637. https://doi.org/10.3109/10253890.2015.1078787

Wynne-Edwards, V. C. (1964). Group selection and kin selection. *Nature*, *201*(4924), 1147–1147.

Xiang, T., Lohrenz, T., & Montague, P. R. (2013). Computational Substrates of Norms and Their Violations during Social Exchange. *Journal of Neuroscience*, *33*(3), 1099–1108.

Xiao, F., Zheng, Z., Zhang, H., Xin, Z., Chen, Y., & Li, Y. (2016). Who are you more likely to help? The effects of expected outcomes and regulatory focus on prosocial performance. *PloS one*, *11*(11), e0165717.

Yamagishi, T., & Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly*, 116–132.

Yamagishi, T., Mifune, N., Liu, J. H., & Pauling, J. (2008). Exchanges of group-based favours: Ingroup bias in the prisoner’s dilemma game with minimal groups in Japan and New Zealand. *Asian Journal of Social Psychology*, *11*(3), 196–207. https://doi.org/10.1111/j.1467-839X.2008.00258.x

Young, M. E. (2017). Discounting: A practical guide to multilevel analysis of indifference data. *Journal of the Experimental Analysis of Behavior*, *108*(1), 97–112. https://doi.org/10.1002/jeab.265

Youssef, F. F., Bachew, R., Bissessar, S., Crockett, M. J., & Faber, N. S. (2018). Sex differences in the effects of acute stress on behavior in the ultimatum game.

Psychoneuroendocrinology, *96*, 126–131. https://doi.org/10.1016/j.psyneuen.2018.06.012

Yu, H., Hu, J., Hu, L., & Zhou, X. (2014). The voice of conscience: Neural bases of interpersonal guilt and compensation. *Social Cognitive and Affective Neuroscience*, *9*(8), 1150–1158. https://doi.org/10.1093/scan/nst090

Yu, H., Siegel, J. Z., Clithero, J. A., & Crockett, M. J. (2021). How peer influence shapes value computation in moral decision-making. *Cognition*, *211*, 104641.

Zeidman, P., Jafarian, A., Corbin, N., Seghier, M. L., Razi, A., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *NeuroImage*, *200*, 174–190. https://doi.org/10.1016/j.neuroimage.2019.06.031

Zeidman, P., Jafarian, A., Seghier, M. L., Litvak, V., Cagnan, H., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *NeuroImage*, *200*, 12–25. https://doi.org/10.1016/j.neuroimage.2019.06.032

Zhang, Y.-Y., Xu, L., Liang, Z.-Y., Wang, K., Hou, B., Zhou, Y., Li, S., & Jiang, T. (2018). Separate Neural Networks for Gains and Losses in Intertemporal Choice. *Neuroscience Bulletin*, *34*(5), 725–735. https://doi.org/10.1007/s12264-018-0267-x

Zheng, H., Wang, X. T., & Zhu, L. (2010). Framing effects: Behavioral dynamics and neural basis. *Neuropsychologia*, *48*(11), 3198–3204. https://doi.org/10.1016/j.neuropsychologia.2010.06.031

Zhou, Y., Jiao, P., & Zhang, Q. (2017). Second-party and third-party punishment in a public goods experiment. *Applied Economics Letters*, *24*(1), 54–57.

Zilioli, S., Ponzi, D., Henry, A., & Maestripieri, D. (2015). Testosterone, Cortisol and Empathy: Evidence for the Dual-Hormone Hypothesis. *Adaptive Human Behavior and Physiology*, *1*(4), 421–433. https://doi.org/10.1007/s40750-014-0017-x

9. Eidesstattliche Versicherung

hhu
Heinrich Heine
Universität Düsseldorf

Ich, Herr M. Sc. Adam Schweda, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Ich versichere zudem, dass diese Dissertation an keiner anderen Fakultät vorgelegt worden ist.

Düsseldorf, 9. September 2021

Unterschrift

10. Appendix

Contribution statements

Study 1

Schweda, A., Faber, N. S., Crockett, M. J., & Kalenscher, T. (2019). The effects of psychosocial stress on intergroup resource allocation. *Scientific reports*, *9*(1), 1–12.

Contributions:

- Design of the experimental procedure
- Compilation of study materials
- Programming, piloting and implementation of the paradigm
- Coordination and implementation of the data collection
- Data management and analysis of psychometric, physiological and behavioral data
- Authoring and editing of the article

Study 2

Sellitto, M., Neufang, S., Schweda, A., Weber, B., & Kalenscher, T. (2021). Arbitration between insula and temporoparietal junction subserves framing-induced boosts in generosity during social discounting. *NeuroImage*, *238*, 118211. https://doi.org/10.1016/j.neuroimage.2021.118211

Contributions:

- Design of the experimental procedure for studies 1-3
- Compilation of study materials for studies 1-3
- Coordination and implementation of the data collection for studies 1-4

Study 3

Schweda, A., Margittai, Z., & Kalenscher, T. (2020). Acute stress counteracts framing-induced generosity boosts in social discounting in young healthy men. *Psychoneuroendocrinology*, *121*, 104860. https://doi.org/10.1016/j.psyneuen.2020.104860

Contributions:

- Design of the experimental procedure
- Compilation of study materials
- Programming, piloting and implementation of the paradigm
- Coordination and implementation of the data collection
- Data management and analysis of psychometric, physiological and behavioral data
- Authoring and editing of the article

www.nature.com/scientificreports

SCIENTIFIC REPORTS

natureresearch

OPEN

The effects of psychosocial stress on intergroup resource allocation

Adam Schweda[1*], **Nadira Sophie Faber**[2,3], **Molly J. Crockett**[4] & **Tobias Kalenscher**[1]

Stress changes our social behavior. Traditionally, stress has been associated with "fight-or-flight" – the tendency to attack an aggressor, or escape the stressor. But stress may also promote the opposite pattern, i.e., "tend-and-befriend" – increased prosociality toward others. It is currently unclear which situational or physiological factors promote one or the other. Here, we hypothesized that stress stimulates both tendencies, but that fight-or-flight is primarily directed against a potentially hostile outgroup, moderated by rapid-acting catecholamines, while tend-and-befriend is mainly shown towards a supportive ingroup, regulated by cortisol. To test this hypothesis, we measured stress-related neurohormonal modulators and sex hormones in male and female participants who were exposed to a psychosocial stressor, and subsequently played an intergroup social dilemma game in which they could reveal prosocial motives towards an ingroup (ingroup-love) and hostility towards an outgroup (outgroup-hate). We found no significant effects of stress on social preferences, but stress-related heart-rate increases predicted outgroup-hostile behavior. Furthermore, when controlling for testosterone, cortisol was associated with increased ingroup-love. Other-regarding behavior was overall higher in male than female participants. Our mixed results are of interest to scholars of the effects of stress on prosocial and aggressive behavior, but call for refinement in future replications.

Stress is known to alter social behavior. The canonical social response to stress is fight-or-flight[1]. Fight-or-flight responses prepare an organism for homeostasis for antagonistic situations[2], thus increasing the individual's propensity to aggress and flee. The fight-or-flight response to stress is a well-documented phenomenon that has been widely observed in humans and non-human animals[3]. For example, in humans, stress has been shown to reduce empathy[4] and financial generosity[5] to foster domestic and general violence[6,7] and to promote aggressive criminal behavior[8]. Aggressive fight-or-flight responses go along with arousal, activation of the sympathetic nervous system and mobilization of energy resources, and they are linked to rapid-acting catecholaminergic, mainly noradrenergic (NA) components of the stress response[1,9–12].

However, recent theoretical and empirical evidence suggests that stress can also induce prosocial behavior[13]. For example, von Dawans and colleagues[14] found that psychosocial stress increased males' trust in others and their sharing of monetary resources. This was interpreted as support for the "tend-and-befriend"[13] hypothesis. The tendency to "tend-and-befriend" is a proclaimed coping strategy that involves investing into social networks after stress, thus offering costly help to a delimited group of people in order to seek and offer mutual protection during anticipated or experienced threats[13–17]. The tend-and-befriend hypothesis, initially only postulated as being female-specific[16], has received empirical support in recent years. For example, stress has been shown to increase acceptance of even unfair offers in the ultimatum game amongst women[18]. A tend-and-befriend response has been found in males, too. For instance, stress has been shown, across sexes, to increase donation rates among participants with pro-environmental attitudes[19], to increase trust and sharing behavior in male participants[14], and generosity in males[20,21]. Consistent with the tend-and-befriend hypothesis, stressed individuals report more social closeness[22], socio-evaluative stress has been shown to enhance emotional empathy[23], and empathy- and prosociality-related brain areas are activated after a hybrid stress task[24]. Furthermore, consistent with the assumption of tending-and-befriending a close social network under stress, induction of psychosocial stress leads to increased generosity towards socially close recipients of help, for instance relatives and friends, but not towards socially distant others, such as strangers[20].

There is suggestive evidence that prosocial tend-and-befriend responses to stress are linked to relatively slow-acting cortisol (CORT), a component of the physiological stress response that is distinct from fast-acting

[1]Comparative Psychology, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany. [2]College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom. [3]Oxford Uehiro Centre for Practical Ethics, University of Oxford, Oxford, United Kingdom. [4]Department of Psychology, Yale University, New Haven, Connecticut, USA. *email: adam.schweda@uni-duesseldorf.de

		Player 1	Player 2	Player 3
Keep pool	Ingroup	+5€*	0€	0€
*Player invests 5€	Outgroup	0€	0€	0€
Within-Group Pool	Ingroup	+2.50€*	+2.50€	+2.50€
*Player invests 5€	Outgroup	0€	0€	0€
Between-Group Pool	Ingroup	+2.50€*	+2.50€	+2.50€
*Player invests 5€	Outgroup	−2.50€	−2.50€	−2.50€

Table 1. Payoff matrix of the IPD-MD for each pool separately. Outcomes are displayed for each player after player 1 from the ingroup invests 5€ (example).

catecholamines. For example, exogenous manipulation of CORT activity has been shown to foster financial altruism towards socially close others[21], and stress-related CORT-levels covary with greater trust[25] and social affiliation[22]. The idea that the separate components of the physiological stress response, NA and CORT, have dissociable effects on social behavior has been supported by the recent observation that CORT-related financial altruism could be counteracted by additional administration of yohimbine (an alpha2-adrenoceptor antagonist that boosts NA release[21]). This discovery is in line with the finding that noradrenergic activity correlates negatively with overall financial generosity[20] and even implicit intergroup bias[26–28].

Thus, existing evidence suggests stress can promote aggressive (fight-or-flight) as well as prosocial tendencies (tend-and-befriend), and these two tendencies are tentatively related to distinct components of the physiological stress response. However, it is currently unclear when and why stressed individuals show tend-and-befriend or fight-or-flight behavior. Here, we propose that stress does not provoke one or the other response, but boosts both tendencies at the same time by supporting prosocial behavior towards socially close others (tend-and-befriend), who, unlike strangers, can potentially provide comfort and support in stressful times[20,21]. Simultaneously, stress could foster aggression against socially distant outgroup members who are more likely to present a threat than ingroup members (fight-or-flight). Because of recent evidence for a role of CORT in promoting generosity towards others[20,21,24,25], and the classic association of fight-or-flight tendencies with sympathetic activation, we further hypothesize that tend-and-befriend and fight-or-flight tendencies are modulated by the dissociable actions of the stress-neuromarkers CORT and NA, with CORT promoting prosociality towards ingroup members, and NA fostering aggressive behavior against outgroup members.

Stress induces a complex, non-linear and time-dependent suite of neurohormonal changes. CORT and NA exhibit different response profiles, with NA peaking shortly after stress onset, and CORT roughly 20–30 minutes later[29]. CORT effects on neural activity can further be categorized into faster-acting non-genomic CORT action and slower, but longer-lasting (up to several days in animals) genomic CORT effects[30]. Here, participants play the IPD-MD within a 10-minute time window after offset of the gTSST, at the time at which we expect CORT and NA-action to overlap[31].

Moreover, since male participants are known to respond to stress differently than females[32,33] and often reveal different, gender-dependent social preferences[34,35] we additionally considered gender in our main analyses, as well as the sex hormones testosterone, estradiol and progesterone[36–39], and a range of other state and trait variables.

To test these hypotheses, we induced psychosocial stress in 100 male and 102 female participants (total n = 202), using the group version of the Trier Social Stress Test (gTSST[40]). After performing the gTSST or control procedures, participants played an adapted version of the intergroup prisoner's dilemma maximizing differences game (IPD-MD[41]). In this game, three participants were assigned to one group, and they were told they would play against another group that participated on the previous day. To manipulate group affiliation, participants were instructed that the members of their own group held similar political views (ingroup), and that the members of the other group held radically opposing political views (outgroup[42]). At the beginning of the game, each participant received an initial economic endowment, which they could distribute across three pools (keep-pool, within-group pool and between-group pool). Contributions to the keep-pool would be kept by the participants; 50% of the total sum of contributions to the within-group pool would be paid out to each in-group member, including the participant; contributions to the between-group pool had the same effect to the ingroup, but each outgroup member would lose the amount each ingroup member received (see Table 1 for payoff matrix and example). Hence, contributions to the keep-pool can be interpreted as the motivation to maximize own profit (own-utility maximizing), and contributions to the within-group pool can be interpreted as costly motivation to maximize ingroup profit (called "ingroup love" in the relevant literature[41–47]). Finally, contributions to the between-group pool can be interpreted as motivation to maximize ingroup profit and, at the same time, harm the outgroup (so called "outgroup-hate"). Note that we opted against including a "pure spite" condition that would allow participants to harm the outgroup without giving benefits to the ingroup, as costly spite in the absence of ingroup favoritism occurs very rarely, if ever, in the laboratory or the field[48,49].

We predicted that stressed participants would contribute more money to the between-group pool than non-stressed participants, reflecting the predicted combination of ingroup-love and outgroup-hate, and that contributions to this pool would be correlated with measures of the sympathetic stress response (salivary marker of NA and heart-rate). Furthermore, we expected that the motivation to contribute to either the within-group or the between-group pool over keep-pool investments would be correlated with the amplitude of the salivary CORT response, reflecting the predicted CORT effects on prosociality. Although our stress induction was successful, as indicated by physiological and subjective stress markers, our results did not confirm these hypotheses, at least not unambiguously. While we found evidence for an association between heart-rate increase (indicating sympathetic

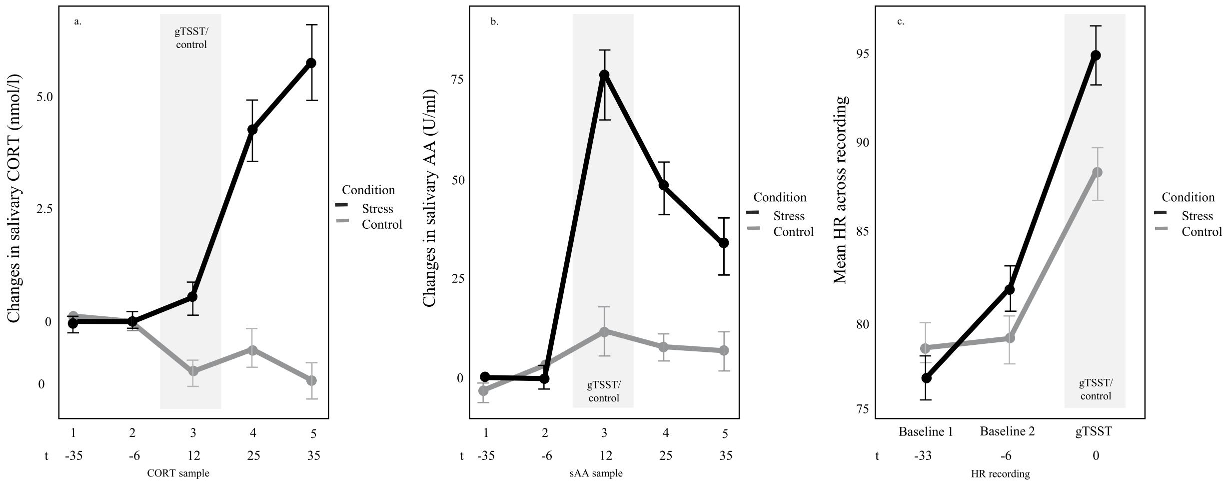


Figure 1. Physiological stress markers. (**a**) Baseline-corrected cortisol (CORT) increased in the stress group, but not the control group ($p < 0.001$). Its peak was reached 20–30 minutes after gTSST onset. (**b**) Stress increased baseline-corrected salivary α-amylase (sAA; $p < 0.001$). SAA increased with stressor-onset. (**c**) Heart-rate increase was more pronounced in stressed than non-stressed participants ($p < 0.001$). All error bars indicate ± 1 SEM.

stress response) and outgroup-hate, the stress manipulation did not significantly affect overall preferences, and there was no correlation between salivary CORT measures and pool investments. However, CORT predicted within-group pool contributions when testosterone was controlled for. Overall, these results suggest a complex relation between stress and intergroup rivalry.

Results

Main results.

Trait measures and group-differences in hormones. To rule out systematic stress-unrelated differences between participants of the stress and control groups, we collected a range of individual trait measures. None of the trait measures differed significantly between groups, with the exception of chronotype and a marginally significant group difference in chronic stress. Moderation analyses with these two factors as potential moderators revealed no significantly influencing role on any of the outcome variables. In addition, there were no significant differences in any of the sex hormone measures (testosterone, progesterone and estradiol) between participants of the stress and control group (cf. SOM for details and analyses).

Manipulation check of stress induction. Compared to controls, participants in the gTSST group had significantly elevated CORT and salivary α-amylase (a measure of central noradrenergic activity[50]) levels (see Fig. 1). In addition, the subjective stress measures also revealed higher levels of psychological stress, such as negative affect, feelings of shame, and insecurity (cf. SOM).

IPD-MD: Main analyses. To assess our main hypotheses, we computed mixed ANOVAs to test for effects of condition (stress vs. control) and gender on investments into the within-group and between-group pools. A priori power analyses for the effect of stress on allocation patterns yield 95% power for at least medium-sized effects (see methods). Consistent with earlier findings[41,45,51], subjects invested more into the within-group pool ($M = 4.17$, $SD = 3.55$) than the between-group pool ($M = 1.51$, $SD = 2.16$, main effect of pool: $F(1, 197) = 59.54$, $p < 0.001$, $\eta_p^2 = 0.232$, see Fig. 2). Male participants contributed more to both pools than female participants, i.e., female participants kept more money for themselves (keep-pool; males: $M = 3.1$, $SD = 3.58$, females: $M = 2.58$, $SD = 2.82$, main effect of gender: $F(1, 197) = 4.79$, $p = 0.030$, $\eta_p^2 = 0.024$, see Fig. 2). However, there were no significant interaction effects between stress, gender and pool (gender x condition: $F(1, 197) = 0.02$, $p = 0.879$, $\eta_p^2 < 0.001$; gender x pool: $F(1, 197) = 0.04$, $p = 0.837$, $\eta_p^2 < 0.001$; gender x pool x condition: $F(1, 197) = 0.12$, $p = 0.724$, $\eta_p^2 = 0.001$). Contrary to our predictions, stress did neither significantly increase within-group, nor between-group pool investments (main effect of condition: $F(1, 197) = 0.04$, $p = 0.844$, $\eta_p^2 < 0.001$, 95% CI [0, 0.019]; interaction effect condition x pool: $F(1, 197) = 0.13$, $p = 0.0.720$, $\eta_p^2 = 0.001$, 95% CI [0, 0.025]). Please note that the latter effect sizes' 95% confidence interval upper bound can still be considered a small effect.

To receive a realistic distribution of plausible stress-related effects and, hence, further elucidate our null finding, we additionally computed Bayesian credibility intervals, which have been shown to produce high coverage of true parameters[52]. To this end, the main effect of condition and its interaction with pool were estimated in a full Bayesian mixed linear model (see SOM for details). The results indicate comparably narrow credibility intervals around zero for stress-related effects, except for the slight increase of investments into the between-group pool (pool: $\beta = -0.82$, 95%-CrI [-1.00, -0.64]; condition: $\beta = 0.02$, 95%-CrI [-0.16, 0.20]; pool x condition: $\beta = 0.08$, 95%-CrI [-0.28, 0.43]). The posterior distributions for condition, pool and their interaction are displayed in Fig. 3. Equivalence tests ("region of practical equivalence", ROPE[53]) based on the stress-related posteriors'

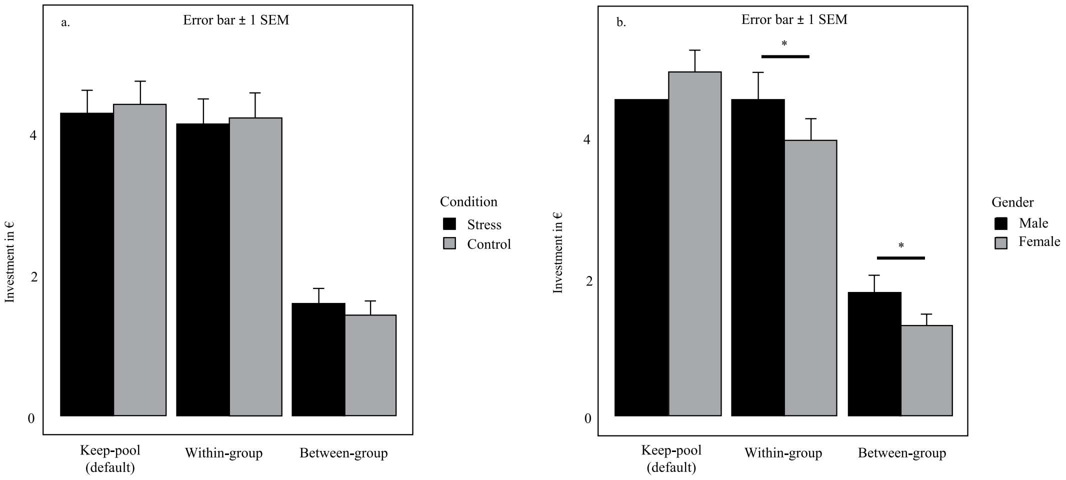


Figure 2. Contributions to the keep-, within-group and between-group pool in the IPD-MD. (**a**) Although participants made more keep- and within-group than between-group investments (more egoism and ingroup-love than outgroup-hate), psychosocial stress did not alter investment patterns. (**b**) Main effect of gender on pool investments. Male participants invested more into the within-group and the between-group pools than females, irrespective of whether they underwent the gTSST procedure or not, suggesting more other-regarding behavior in male than female participants ($p = 0.03$).

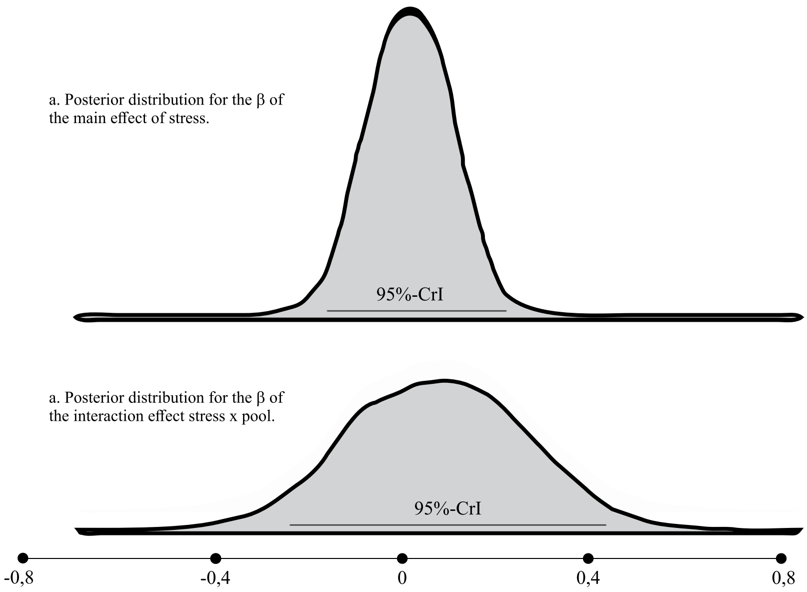


Figure 3. Posterior distributions of the effects of stress on IPD-MD contributions. Horizontal lines mark 95% credibility intervals. (**a**) posterior probability distribution for the standardized regression coefficient of the effect of stress; a marked deviation from zero would indicate that stress influences both, within-group and between-group pool investments, in some, but the same direction. However, it is centered around the mean of 0.02 and the 95% credibility intervals are bounded at -0.16 and 0.20, indicating that – given a 95% criterion – the standardized difference is unlikely to be larger than 0.20 (effect size at the boundary of the heavier tail). (**b**) posterior probabilities of the interaction term pool x condition. Heavy deviations from zero would indicate that stress affects within- and between-group pool investments differentially, for example by only increasing outgroup-hate. Although the posterior mean is close to zero (0.08) and the left tail's 95%-CrI bounds at -0.28, the posterior distribution is right-tailed with an interval boundary at 0.43. This very likely results from a slight, but, in frequentistic terms, statistically insignificant increase of between-pool investments in the stress vs. the control group.

95%-highest density intervals (HDI) accept the null hypothesis if defined as HDIs congruent with an interval of ~20% (for the main effect of stress) or ~45% (for the interaction pool x stress) of the grand standard deviation around zero[53].

To further corroborate these null results, we applied Bayesian hypothesis testing to obtain a quantitative estimate of the evidence for the null hypothesis, using a model comparison approach (see SOM for details). For the following calculations, we set up a mixed linear model with pool and condition as fixed effects and a varying intercept per subject (see SOM for details). Taking into account the reverse of the Bayes factor in favor of the alternative hypothesis - the BF01 - we find evidence for our null hypothesis (only condition: BF01 = 8.523, "moderate" evidence[54]; condition and the interaction term condition x pool: BF01 = 51.476, "very strong" evidence[54]), and this result remains stable for a wider array of prior definitions (see SOM for prior robustness checks). Hence, given the centeredness of the posterior distributions of the stress-related effect-estimates around zero, as well as the Bayes Factors for the null hypothesis, the most reasonable conclusion is that stress does either have no or only a small effect on IPD-MD investments.

Stress markers, sex hormones and investment patterns in the IPD-MD. Next, we asked if investment decisions in the IPD-MD were moderated by changes in the levels of stress markers CORT, α-amylase and heart-rate, independent of a main effect of stress. To this end, we regressed stress marker estimates (α-amylase and CORT, and heart-rate) on the contributions to the pools. As estimates, we considered the area under the curve with respect to increase (AUCi[55]) for CORT and α-amylase, and the increase in heart-rate from the average of the two baseline recordings to the gTSST/control procedure. A mixed linear model with pool, stress markers CORT, α-amylase and HR, as well as their interactions with pool and condition as fixed effects was calculated. Intercepts varied per subject. Our results revealed that changes in heart-rate, but not CORT or α-amylase, modulated decision behavior. We found a significant interaction effect between pool and heart-rate increase ($\beta = 0.196$, $t(298) = 3.141$, $p = 0.002$). Simple regressions on the between- and within-group pool separately indicate an association between heart-rate increase and decreasing ingroup-friendly, as well as increasing outgroup-hostile investments (within-group pool: $\beta = -0.238$, $t(149) = -2.223$, $p = 0.028$, uncorrected; between-group pool: $\beta = 0.153$, $t(149) = 2.41$, $p = 0.017$, uncorrected). There were no main or interaction effects of α-amylase or CORT on pool investments (all $p > 0.20$.). Additional robust Bayesian parameter estimations revealed identical results (cf. SOM).

This finding is partly consistent with one of our main hypotheses that sympathetic activity, with heart-rate as a proxy, should be related to an increased tendency to cause outgroup harm. However, our second proxy of sympathetic activity, α-amylase estimates, was not significantly correlated with outgroup harm, thus limiting our ability to make a definite decision on our hypothesis.

None of the sex hormones testosterone, progesterone & estradiol, nor their interactions explained variance in IPD-MD contributions (cf. SOM). Inspired by the dual-hormone hypothesis[56,57] that predicts an interaction effect of CORT and testosterone on behavior, we investigated if any of the sex hormones, particularly testosterone, moderated (hidden) effects of CORT on IPD-MD decisions. For interpretability, we constructed three different mixed linear models in an exploratory analysis in which we regressed pool (within-/between-group), the respective sex hormone, the area under the curve of CORT, and their interaction terms on the investments. In order to condition on participants' gender and control for different effects of the sex hormones for males and females, we additionally entered gender as a factor. We let the intercept vary for each participant. Only the model including testosterone yielded significant findings: when testosterone was considered in the model, CORT increase (AUCi), as well as testosterone levels predicted pool investments. CORT levels were associated with an increase of allocations into the within-group pool, and testosterone boosted both within- and between-group pool investments (interaction pool x CORT $\beta = -0.149$, $t(329) = -2.444$, $p = 0.015$; testosterone $\beta = 0.207$, $t(329) = 2.039$, $p = 0.042$). CORT itself reached marginal significance ($\beta = 0.115$, $t(329) = 1.876$, $p = 0.062$). This suggests that CORT and testosterone levels explained the variance in within- and between-group investments that was not accounted for by each hormone alone. There was no significant interaction between testosterone and CORT on pool investments ($\beta = -0.018$, $t(329) = 0.291$, $p = 0.711$), testosterone and pool ($\beta = -0.044$, $t(329) = -0.808$, $p = 0.420$), nor any of the higher order interactions (all $p > 0.17$). Also, gender has not reached significance in this model (gender $\beta = 0.107$, $t(329) = 1.104$, $p = 0.270$). Robust Bayesian models show similar results (cf. SOM). Thus, in summary, we found that heart-rate increase predicted a shift from within- to between-group pool investments, and that CORT and testosterone levels explained the variance in within- and between-group investments if considered in a model that conditions on both hormones. This finding is partly consistent with our hypothesis predicting a double-dissociation of sympathetic activity and CORT on ingroup-love (within-group investments) and outgroup-hate (between-group investments). However, it has to be interpreted with caution because of the complexity of the results and the inconsistency in redundant stress marker effects, e.g., the lack of correlation of α-amylase with pool investments.

Exploratory Analyses

Sex-hormonal underpinnings of gender differences in IPD-MD investments. We ran additional exploratory analyses to further elucidate the gender differences in pool investments described above, suggesting that male participants invested more into both the within- and between-group pool than females (see SOM for details). We asked if these gender differences in behavior can be explained by diverging sex hormone profiles or trait measures that we collected. However, these analyses revealed that the gender differences in pool investments were unrelated to differences in the sex hormone compositions and trait measures, except for the emerging value of testosterone as a predictor when cortisol is considered (see above). Since this result needs to be considered with caution, and sex hormones are not directly related to pool investments, our data suggest that males' higher contributions to both the within- and between-group pool likely reflected factors not considered in this study such as, for example, physiology-independent gender differences or same-sex group composition.

IPD-MD investments and chronic stress. It is possible that chronic, but not acute stress (as induced by the gTSST procedure), altered investment behavior in the IPD-MD. Indeed, we found that an increase in chronic stress – as measured by the Trier Inventory of Chronic Stress[58] (TICS; see SOM) - was associated with an overall decrease in ingroup-love and outgroup-hate, and thus, in other-regarding preferences (TICS $\beta = -0.089$, $t(394) = -2.024$, $p = 0.043$). This suggests that chronic stress, as opposed to acute psychosocial stress, is related to an overall disengagement from other-regarding investments. See SOM for analyses on other trait variables.

Discussion

We measured the effects of psychosocial stress on social decision-making in an intergroup rivalry setting. We predicted that stress would increase ingroup-love, and, at the same time, promote outgroup-hate. We further expected that the prosocial effects of stress towards the ingroup would be related to the CORT-component of the neuroendocrine stress response, while aggressive tendencies of outgroup harm would be associated with the sympathetic part of the stress response, mainly NA action. We further considered the potentially moderating role of a range of other endocrine, trait and state variables, including gender, and sex hormones. Although our results are mixed, as discussed below, we found no straightforward support for our hypotheses.

Regarding our main hypothesis, there was no significant effect of psychosocial stress on ingroup-love, outgroup-hate, or selfish choice. A priori power analyses indicated sufficient power to capture small to medium-sized effects with our IPD-MD design. The Bayesian credibility interval of the main effect of stress on investments into the within- and between-pool is considerably narrow; if the heavier tailed 95%-bound of the stress-related posterior distributions of the β-estimates is considered an upper limit of a standardized measure of difference ($\beta = 0.20$), it only yields small plausible effects[59], if any. The posterior distribution of the interaction term is tailed towards higher credibility. Here, plausible effect sizes based on the 95% bound of the longer tail ($\beta = 0.43$) of the posterior still range in the medium category[59]. Further calculations of Bayes factors show reasonable evidence in favor of our null hypotheses. This, as well as the narrow frequentist confidence intervals around the effect sizes, point to no – or a non-detectable - effect of our stress manipulation on IPD-MD decisions.

Of course, this analysis still leaves room for doubt of a true null effect. We could accept the HDI-based null hypothesis using criteria of 20% (for the main effect of stress) or 45% (for the interaction between pool and stress) of the standard deviation around 0 in an equivalence test, but these criteria are still very liberal[53]. The variability in individual contributions to the pools was large, so that we cannot exclude the possibility with certainty that we simply failed to detect small or noisy stress effects on choice. However, our sample size is large, and given our initial power, the centeredness of the posterior distributions, and the Bayes Factor analyses the most likely interpretation of our results is that psychosocial stress had very small or non-existent effects on investment behavior in the IPD-MD game.

We further hypothesized that outgroup-hate was linked to catecholaminergic activity, so that catecholamine action, especially noradrenaline, would boost between-group investments. Our secondary analyses yielded at least partial evidence for this hypothesis. Heart-rate increase from baseline to stress, which is heavily influenced by catecholaminergic action[60,61], predicted a decrease of investment into the within-group pool, and a slight, statistically significant increase into the between-group pool. This suggests that sympathetic activity correlates with decreased ingroup-love, consistent with previous findings[21], as well as increased outgroup-hate. We note that heart-rate responses to stress were associated with reduced prosociality and enhanced harm infliction in the absence of a main effect of stress on social decision-making.

However, while sympathetic activity is associated with enhanced NA release, our other marker of NA activity, α-amylase, did not significantly correlate with pool investments in the IPD-MD game. Hence, the question remains why heart-rate, but not α-amylase levels, predicted changes in ingroup-love and outgroup-hate. The heart-rate response is a temporally well-resolved measure of sympathetic activity, including, but not restricted to, noradrenergic release, and is also to a degree influenced by the parasympathetic nervous system[60]. α-amylase is mainly secreted by the parotid glands, it is directly controlled by sympathetic input, and linked to plasma noradrenaline[50,62]. Correlations between sympathetic indicators (such as skin conductance level[63] and ventricular ejection time[64]) and α-amylase are moderate[50] and often noisy. Also in our study, heart-rate and α-amylase correlate significantly, but weakly ($r = 0.19$, $p = 0.018$). Thus, both measures show complex relationships with sympathetic activation, and each have their own caveats[65,66] because they measure different subprocesses of arousal. This might be one reason why other studies found behavioral measures to be correlated with one marker, but not the other[20]. Thus, the relationship between the sympathetic stress response and a shift away from ingroup-love to outgroup-hate is possibly real, but the specific mechanisms are complex and need to be illuminated in future studies.

We further hypothesized that the CORT response to the stress manipulation would be correlated with ingroup-love[21,25]. Neither stress nor CORT were directly associated with ingroup-love, but when controlling for the variance explained by testosterone, CORT indeed positively predicted within-group pool investments. Also, if testosterone levels were conditioned on CORT, testosterone predicted an increase in overall within- and between-group pool contributions. Note that, because the default option was to keep the investment (contributions to the keep-pool; see methods), increased investments into within- or between-group pools reflect a dominance of other-regarding over selfish motives. Hence, this finding suggests that, when considering the variance explained by either hormone, CORT and testosterone indeed predict ingroup-love and other-regarding behavior.

At first glance, this result seems consistent with the dual-hormone hypothesis[42,53]. This hypothesis postulates that testosterone needs low CORT to predict aggression, while CORT would boost empathy, but only at high testosterone levels[54]. However, despite the fact that, here, CORT levels predicted IPD-MD decisions only when controlling for testosterone (and vice versa), we found no statistical interaction between CORT and testosterone, nor any three-way interaction with pool; our results therefore cannot be readily interpreted as a CORT-testosterone moderation effect on social choice. Thus, our results are, once more, exploratory, complex and call for further investigation.

Not unexpectedly, male participants revealed more outgroup-hate, but also more ingroup-love (higher within- and between-group than keep investments) than female participants. Hence, male participants showed more other-regarding behavior while females were more selfish. This finding is consistent with much of the literature on gender differences in cooperative and competitive behavior, showing that men are often more competitive than women, but can also form cooperative alliances to reach a common goal and protect resources and social status[34,67–71]. The ultimate cause of such behaviors is often discussed in terms of the evolutionary importance of forming strong male bonds to enhance chances of success in intergroup conflicts[70]. Especially testosterone has been linked to rapidly and adaptively increasing the drive for competition in social settings[32]. The fact that we found a simultaneous increase in generosity and hostility in males might conform with possible bidirectionalities caused by testosterone. For example, exogenous testosterone administration has been found to both increase[61] and decrease[72] generosity in the ultimatum game. However, the evidence that testosterone mediated the gender effects on other-regarding behavior in the present study was weak: none of our sampled sex hormones were directly correlated with the male participants' increased ingroup-love and outgroup-hate, except when testosterone and CORT were considered in one exploratory model, and none of our trait measures, such as psychopathy or social value orientation, explained the gender-dependent variance in IPD-MD choices. Hence, the proximal mechanisms underlying the gender differences in other-regarding behavior in the present study remain elusive and might be caused by other factors, such as non-physiological gender differences, that were not considered in this study. For example, a recent meta-analysis on the effects of same- versus mixed-sex group compositions on social behavior found that female participants are slightly less cooperative than men ($d = 0.16$) in same-sex settings[73], suggesting that non-biological, environmental factors matter in shaping prosocial attitudes. Moreover, the interpretation of the small size of the gender effect (partial $eta^2 = 0.024$) requires caution, too.

Interestingly, exploratory analyses reveal an association between subjective chronic stress and an overall disengagement from other-regarding investments. Chronic stress has been frequently discussed as a trigger for social idleness, or even aggression; for example, rodents exposed to chronic physical stressors show a shutdown in social motivation, as well as antagonistic behavior against conspecifics[3,74], and in humans, depressive periods that correlate with chronic stress might be accompanied by social isolation[75], and sometimes even with sudden aggressive outbursts[76]. Of course, the correlative nature of our finding prompts caution, but if chronic stress could be identified as causal to decline of social engagement, this would shed light on the causes of some of the core symptoms of chronic stress disorders, and it would even have significant implications for our understanding of societal cohesion. Future studies should investigate the relationship between chronic stress in social and intergroup contexts e.g. by using modern approaches to causal inference in longitudinal data. Also, note that caution is required here because of our result's exploratory nature, and the inherent potential for false-positive conclusions. Yet, our result is consistent with evidence from the animal literature suggesting increased social apathy, or even agression, with chronic stress[3].

The most robust result of the present study is the null effect of psychosocial stress on pool investment in the IPD-MD game. The absence of any acute stress effect on social decision making is puzzling, given the vast number of studies, including experiments from our own lab, that found such effects on social choice[14,17,18,20–22,24,77,78]. It is unlikely that the null-effect of stress on social choices was due to ineffective or insufficient stress induction because all physiological and psychological stress markers indicate the contrary, i.e., successful stress induction in our participants. There are several other reasons why our manipulation might have failed to work. For instance, it is plausible that our implementation of the IPD-MD was not sufficiently sensitive to the social constructs it was supposed to measure, despite recent claims to the contrary with other implementations[42,79]. For instance, our task might have prompted deliberate, strategic thinking, but relied to a much lesser degree on social sentiments, such as ingroup-love and outgroup-hate[80,81]. That is, our participants' predominant motive might have been payoff maximization, achieved by risk assessments and reciprocity expectations, and an attempt to balance these to make an optimal decision. Hence, social feelings like ingroup-love and outgroup-hate might have been overshadowed by these strategies, or they might have been outright irrelevant for task performance. Additional analyses (cf. SOM) partially support this possibility, suggesting that the IPD-MD might indeed be a predominantly strategic decision game, and less as an instrument to capture affectively tinted ingroup-love and outgroup-hate.

In addition, it is also possible that our ingroup-outgroup manipulation was not salient enough to produce true intergroup rivalry. We used political voter preferences to induce social group affiliation. Although our political framing induced relatively strong outgroup harm (between-subject pool investments) as compared to other studies (15.1% in our study vs. 6% in the original work by Halevy and colleagues), it was, overall, still low (a recent replication[82] also reached 18.33% on average). And only half of our participants allocated money into the between-group pool at all. Consistent with Weisel and Böhm[42], we conclude that political rivalries in Germany may not be strong enough to produce a robust outgroup harm response. The expression of any covert tendencies for outgroup harm might, additionally, be shackled by prevailing social norms prohibiting interpersonal aggression. Furthermore, it is uncertain whether similar political preferences generate the group cohesion necessary to produce ingroup affiliation, or even ingroup-love. Future studies could replicate the present experiment with a more emotionally salient intergroup manipulation aimed at overcoming harm aversion, such as the recruitment of members of minorities[83] or rivalling ethnic tribes[84].

Also, the temporal dynamics of the stress response need to be considered in future research. Here, participants performed the IPD-MD directly after stressor offset where CORT and NA-action are supposed to act in concert[31]. In order to obtain further evidence for our postulated dissociation between CORT for tend-and-befriend and NA for fight-or-flight, it would be interesting to administer the IPD-MD in different time windows after stress. If our hypothesis was true, then fight-or-flight tendencies should preferably occur during and shortly after the stressor, when catecholaminergic action peaks, while tend-and-befriend behavior should be predominantly found in the aftermath of stress during (genomic or non-genomic) CORT action. Also, future research should target the proximate neural mechanisms of "tend-and-befriend" and "fight-or-flight" in social decision making.

Recent neuroimaging literature suggests that processes involving cooperation and intergroup cognition reliably recruit frontolimbic brain networks[85]. At the same time, acute stress was found to cause a massive reorganization of functional network dynamics in the brain, including the prefrontal cortex, the amygdala, as well as subcortical and orbitofrontal valuation-related areas[31,86,87], suggesting that stress-related changes in those frontolimbic networks might underlie the promotion of tend-and-befriend or fight-or-flight tendencies. In addition, social decision making is known to be moderated by a range of other mental faculties, including reasoning, cognitive control, time-preference, memory and attention[88–94] – faculties that are subserved by the same neural networks that are altered by stress. Hence, it is an open question if and how these mental functions and their underlying neural processes moderate or even mediate stress-effects on social behavior.

Overall, our study shows no direct effect of socio-evaluative stress on social decisions in the IPD-MD. However, heart-rate changes in response to stress were associated with a shift from ingroup-love to outgroup-hate. In addition, when considered in one model, CORT and testosterone action was associated with ingroup-love or both, ingroup-love and outgroup-hate - thus, other regarding behavior. In general, male participants revealed more other-regarding preferences than female participants, and this gender-effect was neither explained directly by sex hormones nor gender-dependent traits. Our study contributes to the literature on stress and decision making by highlighting the boundaries of stress effects on social choice; we conclude that our version of the IPD-MD failed to capture stress-related effects on social sentiments, probably due to its strategic character. Future studies are needed to further elucidate putative stress effects on social behavior.

Methods

Participants. One hundred and three male (age: $M = 24.85$, $SD = 4.77$), and 105 female participants (age: $M = 24.71$, $SD = 5.13$) were recruited within the Düsseldorf (Germany) area. Three male and three female subjects were excluded, either for incidentally knowing other members in their group (three females and two males), or for misunderstanding the intergroup prisoners' dilemma rules, as revealed by the comprehension questions (one male). Based on a priori power analyses (G*Power[95]), we opted for a sample size that yielded 95% power for small to medium effect sizes for our pool x condition interaction ($n = 206$, Cohen's $f = 0.15$, repeated measures correlation $= 0.3$). In case of a between-group effect, our experiment could detect effect sizes of 0.205 with a power of 0.95 ($n = 204$, repeated measures correlation $= 0.3$). With a final sample size of $n = 202$, our experiment is comparably well-powered.

Before participation, participants were screened via telephone interview for a number of eligibility criteria. Participants required to hold at least moderate sympathy for one of Germany's five political parties with seats in the German parliament, except the right-wing populist party (*Alternative für Deutschland*, AfD). We applied a number of further eligibility criteria and participation rules, as outlined in the SOM.

The study was approved by the ethical committee of the Heinrich Heine University Düsseldorf, and our methods were performed in accordance with the committee's rules and guidelines.

Material. *Trait measures and cognitive reflection.* To exclude potential confounds and ensure similarity between participants in stress- and decision-making related traits, we collected a number of trait measures before the laboratory experiment and before the stress induction using online survey tools. These trait measures are described and summarized in the SOM.

Stress induction. All experimental sessions took place between 14:00 and 17:00 h to control for diurnal variation in CORT levels. We tested all participants in groups of three. They were randomly assigned to a stress condition or a control condition. Socio-evaluative stress was induced using the Trier Social Stress Test for groups[40]. In the stress condition, participants were exposed to a fictional job interview (net speaking time three minutes per participants) and, subsequently, to a mental arithmetic task in presence of other participants and in front of a non-responsive evaluation panel of experimenters, while their performance was video-taped. In the control condition, participants prepared a short talk about their friends, and they also performed an arithmetic task, but the evaluation panel paid ostensibly no attention to the participants. Participants spoke simultaneously and they were not videotaped. Stress and control conditions were matched in terms of cognitive load, speaking time, participant engagement etc., but differed in the socio-evaluative component[20,40].

Physiological and subjective stress measures. We collected multiple saliva samples to determine stable baseline measures of the sex hormones progesterone, estradiol and testosterone, and to quantify the impact of our stress manipulation based on CORT and α-amylase. For the sex hormones, subjects filled ultra-pure polypropylene spit-in samples (SaliCaps, IBL International GmbH, Hamburg, Germany) with 1 mL of clear saliva. Three separate sex-hormone samples were collected in the first half of the experiment before subjection to the stress or control procedure (see Fig. 4). CORT/α-amylase samples (Salivette®, Sarstedt AG & Co. KG, Nuernbrecht, Germany) were collected throughout the entire experiment (see Fig. 4 for exact sample time points and SOM for a description of the timeline). Further details of saliva sampling and analysis procedures are provided in the SOM. As an additional measure of sympathetic activity, we recorded heart-rate using a commercially available HR-monitor (Polar A370, Polar Electro Oy, Kempele, Finland). Heart-rate was measured at several time points during the experiment (see Fig. 4 and SOM). Participants indicated positive and negative mood by completing a Positive and Negative Affect scale (PANAS[96]) before, during and after the gTSST/control procedure (see Fig. 4 and SOM), and, they also indicated current feelings of shame, insecurity, stress and confidence on visual analogue scales before, during and after the experimental procedures (VAS, 1–100).

Intergroup prisoners' dilemma - maximizing differences (IPD-MD). We used political preferences to induce intergroup rivalry. Unlike other means of group manipulations[49], group assignments based on political

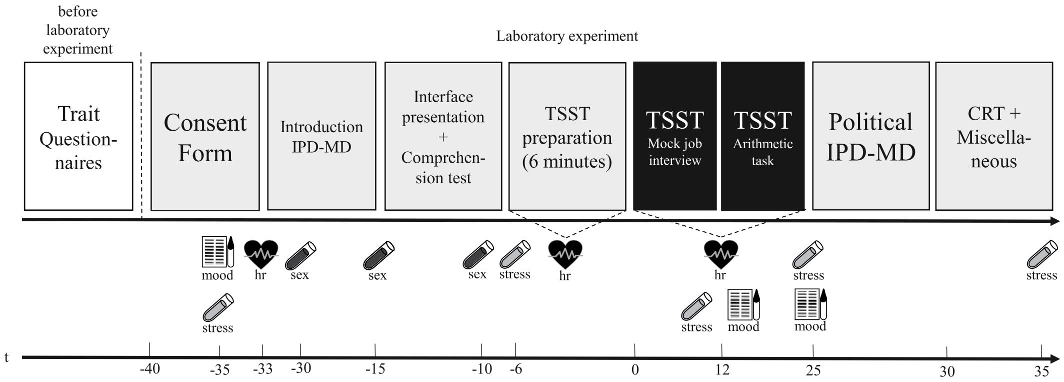


Figure 4. Illustration of experimental timeline. After giving the phone interview and completing the online questionnaires, participants were invited to the laboratory. All experimental sessions took place between 14:00 and 17:00. Participants gave informed consent and were familiarized with the different types of saliva sampling and the HR-monitor. Then, the first heart-rate baseline recording of 3 minutes started and the sex hormone and baseline CORT and α-amylase samples were collected together with the first PANAS measurement. The IPD-MD instructions were then given individually and comprehension was exhaustively tested, interspersed with the 2nd and 3rd sex-hormone samples. This was followed by heart-rate baseline monitoring for six minutes and gTSST/control procedure instructions, after which the gTSST/control procedure began. Participants completed the PANAS and gave further stress marker saliva samples during and directly after the gTSST/control procedure. Directly following the gTSST/control procedure, subjects played the IPD-MD. The IPD-MD lasted for no more than 10 minutes. The cognitive reflection task (CRT; cf. SOM), as well as a set of decision- and demographics-related questionnaires followed. 10 minutes after the TSST, we collected the last stress-marker saliva sample. The experiment concluded with a debriefing and individual, anonymous payouts.

preferences have been shown to induce ingroup affiliation and outgroup harm while being realistic, feasible and credible[42]. The three participants in each testing session were instructed that they would form a group and play against another group of three other participants who performed the game one day before. They were told that group assignment was based according to the participants' political voting preferences which were assessed before in the online questionnaires. The instructions explicitly stated that all members of the participants' own group held similar political views (ingroup) and that the members of the other group were supporters of the political party "*Alternative für Deutschland*" (AfD; a German right-wing populist party). Outgroup decisions were not real and shammed by the experimenters. We used an adapted version of the IPD-MD to simulate intergroup behavior in the laboratory[41,42,79]. As mentioned above, in this game, two groups of three participants play against each other. Each participant receives the initial monetary endowment of 10 EUR, which they can freely distribute between three pools. Money contributed to the first pool (the "keep" pool), is kept by the player. For example, 5 EUR investment into the keep-pool would imply that the participant can keep those 5 EUR for herself. Fifty percent of the total sum of contributions to the second pool (the "within-group" pool) are paid out to each ingroup member, including the participant. For example, if a participant contributes 5 EUR to the within-group pool, each ingroup member, including the participant, would receive 2.50 EUR payback. Thus, contributions to the within-group pool are potentially costly to the participant because she only receives a back-payment of 50% of the invested sum if no one else contributes, but the overall sum of all payoffs to all group members is higher than individual contributions to the keep pool (the sum of the payoffs to all ingroup members in the above example is 7.50 EUR, see Table 1). Hence, the dominant group strategy[41] would imply that every member contributes to the within-group pool since this would maximize the overall sum of payoffs to all players. For example, if all three ingroup members contributed their entire endowment of 10 EUR each to the within-group pool, the total sum of contributions would amount to 30 EUR, thus each ingroup member would receive a back-payment of 15 EUR.

Contributions to the third pool (the "between-group"-pool) have the same effect to the ingroup members as contributions to the within-group pool, but each outgroup member loses the amount each ingroup member receives. For example, if a participant invests 5 EUR to the between-group pool, each ingroup member, including the participant, would receive 2.50 EUR, and each outgroup member would lose 2.50 EUR. Thus, contributions to the between-group pool represent the same social dilemma as contributions to the within-group pool, but additionally entail the possibility to harm the outgroup.

Contributions to the keep-pool can be interpreted as the motivation to maximize own profit (own-utility maximizing), contributions to the within-group pool can be interpreted as potentially costly motivation to maximize ingroup profit (ingroup-love, or ingroup trust) and contributions to the between-group pool can be interpreted as costly motivation to maximize ingroup profit and, at the same time, harm the outgroup (outgroup-hate).

To ensure participants' full understanding of the game, we presented extensive instructions covering multiple exemplary scenarios before stress induction. Subjects' comprehension was controlled using a set of questions. If subjects revealed difficulties in understanding the rules of the game or its financial implications for the ingroup and outgroup, the respective parts of the game were explained again by the experimenters. The experiment was incentive-compatible. Subjects received a fixed participation fee of 20 EUR plus the gains from the in-group investments in the IPD-MD, minus 2 EUR as result of simulated between-group pool investments by the fictional outgroup decisions. Task completion took less than 10 minutes.

Data analysis. To analyze IPD-MD distribution patterns, we applied a 2 × 2 × 2 mixed-factorial analysis of variance (ANOVA) with pool as a repeated-measures (within-group vs. between group; note that investments into all three pools were not independent, we therefore considered investments into the keep-pool as the default option and compared investments between the within- and between-group pools only) as well as condition (stress vs control) and gender (male vs female) as between-subject factors. To complement this analysis, we ran additional Bayes-Factor analyses on a mixed linear model with pool and condition as fixed factors and subject as a random intercept (using the software JASP[97]). Furthermore, full Bayesian parameter estimation of the same model was used to estimate posterior parameter distributions and to yield information on credible parameter values (R-Package *brms*[98]). We used the R-package *afex*[99] for mixed linear models and ANOVAs.

Ethics approval and consent to participate. The study was approved by the ethics committee of the Medical Faculty of the Heinrich-Heine-University in Dusseldorf. All participants gave their informed consent.

Statement of responsibility. AS designed the task, ran the data collection, analyzed the data and wrote the paper, TK designed the task, wrote the paper, provided consultation at all stages of the project and funded the project, NF and MC edited the paper and provided consultation at all stages of the project[100–102].

Received: 26 March 2019; Accepted: 20 November 2019;

References

1. Cannon, W. B. Natural Defenses of the Organism. In *The wisdom of the body* 213–215 (W W Norton & Co, 1932).
2. Goldstein, D. S. Adrenal Responses to Stress. *Cell. Mol. Neurobiol.* **30**, 1433–1440 (2010).
3. Sandi, C. & Haller, J. Stress and the social brain: behavioural effects and neurobiological mechanisms. *Nat. Rev. Neurosci.* **16**, 290–304 (2015).
4. Negd, M., Mallan, K. M. & Lipp, O. V. The role of anxiety and perspective-taking strategy on affective empathic responses. *Behav. Res. Ther.* **49**, 852–857 (2011).
5. Vinkers, C. H. *et al.* Time-dependent changes in altruistic punishment following stress. *Psychoneuroendocrinology* **38**, 1467–1475 (2013).
6. Silver, E. & Teasdale, B. Mental Disorder and Violence: An Examination of Stressful Life Events and Impaired Social Support. *Soc. Probl.* **52**, 62–78 (2005).
7. Klaw, E. L., Demers, A. L. & Da Silva, N. Predicting Risk Factors for Intimate Partner Violence Among Post-9/11 College Student Veterans. *J. Interpers. Violence* **31**, 572–597 (2016).
8. Agnew, R. Building on the foundation of general strain theory: Specifying the types of strain most likely to lead to crime and delinquency. *J. Res. Crime Delinquency* **38**, 319–361 (2001).
9. Geen, R. G. & O'Neal, E. C. Activation of cue-elicited aggression by general arousal. *J. Pers. Soc. Psychol.* **11**, 289–292 (1969).
10. Allen, J. J., Anderson, C. A. & Bushman, B. J. The General Aggression Model. *Curr. Opin. Psychol.* **19**, 75–80 (2018).
11. Nelson, R. J. & Trainor, B. C. Neural mechanisms of aggression. *Nat. Rev. Neurosci.* **8**, 536–546 (2007).
12. Miczek, K. A., DeBold, J. F., Gobrogge, K., Newman, E. L. & Almeida, R. M. M. The Role of Neurotransmitters in Violence and Aggression. In *The Wiley Handbook of Violence and Aggression* 1–13 (Wiley, 2017).
13. Taylor, S. E. Tend and Befriend: Biobehavioral Bases of Affiliation Under Stress. *Curr. Dir. Psychol. Sci.* **15**, 273–277 (2006).
14. Von Dawans, B., Fischbacher, U., Kirschbaum, C., Fehr, E. & Heinrichs, M. The social dimension of stress reactivity: acute stress increases prosocial behavior in humans. *Psychol. Sci.* **23**, 651–660 (2012).
15. Baumeister, R. F. & Leary, M. R. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* **117**, 497–529 (1995).
16. Taylor, S. E. *et al.* Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychol. Rev.* **107**, 411–429 (2000).
17. Buchanan, T. W. & Preston, S. D. Stress leads to prosocial action in immediate need situations. *Front. Behav. Neurosci.* **8**, 5 (2014).
18. Youssef, F. F., Bachew, R., Bissessar, S., Crockett, M. J. & Faber, N. S. Sex differences in the effects of acute stress on behavior in the ultimatum game. *Psychoneuroendocrinology* **96**, 126–131 (2018).
19. Sollberger, S., Bernauer, T. & Ehlert, U. Stress influences environmental donation behavior in men. *Psychoneuroendocrinology* **63**, 311–319 (2016).
20. Margittai, Z. *et al.* A friend in need: Time-dependent effects of stress on social discounting in men. *Horm. Behav.* **73**, 75–82 (2015).
21. Margittai, Z., van Wingerden, M., Schnitzler, A., Joëls, M. & Kalenscher, T. Dissociable roles of glucocorticoid and noradrenergic activation on social discounting. *Psychoneuroendocrinology* **90**, 22–28 (2018).
22. Berger, J., Heinrichs, M., von Dawans, B., Way, B. M. & Chen, F. S. Cortisol modulates men's affiliative responses to acute social stress. *Psychoneuroendocrinology* **63**, 1–9 (2016).
23. Wolf, O. T. *et al.* Enhanced emotional empathy after psychosocial stress in young healthy men. *Stress* **18**, 631–637 (2015).
24. Tomova, L. *et al.* Increased neural responses to empathy for pain might explain how acute stress increases prosociality. *Soc. Cogn. Affect. Neurosci.* **12**, 401–408 (2017).
25. Steinbeis, N., Engert, V., Linz, R. & Singer, T. The effects of stress and affiliation on social decision-making: Investigating the tend-and-befriend pattern. *Psychoneuroendocrinology* **62**, 138–148 (2015).
26. Terbeck, S. *et al.* Propranolol reduces implicit negative racial bias. *Psychopharmacology (Berl.)* **222**, 419–424 (2012).
27. Terbeck, S., Savulescu, J., Chesterman, L. P. & Cowen, P. J. Noradrenaline effects on social behaviour, intergroup relations, and moral decisions. *Neurosci. Biobehav. Rev.* **66**, 54–60 (2016).
28. Terbeck, S. *et al.* β-Adrenoceptor blockade modulates fusiform gyrus activity to black versus white faces. *Psychopharmacology (Berl.)* **232**, 2951–2958 (2015).
29. Dickerson, S. S. & Kemeny, M. E. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychol. Bull.* **130**, 355–391 (2004).
30. Joëls, M. Corticosteroids and the brain. *J. Endocrinol.* **238**, R121–R130 (2018).
31. Hermans, E. J., Henckens, M. J. A. G., Joëls, M. & Fernández, G. Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends Neurosci.* **37**, 304–314 (2014).
32. Ter Horst, J. P., de Kloet, E. R., Schächinger, H. & Oitzl, M. S. Relevance of stress and female sex hormones for emotion and cognition. *Cell. Mol. Neurobiol.* **32**, 725–735 (2012).
33. Soldin, O. & Mattison, D. Sex Differences in Pharmacokinetics and Pharmacodynamics. *Clin. Pharmacokinet.* **48**, 143–157 (2009).
34. Soutschek, A. *et al.* The dopaminergic reward system underpins gender differences in social preferences. *Nat. Hum. Behav.* **1**, 819 (2017).

35. Rand, D. G., Brescoll, V. L., Everett, J. A. C., Capraro, V. & Barcelo, H. Social heuristics and social roles: Intuition favors altruism for women but not for men. *J. Exp. Psychol. Gen.* **145**, 389–396 (2016).
36. Casto, K. V. & Edwards, D. A. Testosterone, cortisol, and human competition. *Horm. Behav.* **82**, 21–37 (2016).
37. Losecaat Vermeer, A. B., Riečanský, I. & Eisenegger, C. Competition, testosterone, and adult neurobehavioral plasticity. in *Progress in Brain Research* (eds Studer, B. & Knecht, S.) 213–238 (Elsevier, 2016).
38. Geniole, S. N., Bird, B. M., Ruddick, E. L. & Carré, J. M. Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis. *Horm. Behav.* **92**, 37–50 (2017).
39. Dreher, J.-C. *et al.* Testosterone causes both prosocial and antisocial status-enhancing behaviors in human males. *Proc. Natl. Acad. Sci.* **113**, 11633–11638 (2016).
40. Von Dawans, B., Kirschbaum, C. & Heinrichs, M. The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology* **36**, 514–522 (2011).
41. Halevy, N., Bornstein, G. & Sagiv, L. "In-Group Love" and "Out-Group Hate" as Motives for Individual Participation in Intergroup Conflict: A New Game Paradigm. *Psychol. Sci.* **19**, 405–411 (2008).
42. Mehta, P. H. & Josephs, R. A. Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Horm. Behav.* **58**, 898–906 (2010).
43. Brewer, M. B. The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *J. Soc. Issues* **55**, 429–444 (1999).
44. De Dreu, C. K. W. *et al.* The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* **328**, 1408–1411 (2010).
45. Halevy, N., Weisel, O. & Bornstein, G. "In-Group Love" and "Out-Group Hate" in Repeated Interaction Between Groups. *J. Behav. Decis. Mak.* **25**, 188–195 (2012).
46. Hamley, L., Houkamau, C. A., Osborne, D., Barlow, F. K. & Sibley, C. G. Ingroup Love or Outgroup Hate (or Both)? Mapping Distinct Bias Profiles in the Population. *Pers. Soc. Psychol. Bull.*, https://doi.org/10.1177/0146167219845919 (2019).
47. Yamagishi, T. & Mifune, N. Social exchange and solidarity: in-group love or out-group hate? *Evol. Hum. Behav.* **30**, 229–237 (2009).
48. West, S. A. & Gardner, A. Altruism, Spite, and Greenbeards. *Science* **327**, 1341–1344 (2010).
49. Yamagishi, T. & Mifune, N. Parochial altruism: does it explain modern human group psychology? *Curr. Opin. Psychol.* **7**, 39–43 (2016).
50. Nater, U. M. & Rohleder, N. Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: Current state of research. *Psychoneuroendocrinology* **34**, 486–496 (2009).
51. De Dreu, C. K. W., Dussel, D. B. & Velden, F. S. T. In intergroup conflict, self-sacrifice is stronger among pro-social individuals, and parochial altruism emerges especially among cognitively taxed individuals. *Front. Psychol.* **6**, 572 (2015).
52. Nalborczyk, L., Bürkner, P.-C. & Williams, D. R. Pragmatism should not be a substitute for statistical literacy, a commentary on Albers, Kiers, and van Ravenzwaaij. Collabra: Psychology, **5**(1), (2018).
53. Mehta, P. H. & Prasad, S. The dual-hormone hypothesis: a brief review and future research agenda. *Curr. Opin. Behav. Sci.* **3**, 163–168 (2015).
54. Zilioli, S., Ponzi, D., Henry, A. & Maestripieri, D. Testosterone, Cortisol and Empathy: Evidence for the Dual-Hormone Hypothesis. *Adapt. Hum. Behav. Physiol.* **1**, 421–433 (2015).
55. Pruessner, J. C., Kirschbaum, C., Meinlschmid, G. & Hellhammer, D. H. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology* **28**, 916–931 (2003).
56. Weisel, O. & Böhm, R. "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. *J. Exp. Soc. Psychol.* **60**, 110–120 (2015).
57. Carré, J. M. & Mehta, P. H. Importance of considering testosterone–cortisol interactions in predicting human aggression and dominance. *Aggress. Behav.* **37**, 489–491 (2011).
58. Schulz, P. & Schlotz, W. The Trier Inventory for the Assessment of Chronic Stress (TICS). Scale construction, statistical testing, and validation of the scale work overload. *Diagnostica* **45**, 8–19 (1999).
59. Cohen, J. A power primer. *Psychol. Bull.* **112**, 155 (1992).
60. Hall, J. E. Rhythmical Excitation of the Heart. In *Guyton and Hall Textbook of Medical Physiology* 123–129 (Elsevier Health Sciences, 2016).
61. Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M. & Fehr, E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* **463**, 356–359 (2010).
62. Chatterton, R. T., Vogelsong, K. M., Lu, Y., Ellman, A. B. & Hudgens, G. A. Salivary α-amylase as a measure of endogenous adrenergic activity. *Clin. Physiol.* **16**, 433–448 (1996).
63. El-Sheikh, M., Erath, S. A., Buckhalt, J. A., Granger, D. A. & Mize, J. Cortisol and Children's Adjustment: The Moderating Role of Sympathetic Nervous System Activity. *J. Abnorm. Child Psychol.* **36**, 601–611 (2008).
64. Bosch, J. A., de Geus, E. J. C., Veerman, E. C. I. & Nieuw Amerongen, A. V. Innate secretory immunity in response to laboratory stressors that evoke distinct patterns of cardiac autonomic reactivity. *Psychosom. Med.* **62**, 245–258 (2000).
65. Nagy, T. *et al.* A fluid response: Alpha-amylase reactions to acute laboratory stress are related to sample timing and saliva flow rate. *Biol. Psychol.* **109**, 111–119 (2015).
66. Picard, G., Tan, C. O., Zafonte, R. & Taylor, J. A. Incongruous changes in heart period and heart rate variability with vagotonic atropine: implications for rehabilitation medicine. *PM R* **1**, 820–826 (2009).
67. Wakefield, K. L. & Wann, D. L. An Examination of Dysfunctional Sport Fans: Method of Classification and Relationships with Problem Behaviors. *J. Leis. Res.* **38**, 168–186 (2006).
68. Bourgeois, C. A. & Fisher, M. L. More "bros," more woes? The prevalence of male coalitions in crimes of robbery. *Evol. Behav. Sci.* **12**, 126–131 (2018).
69. Massen, J. J. M., Bauer, L., Spurny, B., Bugnyar, T. & Kret, M. E. Sharing of science is most likely among male scientists. *Sci. Rep.* **7**, 12927 (2017).
70. Van Vugt, M., De Cremer, D. & Janssen, D. P. Gender differences in cooperation and competition: the male-warrior hypothesis. *Psychol. Sci.* **18**, 19–23 (2007).
71. McDonald, M. M., Navarrete, C. D. & Van Vugt, M. Evolution and the psychology of intergroup conflict: the male warrior hypothesis. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 670–679 (2012).
72. Zak, P. J. *et al.* Testosterone Administration Decreases Generosity in the Ultimatum Game. *PLOS ONE* **4**, e8330 (2009).
73. Balliet, D., Li, N. P., Macfarlan, S. J. & Van Vugt, M. Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychol. Bull.* **137**, 881–909 (2011).
74. Mineur, Y. S., Prasol, D. J., Belzung, C. & Crusio, W. E. Agonistic Behavior and Unpredictable Chronic Mild Stress in Mice. *Behav. Genet.* **33**, 513–519 (2003).
75. Glass, T. A., De Leon, C. F. M., Bassuk, S. S. & Berkman, L. F. Social Engagement and Depressive Symptoms in Late Life: Longitudinal Findings. *J. Aging Health* **18**, 604–628 (2006).
76. Fava, M. Depression with anger attacks. *J. Clin. Psychiatry* **59**, 18–22 (1998).
77. Von Dawans, B., Ditzen, B., Trueg, A., Fischbacher, U. & Heinrichs, M. Effects of acute stress on social behavior in women. *Psychoneuroendocrinology* **99**, 137–144 (2019).
78. Starcke, K., Polzer, C., Wolf, O. T. & Brand, M. Does stress alter everyday moral decision-making? *Psychoneuroendocrinology* **36**, 210–219 (2011).

79. De Dreu, C. K. W. Social value orientation moderates ingroup love but not outgroup hate in competitive intergroup conflict. *Group Process. Intergroup Relat.* **13**, 701–713 (2010).
80. Loewenstein, G., O'Donoghue, T. & Bhatia, S. Modeling the interplay between affect and deliberation. *Decision* **2**, 55–81 (2015).
81. Hofmann, W., Friese, M. & Strack, F. Impulse and Self-Control From a Dual-Systems Perspective. *Perspect. Psychol. Sci.* **4**, 162–176 (2009).
82. Open Science Collaboration. Replication of N Halevy, G Bornstein, L Sagiv (2008) by Thomae, Wood & Immelman as part of 'Estimating the reproducibility of psychological science'. *Science* **349**, aac4716 (2015).
83. Hein, G., Engelmann, J. B. & Tobler, P. N. Pain relief provided by an outgroup member enhances analgesia. *Proc. R. Soc. B Biol. Sci.* **285**, 20180501 (2018).
84. Bernhard, H., Fischbacher, U. & Fehr, E. Parochial altruism in humans. *Nature* **442**, 912–915 (2006).
85. Cikara, M. & Van Bavel, J. J. The Neuroscience of Intergroup Relations: An Integrative Review. *Perspect. Psychol. Sci. J. Assoc. Psychol. Sci.* **9**, 245–274 (2014).
86. Kinner, V. L., Wolf, O. T. & Merz, C. J. Cortisol alters reward processing in the human brain. *Horm. Behav.* **84**, 75–83 (2016).
87. Porcelli, A. J., Lewis, A. H. & Delgado, M. R. Acute Stress Influences Neural Circuits of Reward Processing. *Front. Neurosci.* **6**, 157 (2012).
88. Amodio, D. M. Social Cognition 2.0: An Interactive Memory Systems Account. *Trends Cogn. Sci.* **23**, 21–33 (2019).
89. Crockett, M. J. *et al.* Serotonin Modulates Striatal Responses to Fairness and Retaliation in Humans. *J. Neurosci.* **33**, 3505–3513 (2013).
90. Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G. & Robbins, T. W. Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. *Emot. Wash. DC* **10**, 855–862 (2010).
91. Peters, J. & Büchel, C. Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron* **66**, 138–148 (2010).
92. Ruff, C. C. & Fehr, E. The neurobiology of rewards and values in social decision making. *Nat. Rev. Neurosci.* **15**, 549–562 (2014).
93. Wing, E. A. *et al.* Neural mechanisms underlying subsequent memory for personal beliefs:An fMRI study. *Cogn. Affect. Behav. Neurosci.* **18**, 216–231 (2018).
94. Yao, S. *et al.* Oxytocin Modulates Attention Switching Between Interoceptive Signals and External Social Cues. *Neuropsychopharmacology* **43**, 294–301 (2018).
95. Erdfelder, E., Faul, F. & Buchner, A. G*Power: A general power analysis program. *Behav. Res. Methods Instrum. Comput.* **28**, 1–11 (1996).
96. Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* **54**, 1063–1070 (1988).
97. The JASP Team. *JASP*. (2018).
98. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
99. Singmann, H. afex: A User-Friendly Package for the Analysis of Factorial Experiments. (in revision).
100. Kruschke, J. K. Rejecting or Accepting Parameter Values in Bayesian Estimation. *Adv. Methods Pract. Psychol. Sci.* **1**, 270–280 (2018).
101. Kass, R. E. & Raftery, A. E. Bayes Factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
102. Kirschbaum, C., Pirke, K.-M. & Hellhammer, D. H. The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology* **28**, 76–81 (1993).

Acknowledgements

We thank Guy Kahane for proofreading the manuscript. The study was supported by a grant from the Deutsche Forschungsgemeinschaft to TK (DFG-KA 2675/4-3), as well as a grant by the Oxford Martin School to NF and MK (Oxford Martin Programme for Collective Responsibility for Infectious Disease). We acknowledge the support of the article processing charges by the Heinrich Heine University Düsseldorf.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-54954-w.

Correspondence and requests for materials should be addressed to A.S.

NeuroImage 238 (2021) 118211

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Arbitration between insula and temporoparietal junction subserves framing-induced boosts in generosity during social discounting

Manuela Sellitto[a,*], Susanne Neufang[b], Adam Schweda[a], Bernd Weber[c], Tobias Kalenscher[a]

[a] Comparative Psychology, Institute of Experimental Psychology, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany
[b] Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich Heine University Düsseldorf, Bergische Landstraße 2, 40629 Düsseldorf, Germany
[c] Institute of Experimental Epileptology and Cognition Research, University Hospital Bonn, Sigmund-Freud Straße 25, 53127 Bonn, Germany

ARTICLE INFO

Keywords:
DCM
Framing effect
Insular cortex
Social discounting
TPJ
VMPFC

ABSTRACT

Generosity toward others declines across the perceived social distance to them. Here, participants chose between selfish and costly generous options in two conditions: in the gain frame, a generous choice yielded a gain to the other; in the loss frame, it entailed preventing the loss of a previous endowment to the other. Social discounting was reduced in the loss compared to the gain frame, implying increased generosity toward strangers. Using neuroimaging tools, we found that while activity in the temporoparietal junction (TPJ) and the ventromedial prefrontal cortex (VMPFC) was associated with generosity in the gain frame, the insular cortex was selectively recruited during generous choices in the loss frame. We provide support for a network-model according to which TPJ and insula differentially subserve generosity by modulating value signals in the VMPFC in a frame-dependent fashion. These results extend our understanding of the insula role in nudging prosocial behavior in humans.

1. Introduction

Most human societies are collaborative. Collaboration offers benefits to their members that they would not be able to achieve individually. However, societies can only function efficiently if their members are willing to contribute to causes whose beneficiaries are abstract and anonymous, such as public goods, and/or to causes whose beneficiaries are socially remote, as it is often the case with wealth redistribution for social welfare, public health insurance, or state pension systems (see also Kalenscher, 2014). Most people are indeed willing to sacrifice own resources for the welfare of others (Nowak, 2006; Rilling and Sanfey, 2011), but their generosity typically declines steeply with social distance between them and the recipients of help, a phenomenon called social discounting (Jones and Rachlin, 2006; Strombach et al., 2015). Hence, while people are ready to provide costly support to friends, relatives, and acquaintances, they are less inclined to help remote strangers.

The social discount function is idiosyncratic (Kalenscher 2017; Vekaria et al., 2017; Archambault et al., 2019), but it is far from stable within and across individuals. For instance, we and others have shown that participants from individualistic or collectivistic cultures (Strombach et al., 2014) differ in their attitude towards the welfare of socially close peers; that psychosocial stress (Margittai et al., 2015) and neurohormonal stress action (Margittai et al., 2018) can increase generosity towards socially close friends and acquaintances; and that the level of prosociality towards socially close others depends on gender and cognitive load (Soutschek et al., 2017; Strombach et al., 2016). We further showed that disrupting the temporoparietal junction (TPJ) – a brain region we recently identified as a central hub orchestrating the balance between egocentric and other-regarding preferences in social discounting (Strombach et al., 2015), and which is also associated with perspective taking (Tusche et al., 2016) and theory of mind (Saxe and Kanwisher, 2003) – by means of transcranial magnetic stimulation increases the steepness of social discounting (Soutschek et al., 2016), thus lowering the willingness to support socially remote strangers.

This body of evidence suggests that the degree by which individuals value socially close and distant others' well-being is highly malleable. However, despite its paramount theoretical and societal significance, means to *increase* the inclination for costly support of socially remote beneficiaries are elusive.

Here, we provide behavioral and neural evidence for a simple manipulation that aims at significantly increasing individuals' willingness to costly support socially remote others. We make use of the observation that people are more sensitive to others' losses than gains (Bardsley, 2008; Dreu, 1997; Evans and Beest, 2017; Everett et al., 2015; Li et al., 2017; Liu et al., 2020; Schweda et al., 2020; Sip et al., 2015; Smith et al., 2015; Wang et al., 2017; Xiao et al., 2016; Zheng et al., 2010), and are consequently strongly reluctant to increase their own payoff at the expense of others' welfare (Baumeister et al., 1994; Chang et al., 2011; Chang and Sanfey, 2013; Crockett et al., 2014; List, 2007). We hypothesized that participants would be more altruistic

---

[*] Corresponding author.
E-mail address: manuela.sellitto@hhu.de (M. Sellitto).

https://doi.org/10.1016/j.neuroimage.2021.118211.
Received 11 November 2020; Received in revised form 29 April 2021; Accepted 9 May 2021

towards others, including socially remote strangers, if a costly generous choice was framed as preventing a monetary loss to others rather than granting them a gain, even if actual economic outcomes were equivalent. In other words, we expected that the way a prosocial decision problem was framed mattered for the shape of the social discount function.

To test this hypothesis, we elicited social preferences in a standard version of the social discounting task (gain frame; Strombach et al., 2015) as well as in a loss frame variant. In each trial, participants decided to share money with other individuals on variable social distance levels. They chose between a selfish option, yielding high own-payoff and zero other-payoff, and a generous option, yielding a lower own-payoff and a non-zero other-payoff. The main difference between conditions was the way the decision problem was described: in the gain frame, a costly generous choice would yield an equivalent gain to the other player, while, in the loss frame, it would imply preventing the loss of a previous endowment to the other player. Importantly, the payoff structure was mathematically identical across frame conditions, i.e., the choice alternatives in the loss frame yielded identical own- and other-payoffs to those in the gain frame. Participants were explicitly instructed that the other persons would only be informed about the final outcome, but not about their endowment, or the loss of it; hence, they knew about the economic equivalence across frames.

We show in a series of independent studies that participants were more reluctant to make a selfish choice if this implied a loss of the endowment to the other, resulting in a substantially flatter social discount function in the loss than the gain frame and, hence, higher generosity towards socially remote others.

To obtain further insights into the psychological and neural mechanisms underlying this framing effect on social discounting, we measured blood oxygen level-dependent (BOLD) responses while participants performed both frame conditions of the social discounting task. We hypothesized that the psychological motives underlying generosity were frame-dependent and dissociable on the neural level. Consistent with our previous work (Strombach et al., 2015), we predicted that generosity in the gain frame was vicariously rewarding and the result of the resolution of the conflict between selfish and altruistic motives. Specifically, generosity in (Strombach et al., 2015) was associated with activity in TPJ, which suggested facilitation in overcoming the egoism bias via the modulation of value signals in the ventromedial prefrontal cortex (VMPFC), a brain structure known to represent own and vicarious reward value (Bartra et al., 2013; Mobbs et al., 2009), through the integration of other-regarding utility. In line with (Soutschek et al., 2016; Strombach et al., 2015), we therefore expected that generous choices in the gain frame would elicit activation of the VMPFC along with TPJ. Conversely, in the loss frame, we expected that the disinclination to maximize own-gain at the expense of other-loss was motivated by the desire to comply to social norms, such as the respect of others' property rights, or the do-no-harm principle. We therefore hypothesized increased activity in brain regions that are implicated in the social sentiments that motivate individuals to comply to social norms, such as the negative emotions experienced during social norm transgressions, e.g., guilt and shame, as well as the aversive experience of unfairness and inequality (Montague et al., 2007; Xiang et al., 2013). Such social sentiments have been consistently associated with the insular cortex (Chang et al., 2011; Chang and Sanfey, 2013; Civai et al., 2012; Corradi-Dell'Acqua et al., 2013; Gu et al., 2015; Lallement et al., 2013; Oldham et al., 2018; Samanez-Larkin et al., 2008; Spitzer et al., 2007; Tomasino et al., 2013; Von Siebenthal et al., 2017; Wang et al., 2017; Yu et al., 2014). Results support our main hypothesis that frame-dependent choice motives were associated with distinct neural signatures. During generous choice in the gain frame we found the involvement of VMPFC and TPJ (Hutcherson et al., 2015; Strombach et al., 2015), while we identified the insular cortex as the core component of a network associated with generous choice in the loss frame.

2. Material and methods

2.1. Participants

2.1.1. Studies 1–3

Three separate behavioral studies were carried out to test the validity of our paradigm in different settings and with different compensation procedures. For these studies we did not calculate the sample size in advance as we were not aware of any previous similar manipulation of social discounting. Study 1 was run online (n = 61; seven participants later excluded from the analyses due to bad fitting; 28 females; mean age = 36 years, ±11 standard deviation) and participants were paid a fixed allowance of €8.5. Study 2 was run online (n = 36; 32 females; mean age = 21 years, ±2.6) and participants, all psychology students on campus, were reimbursed for their time with a fixed amount of university credits. Study 3 (n = 39; eight participants later excluded from the analyses; 20 females; mean age = 26 years, ±6.0) was run in the laboratory and participants were paid out with the same fully incentive-compatible procedure as in the fMRI study 4 (see below). All three studies were conducted according to the Declaration of Helsinki and they were approved by the local ethics review board of the Heinrich-Heine University Düsseldorf. For studies 1 and 2 we did not collect informed consent, as this was allowed by the local ethics committee for online studies, which were fully anonymized, whereas we collected written informed consent in study 3, in the laboratory.

2.1.2. Study 4

After having replicated our results across the three behavioral studies with a within-subject design (see Results), for the fMRI study we estimated, via G*Power, assuming a medium-to-large effect size, that the sample size necessary to achieve a power of 0.95 was n = 23. Considering frequent participants' drop out during long scanning sessions as ours, or due to excessive movement, we opted for n = 40. Forty healthy young volunteers were therefore recruited at the Life&Brain Research Center in Bonn for an fMRI study. All participants met MR-compatible inclusion criteria in addition to no self-reported current or history of neurological or psychiatric disorder, as well as no current use of medication affecting the central nervous system. Due to excessive head motion during measurements (>4 mm, >4° rotation, as computed through Artrepair Toolbox; Stanford Psychiatric Neuroimaging Laboratory, see (Cho et al., 2013; Strombach et al., 2015; Wendelken et al., 2011), 10 participants were excluded from all analyses. Thus, the final sample included 30 subjects (21 females; mean age = 25 years, ±4.6, range: 19–35 years) with high-education level (mean education = 14 years, ±1.9, range: 12–18 years, from high school to university master degree). Fifteen participants had a net monthly income between €0 and €499, eight between €500 and €999, five between €1000 and €1499, one between €1599 and €1999, and one larger than €2500.

All participants were fluent German speakers, right-handed, and had normal or corrected-to-normal vision. As reimbursement, they were paid €20 as participation fee, plus earnings from the social discounting task. Therefore, participants' payoff ranged from €27.5 up to €35.5 (see Social discounting task).

The study was conducted according to the Declaration of Helsinki and it was approved by the local ethics review board of the Universitätsklinikum Bonn. All volunteers gave written informed consent to participate in the study.

2.2. Social discounting task

In this task (adapted from Strombach et al., 2015), participants were first asked to imagine people from their social environment represented on a scale ranging from 1 (the person socially closest to them) to 100 (a random stranger), where a person at rank 50 was described as a person that the subject had seen several times without knowing the name. They were instructed to select six real persons located at social distances of 1,

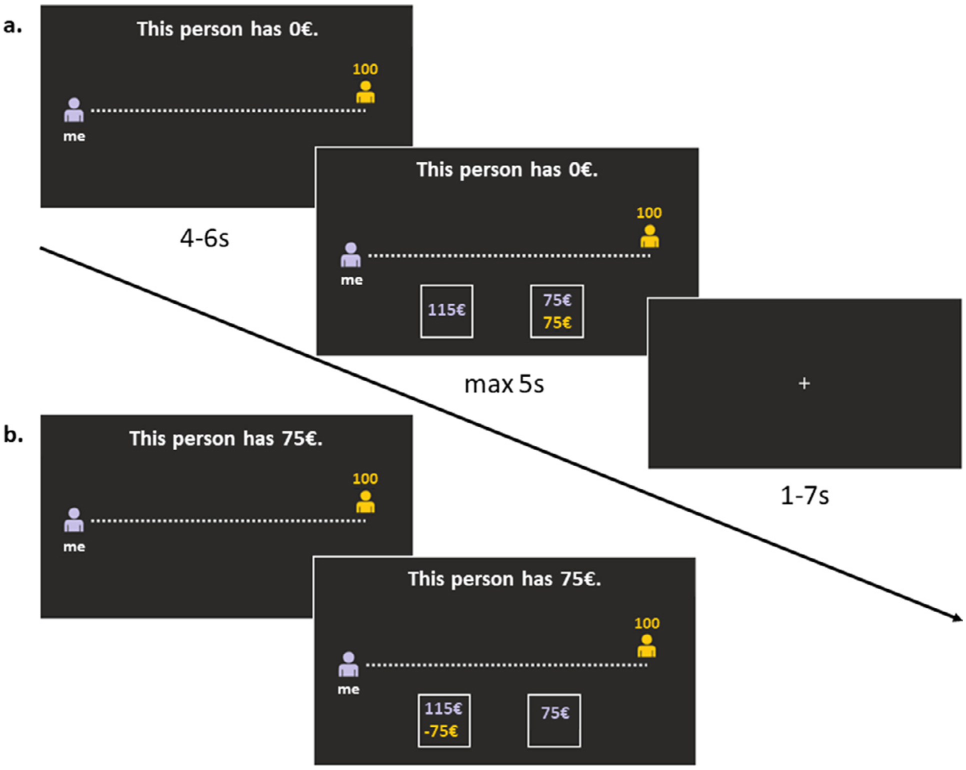


Fig. 1. Social discounting task (fMRI study 4). a. Trial example of the *gain frame*. **b.** Trial example of the *loss frame*. Each trial started with the presentation of a ruler-based representation of social distance to the other-person, with a left-most purple icon representing the participant and a yellow icon indicating the social distance of the other-person in the current trial (100 in this example). Additionally, participants received information as to the endowment of the other person, i.e., "This person has €0″ for the gain frame (**a**), or "This person has €75″ for the loss frame (**b**) (4–6 s). Afterwards, the two choice options appeared. The selfish alternative was displayed in purple fonts, indicating the own-reward magnitude to the participant (here: €115). Selfish choices implied a null gain for the partner in the gain frame (**a**) or the loss of the initial €75-endowment (in yellow) for the other person in the loss frame (**b**). The generous alternative was displayed in yellow fonts, and always yielded an equal €75 own-reward and €75 other-reward split in the gain frame (**a**), or a €75 own-reward gain and the possibility to keep the €75 other-endowment in the loss frame (**b**). As soon as the two choice options appeared, participants had 5 s to choose one of the two alternatives. After a choice was made, or after the 5 s had passed, a blank screen with a fixation cross appeared (1–7 s), and then a new trial started.

5, 10, 20, 50, and 100 (with no need of specifying the name and their social relationship for the social distances 50 and 100). Participants were encouraged to avoid thinking of people that they felt negatively toward and people they shared a bank account or household with. Each trial began with the display of the social distance level of the partner the participant was playing with. Social distance was represented with a ruler scale consisting of 101 icons. The left-most icon, highlighted in purple, depicted the participant. One of the remaining 100 other icons was highlighted in yellow, indicating the social distance of the partner. Furthermore, social distance information was additionally indicated as a number on top of the highlighted yellow icon to prevent perceptual inaccuracies in estimating social distance (cf. Fig. 1 for an example on a partner on social distance 100).

We included two experimental conditions, a *gain frame* and a *loss frame*. The *gain frame* manipulation was near-identical to the task used in (Strombach et al., 2015). Briefly, after presenting the social distance information as described above, participants were instructed that, in this trial, the experimenter gave an initial endowment of €0 to the other person ("This person has €0″; Fig. 1a). Participants were explicitly and repeatedly instructed that the other person was not aware of her zero endowment, she would only be informed of the final payoff after implementing the participant's choice. Then, two monetary options appeared, a selfish and a generous option. The selfish option (on the left in Fig. 1a, in purple letters) indicated the reward magnitude for the participant, if chosen (e.g., €115 to the participant and no other-reward to other). The generous option contained a smaller own-reward to the participant (€75) and an other-reward to the other person (€75) (on the right in Fig. 1a). Own-rewards were always indicated in purple and other-rewards were always indicated in yellow. Participants indicated their choice of the selfish or the generous alternative by a left or right button press.

In the *loss frame*, participants were informed, after the social distance presentation, that the other person has received an initial endowment of €75 ("This person has €75″; Fig. 1b). As before, participants were explicitly and repeatedly instructed that the other person was not aware of her initial endowment, or the potential loss of it. On the next screen, a selfish (on the left in Fig. 1b) and a generous alternative (on the right in Fig. 1b) appeared. When choosing the selfish alternative, the participant received the own-reward amount indicated in purple (here, €115), and the other person lost her initial endowment, as indicated in yellow (-€75), thus leaving her empty-handed. When choosing the generous alternative, the participant received a smaller own-reward indicated in purple (€75), implying that the other person would keep her endowment.

In addition to the framing (gain frame, loss frame) and the social distance levels of the other (1, 5, 10, 20, 50, 100), in each condition, we manipulated the magnitude of the own-reward across trials: we used nine selfish reward amounts per frame condition, ranging from €75 to €155 in steps of €10. The generous alternative's payoff was invariant, always yielding €75 own-reward and €75 other-reward in all conditions and trials.

Thus, in the gain frame condition, the other person always had a €0 endowment, the selfish alternative always yielded a variable own-reward and no reward for the other, and the generous alternative invariantly yielded an equal €75/€75 split between participant and other person. In the loss frame condition, the other-endowment was always €75, the selfish alternative yielded a variable own-reward accompanied by the loss of the €75 endowment to the other, and the generous alternative always yielded €75 own-reward and had no financial consequences for the other, i.e., she could keep her initial endowment of €75.

To summarize the logic of the task, both frames were mathematically equivalent, i.e., they yielded identical final payoff states to the participant and the other person (in the example in Fig. 1: both frames yield an own-reward gain of €115 to the participant and €0 gain to the other person after a selfish choice, or €75 own-reward and €75 other-reward after a generous choice). The only difference between conditions was that a €0 other-reward outcome was framed as a loss of the initial endowment in the *loss frame* vs. a null-gain in the *gain frame,* and a €75 other-reward was framed as keep-endowment in the *loss frame* vs. a €75 gain in the *gain frame*.

The order of frame conditions, selfish-reward presentations, as well as the left or right screen- position of the selfish and generous alternative were randomized and counterbalanced across trials. The task of studies 1–3 had a total of 108 self-paced trials (54 trials per each frame). The task of study 4 had a total of 216 trials as each trial type was repeated twice to allow for full left/right position counterbalancing. For events duration of study 4, please refer to Fig. 1.

2.2.1. Incentivization procedure

In studies 3 and 4 the social discounting task was fully incentive-compatible. At the end of the session, one of the participant's choices was randomly drawn and 10% of the own-reward amount was paid out, as well as, in case of a generous choice, 10% of the other-reward amount was paid out to the other person in that trial, either via cheque, for the other-persons indicated by the participants at social distance 1, 5, 10, 20, or in cash to a random person on site in the case of other-persons at social distance 50 and 100. Note that the recipients of other-reward were only notified in case of a positive payoff, but not in case of a zero payoff or in case a trial was randomly chosen that did not consider them; in addition, they were not informed beforehand about this experiment, and, thus, had no prior outcome expectations. Hence, our incentivization procedure made it logically impossible for the other persons to know about their endowment, or the loss of it.

2.3. General procedure

2.3.1. Studies 1–3

All participants performed a social discounting task and, at the end, they completed a questionnaire assessing social desirability (see Supplementary material and methods). Although the social discounting task of studies 1 and 2 was not incentivized, participants were strongly encouraged to think as if they were making decisions for real. In studies 1 and 2, participants were instructed about the social discounting task, and then, after answering comprehension questions, they assigned other persons (i.e., name and personal relationship with them) from their social environment to the social distances 1, 5, 10, and 20, completed the task (see Social discounting task), and finally filled out a questionnaire (see Supplementary material and methods) through Unipark online survey software (Unipark questback). Participants were provided with a web link to do so, after being recruited via flyers and advertisements on social platforms. Monetary payment for study 1 was implemented via Clickworker (GmbH), whereas university credits reimbursement was carried out on campus for study 2. In study 3, after recruitment, participants were invited to the laboratory, they were instructed on the social discounting task along with the comprehension questions. They then completed the task, implemented in Matlab R2016a (MathWorks) and Cogent toolbox (www.vislab.ucl.ac.uk), and filled out a questionnaire on a laptop. Finally, they were reimbursed for participation contingent on their choices, identical to the incentivization procedure in study 4.

2.3.2. Study 4

Upon arrival, participants received instructions about the social discounting task and then, after applying comprehension questions to check for full understanding of the task, they assigned other persons (i.e., name and personal relationship with them) from their social environment to the social distances 1, 5, 10, and 20 via paper and pencil. Afterwards, participants performed a few sample trials to familiarize with the task structure and they were subsequently cleared for the scanning session. At the end of the scanning session, they answered control questions concerning the social discounting task, and filled out a demographic questionnaire as well as questionnaires assessing social desirability and empathy (see Supplementary material and methods). Finally, participants were debriefed and received their monetary allowance.

2.4. Studies 1–4

2.4.1. Behavioral data analysis

Hyperbolic model: Similar to previous studies (Jones and Rachlin, 2006; Margittai et al., 2015, 2018; Soutschek et al., 2016; Strombach et al., 2015), we approximated the participants' decay in generosity across social distance with a hyperbolic function:

$$v = \frac{V}{(1 + k * SD)} \tag{1}$$

where v represents the discounted value of generosity, SD represents social distance, k represents the degree of discounting, and V is the intercept at social distance 0, thus the origin of the social discount function (Jones and Rachlin, 2006; Margittai et al., 2015; Soutschek et al., 2016; Strombach et al., 2015). While V can be considered an indicator of generosity towards socially close others (Margittai et al., 2015, 2018; Strang et al., 2017), k describes the discount rate, i.e., the steepness by which the social discount function decays across social distance. We estimated k and V for each participant separately, depending on her individual choice pattern.

To estimate V, we titrated the selfish amount to determine, at each social distance, the point at which the subject was indifferent between the selfish and generous options (i.e., indifference point; see Supplementary results). Logistic regression, implemented in Matlab R2016a (MathWorks), was used to determine the indifference points where the likelihood of choosing the selfish and the generous options was 50% (Soutschek et al., 2016; Strombach et al., 2015). Across the four studies, V ranged between 10 and 99, 95% CI [76, 82] for the gain frame, and between 10 and 98, 95% CI [73, 79] for the loss frame. Across the four studies, the median R^2 of the estimated V parameters equalled 0.99, range = 1 for the gain frame, and 0.93, range = 1 for the loss frame.

To fit Eq. (1) and estimate k, we modeled trial-by-trial choices via a softmax function to compute the probability P of choosing the selected option o_i over the other option o_{ii} on a given trial:

$$Po_i = \frac{1}{1 + \exp\left(-1 * m * \left(vo_i - vo_{ii}\right)\right)} \tag{2}$$

given the subjective values v (based on the current selfish amount and social distance) of the current available options o_1 (vo_1) and o_2 (vo_2) as in Eq. (1). The nuisance parameter m reflects the stochasticity of individual performance. The larger the m, the less noisy the choice pattern. Individual discount rates were defined by the respective k value that yielded the best prediction of the observed choice probabilities by applying maximum-likelihood estimation using nonlinear optimization procedures (fminsearch function), implemented in Matlab R2016a (MathWorks). To this end, we minimized the log-likelihood of the choice probabilities to obtain the best-fitting k and m parameter estimates, by summing across trials, given a specific set of model parameters k and m, the logarithm of $P(o_i)$. Across the four studies, k ranged between 5E-12 and 0.69, 95% CI [0.03, 0.06] for the gain frame, and between 5E-13 and 0.48, 95% CI [0.01, 0.03] for the loss frame. Across the four studies, the median log-likelihood of estimated k parameters equalled −21, range = 60 for the gain frame, and −20, range = 58 for the loss frame.

We additionally performed *parameter recovery simulation* to check that the fitting procedure had generated meaningful parameter values. Based on the procedures described in Wilson and Collins (2019), we used obtained individual discount parameters k and their respective noise parameter m to create synthetic participants, computed 10 simulations of responses of these synthetic participants, fitted the simulated data with our model (see Eqs. (1) and (2)), and compared the mean values of the obtained recovered parameters from the simulations against the inputted parameters of all four studies collapsed (see also Studer et al., 2019). Participants with null discounting (no variance in choice) were excluded from the sample as model parameters could not be estimated for them. Four simulations (out of a total of 2440) were excluded as they led to a k value of the order of E+13. The parameter recovery simulations showed adequate recovery of the k parameters, with a Pearson correlation between inputted and mean recovered k parameter estimates of $r = 0.95$, $p < 0.001$, Cohen's $d = 6.08$ for the gain frame and of $r = 0.96$, $p < 0.001$, Cohen's $d = 6.86$ for the loss frame (see Fig. S1a,b). Moreover, we correlated the difference in k values between gain and loss frame for each real participant with the difference in k values between gain and loss frame for their respective average simulation. This correlation was $r = 0.89$, $p < 0.001$, Cohen's $d = 3.90$, a result that is corroborated by comparing simulated k parameters for the gain and the loss frame: once again, k values for the loss frame were significantly smaller than k values for the gain frame (median $\mathrm{k}_{\mathrm{gain}} = 0.02$, range $\mathrm{k}_{\mathrm{gain}} = 0.86$ vs. median $\mathrm{k}_{\mathrm{loss}} = 0.008$, range $\mathrm{k}_{\mathrm{loss}} = 0.52$; Wilcoxon matched-pair test: $Z = 7.10$, $p < 0.001$, $r = 0.70$), matching our results. Additionally, the recovery of the noise parameter m led to a correlation of $r = 0.66$, $p < 0.001$, Cohen's $d = 1.76$ for the gain frame and of $r = 0.41$, $p < 0.001$, Cohen's $d = 0.90$ for the loss frame. Note that the recovery was compromised by low-noise participants (with high m values) because their noise parameters were likely overestimated. Excluding those participants (9 out of 140 for the gain frame and 7 out of 105 for the loss frame) improved the recovery of the noise parameter m to $r = 0.95$, $p < 0.001$ for both the gain and the loss frame, without altering the recovery of the discount parameter k ($r = 0.95$, $p < 0.001$ for the gain frame and of $r = 0.96$, $p < 0.001$ for the loss frame) (see Fig. S1c,d; please note that the excluded participants are not reported in these two figures as they were masking the representation of the rest of the data). Across the four studies collapsed, the median log-likelihood of recovered k parameters was −20, range = 56 for the gain frame, and −19, range = 57 for the loss frame. Thus, in summary, our simulations showed reliable and adequate recovery of the hyperbolic discount parameter k and the noise parameter m.

Both estimated variables V and k were analyzed using non-parametric statistics as they were, in most of the cases across the four studies, not normally distributed even after log-transformation (Kolmogorov-Smirnov, all ds > 0.20, all ps < 0.05). When participants did not discount at all (i.e., they always chose the generous or the selfish option), k was set to 0 and V was set to 80 (i.e., maximum reward amount foregone = maximum selfish amount 155 − generous amount 75) for all generous choices or to 0 for all selfish choices. All behavioral analyses were run in Statistica 12 (StatSoft). For additional analyses (i.e., indifference points, area under the curve, reaction times, and questionnaires) please see the Supplementary information.

2.5. Study 4

2.5.1. fMRI procedures

Magnetic resonance images were collected on a 3T whole-body scanner (Magnetom Trio, Siemens Medical Systems, Erlangen, Germany) with an 8-channel head coil. For functional imaging, gradient-echo echo-planar images (EPI) were acquired at TR = 2500 ms (TE = 30 ms; number of slices = 37; slice thickness = 3 mm; distance factor = 10%; FoV = 192 mm × 192 mm; matrix size = 64 × 64; flip angle = 90°). Slices (voxel size = 2 × 2 × 3 mm) were sampled in transversal orientation covering all of the brain, including the midbrain. The scanning session started with a brief localizer acquisition. Afterwards, functional data were acquired in 3 separate runs of ∼421 volumes each, to allow for brief resting periods in between. In order to get information for B_0 distortion correction of the acquired EPI images, a gradient echo field map sequence (TR = 392 ms; TE 1 = 4.92 ms; TE 2 = 7.38 ms; number of slices = 37; voxel size 3 × 3 × 3 mm) was recorded before each functional run. Structural images were collected at the end (∼5 min), using a T1-weighted sequence (rapid acquisition gradient echo sequence; 208 sagittal images; voxel size = 0.8 × 0.8 × 0.8 mm; 0.8 mm slice thickness).

Head movements were minimized by the use of foam pads and scanner noise was reduced with earplugs. When necessary, vision was corrected-to-normal via fMRI compatible goggles. The social discounting task was programmed via an in-house software and presented via a mirror that projected a screen lying behind the participant, who made their choices via a left and a right button boxes.

2.5.2. fMRI preprocessing

Imaging data were preprocessed and analyzed with Statistical Parametric Mapping (SPM12, Wellcome Trust Centre for Neuroimaging, University College London, UK) implemented in Matlab R2016a (MathWorks). After checking raw data quality for each participant using the SPM Check Reg function (Stanford Psychiatric Neuroimaging Laboratory), all images were preprocessed by reorienting them according to the EPI SPM template and coregistered to the fieldmap via FieldMap toolbox. After the functional images were realigned und unwarped to the middle volume and all volumes for participants' motion correction by using phase correction, ArtRepair toolbox (Mazaika et al., 2009) was run in order to identify bad volumes. Bad volumes of participants included in the final sample were not repaired. However, we modeled these bad volumes as regressors of no-interest in the statistical analyses (see fMRI analyses). Finally, functional and structural images were coregistered and the images were spatially normalized based on segmentation of the anatomical image with resampling to 2 × 2 × 2 mm, and spatially smoothed using a 6 mm FWHM Gaussian kernel. High-pass temporal filtering (using a filter width of 128 s) was also applied to the data.

2.5.3. fMRI analyses

At the first-level analysis, trial-related activity for each participant was modeled by delta functions convolved with a canonical hemodynamic response function to model the effects of interest, as well as six covariates capturing residual motion-related artifacts, and a temporal derivative for each regressor of interest to account for slice timing differences.

For each participant, relevant contrasts were computed for each general linear model (GLM) (see below for details) and entered into second-level random effect analysis. The following variables were considered in the analyses: the loss frame condition; the gain frame condition; generous choices; selfish choices. Comparisons were run via one-way Analyses of Variance (ANOVAs), within subject, and via one-sample t-tests, where appropriate.

GLM1 searched for differences in BOLD activations between frame conditions during generous choices, where the onset of a generous choice was defined as the participant's button press to choose the generous option after the monetary options had appeared on the screen (see Fig. 1). It included an unmodulated regressor of all generous choices made in the loss frame condition and an unmodulated regressor of all generous choices made in the gain frame condition. Additionally, the selfish amount magnitude (see Social discounting task) was included as trial-by-trial parametric modulator of all main regressors, separately. In the main manuscript, we additionally considered the reward foregone as a parametric trial-by-trial regressor. Note that the reward foregone is a linear transformation of, and thus collinear with, the selfish reward magnitude; neural activations identified by this parametric regressor might therefore reflect selfish amount or reward foregone (see main text). Reaction times (RTs) were used as duration to account for differences between gain and loss frames (see Supplementary behavioral results). Additionally, missed trials were included as regressors of no-interest and modeled with duration = 5 s, i.e., the maximum time allowed to respond.

Please note that at the level of choice, where the choice onset was defined as the participant's button press after the release of the monetary options at each trial, a full model including separate regressors for both frames (gain and loss) and both types of choice (selfish and generous), as well as the trial-by-trial selfish amount as parametric regressor, was possible only for sixteen participants. This was due to participants who had to be excluded because they never, or only very rarely (not in all experimental runs) chose the selfish alternative in the loss frame. To address potential statistical power concerns associated with small sample size and to attend to potential selective sampling biases, in addition to generous choice being our main focus, we ran instead the above-mentioned model. Nevertheless, results of this full model (***GLMS1***), as well as of a model including only selfish choices (***GLMS2***), are reported in the Supplementary information for completeness.

GLM2 tested for the effect of frame condition, and therefore included an unmodulated regressor of the onsets of the loss frame condition and an unmodulated regressor of the onsets of the gain frame condition. The frame onset was defined as the trial start (see Fig. 1). The social distance was included as trial-by-trial parametric modulator of the frame onsets, separately for the gain and the loss frames. A stick function was used as duration.

To address potential statistical concerns relative to having modelled, separately, the two main events of our task (i.e. frame onset and participants' response onset), we ran an additional analysis (***GLMS3***) including all main regressors and all parametric regressors of both GLM1 and GLM2. We replicated results of both models, indicating that the frame regressors do not compete for variance with the choice regressors. Thus, the anterior insula activation has been correctly attributed to generous choice in the loss frame, making our original interpretation plausible (see Supplementary information for analysis details and results).

All whole-brain level results as well as ROI-based (see below) results were initially thresholded at $p < 0.001$ (uncorrected), minimum cluster size = 5 voxels, and then corrected at the cluster level for multiple comparisons ($p < 0.05$, family-wise error rate [FWE]). Bad volume onsets (as measured via ArtRepair toolbox; (Mazaika et al., 2009)), modeled with a stick function, were included as regressors of no-interest in all the above GLMs.

We additionally conducted, where relevant (see main text) ROI analyses for VMPFC, TPJ, and insular cortex by using anatomical bilateral masks from the Harvard-Oxford Atlas (Jenkinson et al., 2012), and the SPM Anatomical Automatic Labeling Toolbox, version 3 (Rolls et al., 2020), via SPM12 in Matlab R2016a. The probability maps of the SPM Anatomy Toolbox, version 3.0 (Eickhoff et al., 2007), and Neurosynth (http://neurosynth.org) were used for double checking region localization throughout GLMs.

Dynamic causal modeling (DCM): We used DCM analysis as implemented in SPM12. This analysis focused on the interplay between insula and VMPFC and between TPJ and VMPFC, addressing both (i) regions endogenous connectivity and (ii) condition specific modulation of the regions (driving inputs) and their connections (modulatory inputs). We therefore constructed a hierarchical model with regressors defining both frame conditions activations against the total baseline activation. Thus, we entered in the DCM: a regressor of no-interest for baseline connectivity ('all trials', used to correct for global activation) including onsets of the screen presenting the framing information and the social distance, and the onsets of the screen presenting the monetary options, at all trials; a regressor ('all loss trials') including onsets of the screen presenting the framing information and the social distance, and the onsets of the screen presenting the monetary options, for the loss frame trials; a regressor ('all gain trials') including onsets of the screen presenting the framing information and the social distance, and the onsets of the screen presenting the monetary options, for the gain frame trials.

Subject-specific coordinates were guided by ROI-based group activation maxima in the three network regions from the univariate, group-level results (see Results section). Volumes of interest (VOI) spheres, with a radius of 6 mm, were built around the posterior insula (GLM2, [34, −16, 8]), rTPJ (GLM1, [50, −66, 36]), and VMPFC (GLM2, [2, 50, −8]). Note that we focused our DCM analysis on the posterior insula cluster only, as we were interested in a baseline frame activation; including the anterior insula cluster, specific for generous choice within the loss frame (see Results section), might have biased the results in favor of our hypotheses. Also note that we found TPJ in the choice-related analysis (GLM1) only: in our opinion, it was still preferable to opt for this experimentally driven ROI rather than including an ROI taken from the literature. Regional time series were extracted as the first eigenvariate of the three network regions for 'all trials' and mean-corrected for the effect of interest F-contrast at a liberal threshold of $p = 0.1$. This threshold was lowered for some participants until all regions could be detected (Zeidman et al., 2019a, 2019b).

Based on our univariate results, we constructed bilinear models where the endogenous connectivity across the three regions was always assumed. We specified models with nodes reciprocally connected, where the gain and loss frame were allowed to modulate all connections (Li et al., 2015). The resulting 15 models were grouped in two families: A and B. In family A, both condition-specific driving inputs and condition-specific modulatory inputs were assumed. In family B, only condition-specific driving inputs were assumed.

Family A included eleven models (Fig. 2). In model *1* (sum of the log-evidence SF = −4.0797E+05, exceedance probability xp = 0.1282), we assumed that the gain frame condition had direct inputs on VMPFC and TPJ, and a modulatory input on their connections; the loss frame condition had direct inputs on VMPFC and insula, and a modulatory input on their connections. In model *2* (SF = −4.0864E+05, xp = 0.1294), the gain frame had a driving input on VMPFC, and a modulatory input on its connectivity with TPJ; the loss frame had a driving input on VMPFC and a modulatory input on its connectivity with the insula. In model *3* (SF = −4.0829E+05, xp = 0.005), the gain frame had a driving input on TPJ and a modulatory input on its connectivity with VMPFC; the loss frame had a driving input on the insula and a modulatory input on its connectivity with VMPFC. In model *4* (SF = −4.0826E+05, xp = 0.049), the gain frame had a driving input on VMPFC and a modulatory input on its connectivity with TPJ; the loss frame had driving input on the insula and a modulatory input on its connectivity with VMPFC. In model *5* (SF = −4.0786E+05, xp = 0.6595), the gain frame had a driving input on TPJ and a modulatory input on its connectivity with VMPFC; the loss frame had a driving input on VMPFC and a modulatory input on its connectivity with the insula. Therefore, connectivity between regions in model 1 to 5 is assumed to be bidirectional. Additionally, in model *6* (SF = −4.0892E+05, xp = 0.0048), the gain frame had a driving input on VMPFC and a modulatory input on its connectivity to TPJ; the loss frame had a driving input on VMPFC and a modulatory input on its connectivity to the insula. In model *7* (SF = −4.0953E+05, xp = 0.0009), the gain frame had a driving input on VMPFC and a modulatory input on the connectivity from TPJ to VMPFC; the loss frame had a driving input on VMPFC and a modulatory input on the connectivity from the insula to VMPFC. In model *8* (SF = −4.0867E+05, xp = 0), the gain frame had a driving input on TPJ and a modulatory input on its connectivity to VMPFC; the loss frame had a driving input on the insula and a modulatory input on its connectivity to VMPFC. In model *9* (SF = −4.0890E+05, xp = 0.001), the gain frame had a driving input on TPJ and a modulatory input on the connectivity from VMPFC to TPJ; the loss frame had a driving input on the insula and a modulatory input on the connectivity from VMPFC to the insula. In model *10* (SF = −4.0861E+05, xp = 0), the gain frame had a driving input on TPJ and a modulatory input on the connectivity from VMPFC to TPJ; the loss frame had a driving input on VMPFC and a modulatory input on its connectivity to the insula. In model *11* (SF = −4.0851E+05, xp = 0.0221), the gain frame had a driving input on TPJ and a modulatory input on its connectivity to VMPFC; the loss frame had a driving input on VMPFC and a modulatory input on the connectivity from the insula to VMPFC.

Family B included four models (Fig. 2). In model *12* (SF = −4.0886E+05, xp = 0), the gain frame had driving input on VMPFC and TPJ; the loss frame had driving inputs on VMPFC and the insula. In model *13* (SF = −4.0936E+05, xp = 0), the gain frame had a driving input on TPJ and the loss frame had a driving input on insula. In model *14* (SF = −4.0999E+05, xp = 0.0001), both frame conditions' driving inputs were on VMPFC. In model *15* (SF = −4.0893E+05, xp = 0), the gain and the loss frame had driving inputs on all three regions, insula, TPJ, and VMPC, to check whether at increased number of connections, the model fitted the data better.

All the hypothesized models were entered into Bayesian Model Selection (BMS), as implemented in SPM, to determine the best-fit family and model. The inference method used to compare the models across

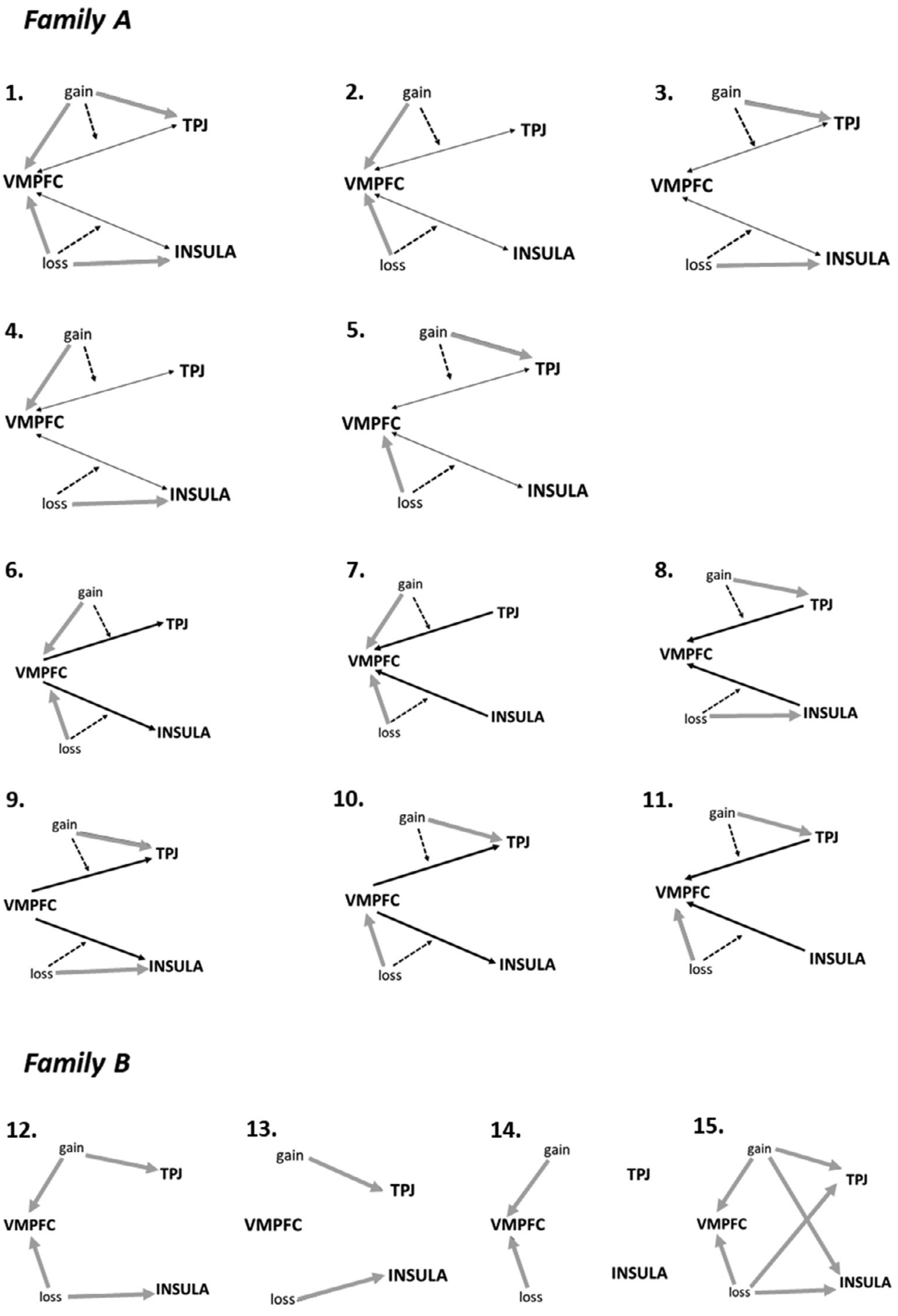


Fig. 2. DCM models. Fifteen models were hypothesized to describe the data. Gray lines represent driving inputs. Dashed black lines represent modulatory inputs. Thin black lines represent bidirectional connectivity. Thick black lines represent unidirectional connectivity. Since endogenous connectivity is always assumed between all three regions in all models, it is not represented here. *Family A*, which assumed both condition-specific driving inputs and condition-specific modulatory inputs, includes models 1 to 11. *Family B*, which assumed only condition-specific driving inputs, includes models 12 to 15.

subjects and session was random effects (2nd-level, RFX). Bayesian Model Averaging (BMA) was used for model comparison. Once the optimal model was selected, the participant-specific parameters for the two frame conditions were averaged across the three runs and entered into group analysis with one-sample and paired-sample t-tests, where appropriate. This allowed us to summarize the consistent findings from the subject-specific DCMs using classical statistics (Cho et al., 2013; Li et al., 2015; Neufang et al., 2016; Wiehler et al., 2017; Zhang et al., 2018).

Mediation analysis: This analysis was run via Hayes's PROCESS-macro (Hayes, 2017) as implemented in the IBM Statistical Package for the Social Sciences (SPSS). The analysis aimed at testing the idea that the gain and the loss frame had an effect on generous behavior through the mediating influence of condition-specific neural activations. The frame condition was included as binary independent variable X (dummy variable: 1 = gain; 2 = loss), the proportion of generous choices (gain frame and loss frame) was entered as dependent variable Y, and the neural activations were entered as mediators. Specifically, beta estimates for the posterior insula [34, −16, 8; GLM2], VMPFC1 [2, 50, −8; GLM2], the anterior insula [42, 4, −4; GLM1], TPJ [50, −66, 36; GLM1], and VMPFC2 [0, 54, 14; GLM1] were extracted, at the single-subject level, for both frames and included in the model, via MarsBaR region of interest toolbox for SPM12 (Brett et al., 2002). Neural activations across both frames were treated as parallel mediators (model template 4, Hayes, 2017). Partially standardized values are reported, and 95% biased-corrected CIs are adopted. Number of bootstrap samples was set to 5000. To determine the statistical power for mediation, the online tool MedPower was used (https://davidakenny.shinyapps.io/MedPower/) using effects of X on mediator (M) (path a), of M on Y (path b), and the direct effect of X on Y (path c′), at alpha=0.05. Total achieved power was ∼0.60.

3. Results

3.1. Social discounting is flatter in the loss than the gain frame

First, in a set of behavioral experiments, we established that our framing manipulation affected generosity towards socially distant others. In a within-subject design, we elicited social preferences in a standard version of the social discounting task (the gain frame; Strombach et al., 2015) as well as in a loss frame variant (see Fig. 1), interleaved in a trial-by-trial fashion. In the gain frame, participants played with other persons at variable social distance levels, and made choices between a selfish alternative, yielding higher monetary payoff to the participant and zero payoff to the other, and a generous alternative, always yielding a lower own-payoff of €75 along with a payoff of €75 to the other. In the loss frame, participants were first informed that the other person had received an initial endowment of €75. The selfish alternative yielded a variable, higher own-payoff as well as the loss of the €75 endowment to the other person, hence, resulting in a zero payoff to the other; the generous alternative yielded a fixed €75 payoff to the participant, and no further consequence to the other, thus leaving her with the initial €75 endowment. Crucially, the payoff structure was mathematically equivalent across both frame conditions, i.e., the choice alternatives in the loss frame yielded identical own- and other-payoffs to those in the gain frame. The main difference between conditions was that, in the gain frame, a generous choice would imply a gain of €75 to the other, while in the loss frame, a generous choice would imply preventing the loss of the previous €75 endowment. Importantly, participants were repeatedly instructed that the other person was unaware of her initial endowment, or the loss of it, and that she would only be informed about the final outcome of the payoff after implementing the participant's choice at the end of the experiment. Task comprehension, in particular regarding participants' understanding that the other person would only be informed about the final outcome, but not about her endowment, or loss of it, was further stressed during the explanation of the incentivization procedure as well as assessed in post-hoc structured interviews (see Material and methods). All participants understood the task well.

3.1.1. Study 1

In a first study, data collection was done online and the task was not incentive-compatible; participants (n = 54) were paid a fixed allowance of €8.5. In the social discounting task, participants can either make a selfish choice or a generous choice, in each frame condition. We adopted the hyperbolic discount model to describe the effect of framing on our participants' behavior as it is the best-documented model to investigate social discounting, with demonstrated better goodness-of-fits in comparison to other, e.g., exponential, models (e.g. Jones and Rachlin, 2006). Specifically, to reconstruct the individual social discount functions, separately for the two frame conditions, we fit a standard hyperbolic model (see Eq. (1) (Jones and Rachlin, 2006; Strombach et al., 2015); Material and methods) to trial-by-trial binary choices (i.e., either selfish or generous) to estimate the parameter k, a measure of the steepness of the social discount function. Additionally, we determined, for each participant and each social distance level, and separately for the two frame conditions, the point at which the participant was indifferent between the selfish and the generous alternative using logistic regression (Strombach et al., 2015). The difference in reward magnitudes for the participant between the two alternatives at the indifference points (see Supplementary results) represented the amount of money a subject was willing to forego (i.e., reward amount foregone) to increase the wealth of another person at a given social distance, and could be construed as a social premium, that is, the price tag participants put on increasing the wealth of the other. We took the estimated parameter V, the intercept at social distance 0, thus the origin of the social discount function (Jones and Rachlin, 2006; Margittai et al., 2015, 2018; Soutschek et al., 2016; Strombach et al., 2015), as an indicator of generosity towards socially close others (Soutschek et al., 2016; Strombach et al., 2015).

Participants' generosity dropped much less steeply in the loss compared to the gain frame (median k_{gain} = 0.022, range k_{gain} = 0.18 vs. median k_{loss} = 4.74E-11, range k_{loss} = 0.10; Wilcoxon matched pairs test: Z = 4.97, p < 0.001; r = 0.68; see supplementary **Fig. S2a**). The difference in social discount functions between frames was most pronounced at high social distance levels, indicating that participants were substantially more generous towards strangers in the loss than the gain frame. We also found a significant difference in V between frame conditions (median V_{gain} = 87, range V_{gain} = 88 vs. median V_{loss} = 77, range V_{loss} = 68; Z = 2.78, p < 0.01; r = 0.38) that, however, disappeared when removing all participants with zero discounting from the analysis (see Supplementary results). These data suggest that participants were strongly more generous towards socially distant others in the loss than the gain frame.

3.1.2. Study 2

In a second study, we replicated the results of our first experiment. Data collection was done online and participants (n = 36) were reimbursed for their time with a fixed amount of university credits. We again found that participants had flatter social discounting in the loss than the gain frame (median k_{gain} = 0.020, range = 0.62 vs. median k_{loss} = 0.0005, range = 0.10; Z = 4.81, p < 0.001; r = 0.80; see supplementary **Fig. S2b**), and we found no difference in V between frame conditions (median V_{gain} = 81, range V_{gain} = 66 vs. median V_{loss} = 80, range V_{loss} = 50; Z = 0.38, p = 0.71). Again, these results held when excluding participants with null discounting.

3.1.3. Study 3

Studies 1 and 2 were not incentive-compatible. To determine whether hypothetical versus real payoffs made a difference in the frame effect on social discounting (Vlaev, 2012), we ran a third fully incentive-compatible study in a laboratory setting (n = 31). Payoff was contingent on the participants' choices, and was paid out to self and other, identical to (Strombach et al., 2015) and to the fMRI study 4 (see next paragraph and Material and methods). Once again, we could replicate the frame effect on k (median k_{gain} = 0.022, range k_{gain} = 0.69 vs. median k_{loss} = 4E-07, range k_{loss} = 0.48; Z = 3.71, p < 0.001; r = 0.67; see supplementary **Fig. S2c**), and there was no difference in the V parameter between frame conditions (median V_{gain} = 81, range V_{gain} = 99 vs. median V_{loss} = 80, range V_{loss} = 88; Z = 1.41, p = 0.16). These results held when excluding participants with null discounting.

Additionally, we plot for all three studies the proportion of generous choices, averaged across participants, as a function of the selfish amount to highlight the flatter decay in generous choices in the loss frame compared with the gain frame, especially at remote social distances (**Fig. S3a,b,c**). Furthermore, the distributions of individual k value differences and V value differences between frames are shown in **Fig. S4a,b,c** for all three studies.

Moreover, social desirability, as measured via the Social Desirability Scale (SDS-17; Stöber, 2001), did not explain the frame effect on social discounting parameters (see Supplementary material).

The result of increased generosity, especially at larger social distances, in the loss compared to the gain frame in the three behavioral studies was also corroborated via a model-free measure, i.e. the area under the curve (AUC), as well as via an analysis of the indifference points (the selfish reward magnitude at which participants were indifferent between the selfish and the generous alternative at each social distance level and in each frame condition; see Supplementary analyses and results).

Overall these results suggest that, while generosity to socially close others was comparable between frame conditions, it decayed significantly less steeply across social distance in the loss than in the gain frame, indicating that participants were considerably more generous towards socially distant others in the loss frame.

3.2. Neural mechanisms underlying the frame effect on social discounting

To obtain more substantial insights into the psychological and neural mechanisms underlying this framing effect on social discounting, in study 4 we measured BOLD responses while participants performed both frame variants of the social discounting task. The fundamental premise of our study is that the decision motives and their neural correlates differ between gain and loss frames. Specifically, we reasoned that generosity in the gain frame was mainly stimulated by other-regarding considerations. Conversely, we predicted that generous decisions in the loss frame were motivated by the desire to comply to social norms, such as the *do-no-harm* principle, or the respect of others' property rights (Sethi et al., 1996), infringements of which are associated with negative social sentiments of guilt and shame. To test this idea, we focused on two main hypotheses. We, first, expected that generosity in the gain frame recruited a network of structures, including VMPFC and TPJ (Hutcherson et al., 2015; Strombach et al., 2015), known to represent vicarious reward value and prosocial behavior. Second, we hypothesized that brain areas implicated in negative social sentiments of social norm transgressions, such as the insular cortex (e.g. Paulus et al., 2003; Chang et al., 2011; Chang and Sanfey 2013; Lallement et al., 2013; Gu et al., 2015; Seara-Cardoso et al., 2016; Sethi and Somanathan 2016; Siebenthal et al. 2017; Wang et al., 2017; Huggins et al., 2018), would be selectively recruited during generous choices in the loss, but not the gain frame.

We first replicated, once more, the behavioral framing effect on social discounting ($n = 30$). As before, the drop in generosity across social distance was pronouncedly flatter in the loss than the gain frame (median $k_{gain} = 0.021$, range $k_{gain} = 0.16$ vs. median $k_{loss} = 0.003$, range $k_{loss} = 0.28$; Wilcoxon matched pairs test: $Z = 3.69$, $p < 0.001$; $r = 0.67$; Fig. 3a), but, again, there was no difference in the V parameter between conditions (median $V_{gain} = 80$, range $V_{gain} = 57$ vs. median $V_{loss} = 80$, range $V_{loss} = 53$; $Z = 0.88$, $p > 0.37$; the results remained identical when excluding participants with null discounting). Additionally, we plot the proportion of generous choices, averaged across participants, as a function of the selfish amount (Fig. 3b) to illustrate the flatter decay in generous choices in the loss compared to the gain frame, especially at remote social distances. Furthermore, the individual distributions of *k* value differences and V value differences between frames are shown in **Fig. S4d**, as well as we plot the choice probability as a function of the difference in value between the generous and the selfish option (**Fig. S5**).

Moreover, neither social desirability (SDS-17; Stöber 2001), nor perspective taking, empathic concern, personal distress, or fantasy (as measured via the Interpersonal Reactivity Index; IRI; (Davis, 1983; Paulus, 2009) explained the frame effect on social discounting parameters (see Supplementary material).

The above results were corroborated, once again, also via an analysis of the AUC, as well as via an analysis of the indifference points at each social distance (see Supplementary analyses and results).

Our first hypothesis predicted activity in brain structures known to represent vicarious reward value and prosocial behavior in the gain frame (similar to Soutschek et al., 2016; Strombach et al., 2015). Our results (**GLM1**; see Material and methods) indeed revealed clusters located in VMPFC (0, 54, 14, whole-brain $p_{FWE\text{-}corr} < 0.001$) as well as right TPJ (rTPJ; 50, −66, 36, whole-brain $p_{FWE\text{-}corr} < 0.035$) to be selectively activated, in addition to other prefrontal regions, when participants made generous choices in the gain frame relative to generous choices in the loss frame. ROI analyses confirmed significant clusters of activation in both VMPFC ($p_{FWE\text{-}corr} < 0.001$) and rTPJ ($p_{FWE\text{-}corr} = 0.01$). Thus, consistent with (Hutcherson et al., 2015; Strombach et al., 2015), a network comprising VMPFC and rTPJ seems to underlie the motivation for costly generosity in the gain frame (Fig. 4; see supplementary **Table S1**). Additionally, the selfish amount magnitude, included as trial-by-trial regressor, did not parametrically modulate activity in VMPFC and rTPJ (GLM1; see Material and methods).

Our second hypothesis predicted that generosity in the loss frame was motivated by social norm compliance rather than other-regarding considerations; generosity should, consequently, go along with a different neural activation pattern in the loss than the gain frame. In a first step, we attempted to isolate frame-dependent neural correlates, independent of participants' choices. To this end, we searched for differential neural activity at trial onset, i.e., when participants learned about the social distance level of the other person and which frame was relevant in the current trial (see Fig. 1), by contrasting neural activity between the two frames (**GLM2**; see Material and methods). We found significant activation in the right posterior insula (34, −16, 8, whole-brain $p_{FWE\text{-}corr} = 0.007$) in the loss vs. gain frame contrast, which was accompanied by significant activations in frontal regions, including VMPFC (2, 50, −8, whole-brain $p_{FWE\text{-}corr} = 0.001$), as well as temporal regions (Fig. 5; see supplementary **Table S2** for a complete list of activations). ROI analyses confirmed significant clusters of activation in the right insula ($p_{FWE\text{-}corr} = 0.03$) as well as in VMPFC ($p_{FWE\text{-}corr} = 0.02$). The opposite contrast, gain frame vs. loss frame, did not reveal any significant activation. Social distance information, included as trial-by-trial regressor, did not parametrically modulate neural activity in any of these contrasts (GLM2; see Material and methods), suggesting that the activations in insula and VMPFC reflected frame but not social distance information.

In support of this conclusion, we found that the right anterior insula (42, 4, −4, ROI analysis, $p_{FWE\text{-}corr} < 0.02$; **GLM1**, see Material and methods), was selectively activated during generous choices in the loss frame relative to generous choices in the gain frame (Fig. 6; see supplementary **Table S1**). The location within the insula mask was slightly anterior to the peak activation we found at trial onset.

Our analysis so far suggests that insula activation reflects the psychological motives underlying generous choice in the loss frame. However, other explanations of our insula finding are conceivable, too. For instance, participants made more generous choices overall in the loss than the gain frame; i.e., they forewent more own-payoff in the loss than the gain frame, and insula activation might reflect the higher level of reward foregone in the loss frame. Yet, the trial-by-trial regressor of reward amount foregone (GLM1; see Material and methods) revealed no parametric modulation of insula activity, nor of activity in any other brain region, during generous choices in either frame condition. Additionally, insula activity is unlikely to reflect the own-reward component of the generous alternative because it was fixed (always €75) and, thus, invariant across trials in both frames.

3.3. Frame-dependent modulation of VMPFC activation by rTPJ and insula

We previously provided empirical support for a network model according to which, in a task similar to our gain frame condition, TPJ would facilitate generous decision-making by modulating basic reward signals in the VMPFC, incorporating other-regarding preferences into an otherwise exclusive own-reward value representation, thus computing the vicarious value of a reward to others (Strombach et al., 2015). Here, we expand on this idea and propose that, in addition to the TPJ-VMPFC connectivity in the gain frame, frame-related information in the loss frame would activate insula, which in turn would down-regulate own-value representations in VMPFC, thus promoting generous choices by decreasing the attractiveness of own-rewards. Hence, in brief, we predicted a complex, frame-dependent pattern of connectivity between insula, TPJ, and VMPFC that reflects the different motives underlying generosity in the gain and the loss frame.

To identify the relations between those regions, we estimated their effective connectivity via DCM analysis (Friston et al., 2003). More specifically, we tested the idea that the frame information at the beginning of each trial would drive increased insula activation selectively in the loss frame, and increased TPJ activation selectively in the gain frame. Additionally, we expected increased endogenous connectivity as

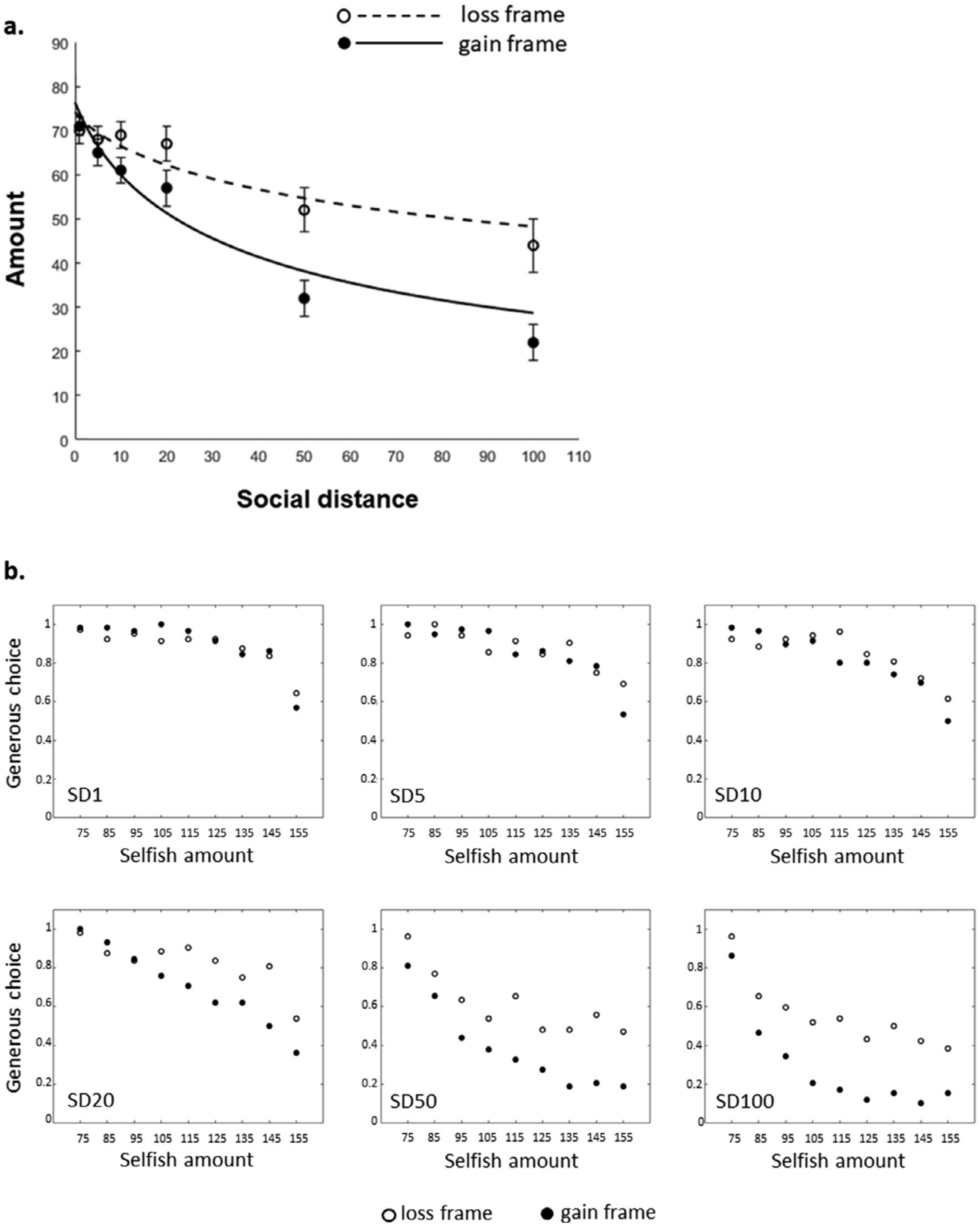


Fig. 3. Hyperbolic discount function fit and proportion of generous choices at each social distance level (fMRI study 4). (**a**) The change in generosity across social distances was captured by a hyperbolic discount model (see main text for details). The figure shows the mean of the participants' individual best-fitting hyperbolic functions, along with the mean amounts foregone (i.e. the social-distance-dependent reward amount that participants were willing to pay to increase the wealth of another person by €75; see main text) at each social distance (i.e. 1, 5, 10, 20, 50, 100), computed separately for the gain frame and the loss frame. The social discounting curve for the loss frame (dashed line) was significantly flatter than the social discounting curve for the gain frame (solid line). Circles represent the amounts foregone for the loss frame, dots represent the amounts foregone for the gain frame. Error bars represent the standard error of the mean. (**b**) Descriptive proportion of generous choices, averaged across participants, as a function of the selfish amount for the loss (circles) and the gain (dots) frame, separately for each social distance (SD).

well as condition-specific modulation between each respective region with VMPFC. Note that we focused our DCM analysis on the posterior insula cluster only, as we were interested in a baseline frame activation; including the anterior insula cluster, specific for generous choice within the loss frame (see above), might have biased the results in favor of our hypotheses.

In total we defined 15 models (see Fig. 2), grouped into two model families: A, which assumed both condition-specific driving inputs and

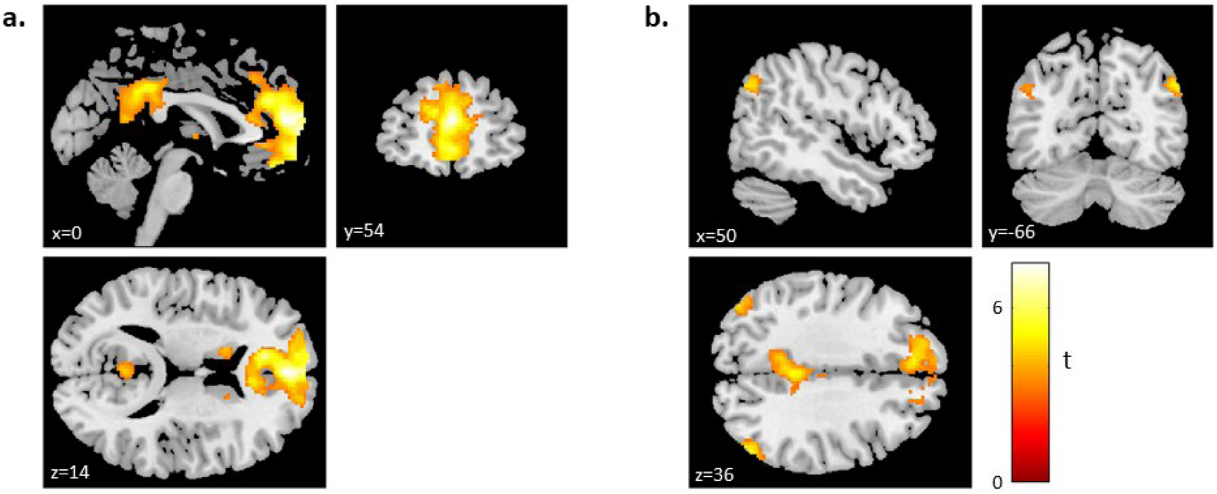


Fig. 4. Generous choices in the gain frame correlate with VMPFC and rTPJ activity. VMPFC (MNI peak [0, 54, 14]) (**a**) as well as right TPJ [50, −66, 36] (**b**) were selectively activated during [generous choice in gain frame > generous choice in loss frame] (GLM1; $p < 0.05$ FWE whole-brain corrected at the cluster level; for illustration purposes, activations are displayed at $p < 0.001$, uncorrected, minimum cluster size ≥ 5). Color bar indicates T-value.

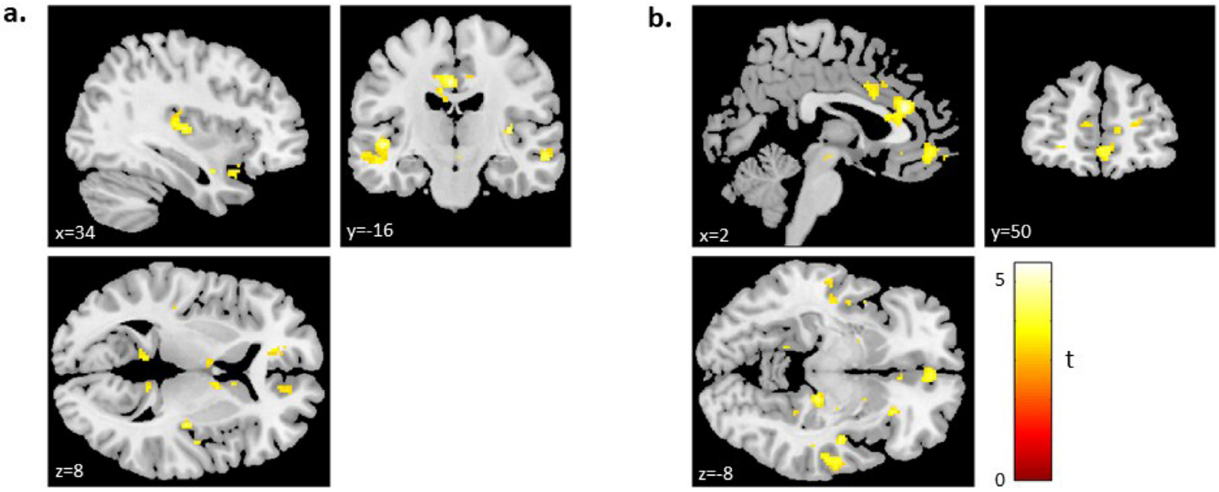


Fig. 5. The loss frame information recruits the insula and VMPFC. Insula [34, −16, 8] (**a**) as well as VMPFC [2, 50, −8] (**b**) were selectively activated during [loss frame > gain frame] onset. (GLM2; $p < 0.05$ FWE-corrected at the cluster level; for illustration purposes, activations are displayed at $p < 0.001$, uncorrected, minimum cluster size ≥ 5). Color bar indicates T-value.

condition-specific modulatory inputs; B, which assumed only condition-specific driving inputs.

Among the two model families tested, model comparison favored family A, i.e., the family of models that assumed condition-specific effects at the level of both driving input and modulatory input (family A expected posterior probability: 0.9678 vs. family B expected posterior probability: 0.0322). The winning model was model number *5* (sum of the log-evidence SF = −4.0786E+05, exceedance probability xp = 0.6595), which assumed that the gain frame had an effect on the TPJ and its connectivity with the VMPFC, while the loss frame had an effect on the VMPFC and its connectivity with the insula (i.e., connectivity between regions is assumed to be bidirectional).

Concerning the driving inputs, we compared the average activity in TPJ in the gain frame against 0, and the average activity in VMPFC in the loss frame against 0 (we checked, beforehand, that no effect of repetition across runs was present; all ps > 0.18), but none of the driving inputs was significantly different from 0 (all ps > 0.26; Table 1).

Next, when addressing the modulatory inputs, the only significant difference was found in the loss frame for modulatory activity from the insula to VMPFC against the endogenous connectivity from the insula

Table 1

DCM estimated parameters of the winning model and statistics. Values are expressed as mean ± standard error (s.e.). Statistics refer to paired t-tests between the modulatory activity and the respective endogenous connectivity, and to one-sample t-tests against 0 for driving inputs. t = t-value; p = p-value; subscript numbers are degrees of freedom; * = $p < 0.05$; endo = endogenous connectivity; mod = modulatory connectivity; drivInp = driving input; Gain = gain frame; Loss = loss frame; TPJ = temporoparietal junction; INS = insula; VMPFC = ventromedial prefrontal cortex. Arrows indicate connectivity direction.

DCM estimated parameters	mean ± *s*.e.	Statistics
endo: TPJ → VMPFC	0.05 ± 0.03	–
endo: VMPFC → TPJ	0.02 ± 0.03	–
endo: INS → VPMFC	0.05 ± 0.02	–
endo: VMPFC → INS	0.01 ± 0.02	–
mod_Gain: TPJ → VMPFC	0.03 ± 0.09	t_{29} = −0.13, p = 0.90
mod_Gain: VMPFC → TPJ	−0.03 ± 0.08	t_{29} = −0.53, p = 0.60
mod_Loss: INS → VMPFC	−0.22 ± 0.09	t_{29} = −2.56, p = 0.02*
mod_Loss: VMPFC → INS	−0.02 ± 0.07	t_{29} = −0.30, p = 0.80
drivInp_Gain: TPJ	0.04 ± 0.03	t_{29} = 1.31, p = 0.20
drivInp_Loss: VMPFC	0.00 ± 0.03	t_{29} = 0.10, p = 0.92

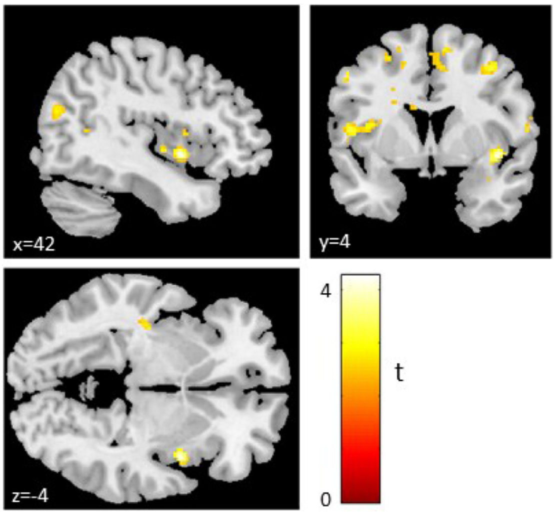


Fig. 6. Insula activation underlies generous choices in the loss frame. Insula [42, 4, −4] was selectively activated during [generous choice in loss frame > generous choice in gain frame] (GLM1; $p < 0.05$ ROI FWE-corrected at the cluster level; for illustration purposes, activations are displayed at $p < 0.01$, uncorrected, minimum cluster size ≥ 5). Color bar indicates T-value.

to VMPFC ($M_{modulatory} = -0.2158$ vs. $M_{endogenous} = 0.04672$, $p = 0.016$, Bonferroni corrected), reflecting a significant modulation of endogenous connectivity by the loss frame information (all other ps > 0.60; Table 1). In addition, the modulatory input was negative, hinting towards an inhibitory influence of insula on VMPFC in the loss frame (as before, there was no effect of repetition across runs in neither modulatory activity nor endogenous connectivity; all ps > 0.13).

3.4. The mediating role of the insula in the frame effect on social discounting

To provide further support to our idea that the frame effect on social discounting was brought about by condition-specific neural activity patterns, we ran a mediation analysis on the relation between frame information, generous behaviour, and neural activation in these regions. More specifically, frame was entered as independent variable X (gain and loss), the proportion of generous choices (gain frame and loss frame) was entered as dependent variable Y, and the neural activations were entered as mediators. We focused on a model where neural activations across both frames were treated as parallel mediators. Neural activations included the posterior insula and the anterior insula, TPJ, and VMPFC (both clusters in GLM1 and GLM2) (see Material and methods for details). Frame condition significantly correlated with all neural activations (all ps < 0.05) with the exception of VMPFC (GLM2) ($p = 0.052$). Additionally, while the direct effect of the frame condition on the proportion of generous choices was not significant ($p = 0.26$), the indirect effect of anterior insula on it was significant, indicating that it influenced frame-specific generosity (partially standardized $B = 0.15$, SE = 0.09, 95% biased-corrected CI 0.003 to 0.36) (Fig. 7).

In conclusion, our results suggest that the frame effect on social discounting was mediated by the interplay between insula and VMPFC in the loss frame, and between TPJ and VMPFC in the gain frame. Thus, we maintain that the most parsimonious explanation of insula activation and its negative modulatory interplay with VMPFC is indeed a frame-dependent downregulation of own-reward values in the valuation network during social discounting, thus decreasing participants' selfishness, while TPJ-VMPFC coupling in the gain frame reflects the upregulation of vicarious reward value signals in VMPFC, hence promoting altruism by increasing the attractiveness of the generous option. Thus, in brief, the different motives underlying generosity in the gain and the loss frame are reflected by differential, frame-dependent activation and connectivity patterns in the brain.

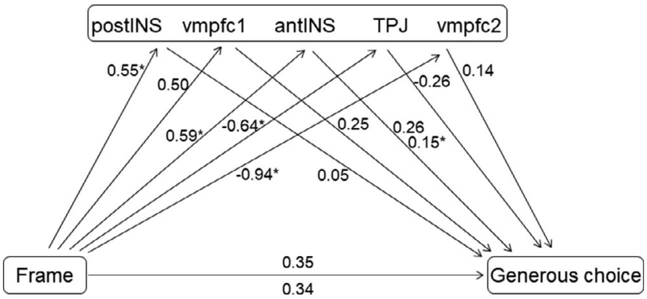


Fig. 7. Mediation analysis. A mediation model was built to clarify the effect of frame conditions (X) on generous choice (Y). Neural activations were entered as parallel potential mediators. Numbers are partially standardized effects. *refers to significant effects ($p < 0.05$). Where two numbers for the same path are reported, the one on the top refers to the direct effect and the one on the bottom refers to the indirect effect (i.e. when the mediators are included in the model).

4. Discussion

We provide behavioral and neural evidence for a simple nudge that aims at increasing individuals' willingness to provide costly support to socially remote others. We adapted a social discounting task where participants chose between a selfish option – a high gain to self and zero-gain to the other – and a generous option – a lower gain to self and equal gain to the other (Soutschek et al., 2016; Strombach et al., 2015). Based on previous evidence that people are strongly reluctant to increase their own payoff at the expense of others' welfare (Baumeister et al., 1994; Chang et al., 2011; Chang and Sanfey, 2013; Crockett et al., 2014), we framed the generous option either as a monetary gain to the other (gain frame), or as the prevention of the loss of a previous monetary endowment to the other (loss frame) (Everett et al., 2015; Li et al., 2017; List, 2007; Liu et al., 2020; Sip et al., 2015; Smith et al., 2015; Wang et al., 2017; Xiao et al., 2016; Zheng et al., 2010). Crucially, between frames, the choice alternatives differed only in the description of the decision problem, but not with regard to their actual economic consequences. In a series of four independent studies, we show that the social discount function was significantly flatter in the loss than the gain frame, indicating that participants were more generous towards socially remote others if a personal gain implied the other's loss of their previous endowment. Notably, our incentivization procedure made it logically impossible for the other persons to know about their endowment, or the potential loss of it, and participants were explicitly instructed about this; all that mattered was the final positive payoff to self and others. Yet, the fact that our participants were still reluctant to inflict losses to others suggests that they had internalized the social norm of not taking away money from others to such a degree that it prevailed even in the absence of any real economic consequences for others.

We hypothesized that the frame-dependent motives underlying generosity are dissociable on the neural level. Consistent with our previous work (Strombach et al., 2015), we found that generosity in the gain frame recruited a network of structures, including VMPFC and rTPJ, known to represent vicarious reward value and prosocial behavior. By contrast, in the loss frame, we expected that the reluctance to maximize own-gain at the expense of other-loss would be ideally mediated by social norm compliance and associated social sentiments, such as the negative emotions experienced during social norm transgressions, e.g., guilt and shame, as well as the aversive experience of unfairness and inequality. We therefore hypothesized that increased activity in brain regions associated with such social sentiments, specifically the insular cortex, would be associated with generous choices in the loss frame specifically (Bellucci et al., 2018; Canessa et al., 2017, 2013; Civai et al., 2012;

Corradi-Dell'Acqua et al., 2013; Huggins et al., 2018; Lallement et al., 2013; Lamm et al., 2011; Montague et al., 2007; Oldham et al., 2018; Paulus et al., 2003; Samanez-Larkin et al., 2008; Siebenthal et al., 2017; Singer et al., 2006; Sokol-Hessner et al., 2013; Sokol-Hessner and Rutledge, 2019; Spitzer et al., 2007; Tomasino et al., 2013; Wagner et al., 2011; Wang et al., 2017; Yu et al., 2014). We indeed found that the anterior insula was significantly more activated when participants made generous choices in the loss frame, relative to the gain frame. Extending these findings, we found that also the posterior part of the insula seemed to be involved in these processes, specifically supporting the representation of the loss frame information even before the decision was made (see also Droutman et al., 2015). Building upon this evidence, we further explored how both activation clusters mediated frame-specific social discounting behavior. We propose and provide empirical support for a network model that predicts that the frame effect on social discounting was associated with a frame-dependent neural connectivity pattern between insula and VMPFC in the loss frame, and TPJ and VMPFC in the gain frame. More specifically, DCM confirmed that posterior insula activation at loss frame onset exerted a negative modulatory effect on VMPFC. It is tempting to speculate that a frame-dependent downregulation of own-reward values in the valuation network during social discounting might lie at the core of the enhanced generosity observed in the loss frame. By contrast, the same analyses confirmed TPJ-VMPFC coupling in the gain frame, consistent with our previous finding (Strombach et al., 2015) that altruism in the gain frame is promoted by increasing the attractiveness of the generous option through TPJ-related upregulation of vicarious reward value signals in the valuation network. Overall, these results call for the idea that the motives behind generosity are likely qualitatively different in the gain and the loss frame, and dissociable on the neural level.

Our analyses revealed two separate clusters within insula; while a more posterior cluster was activated in response to general loss frame information, the more anterior cluster was specific to generous choices in the loss frame. This topographic dissociation within insula is consistent with previous findings suggesting a regional gradient in representing the level of abstraction of social sentiments during moral decision-making (e.g. Droutman et al., 2015; Ying et al., 2018). This pattern of result is in line with the idea that anterior and posterior insula may not subserve qualitatively different functions, but rather reflect different aspects of the same function, such as the interoceptive and visceral aspects of social sentiments in response to vicarious feelings of potential loss (posterior cluster), and their relevance for choice selection (anterior cluster). In addition to this, it is worth noting that, unlike insula, TPJ activity was only found at the decision stage but not at the frame information stage at trial onset. This time difference in activation allows for some speculation on the frame-dependent choice dynamics: it is conceivable that, in the loss frame, the frame information at trial onset signals the frame context, and, hence, prompts the tendency to make generous choices largely independent of social distance or selfish reward magnitude. Thus, the decision to be generous in the loss frame would be determined at trial onset already, and it would not be influenced by subsequently presented social distance or reward information. By contrast, in the gain frame, participants trade off selfish (own-payoff maximization) with other-regarding motives (granting others a gain) in a social-distance-dependent way (Strombach et al., 2015). This conflict between selfish and other-regarding motives can only be resolved when all information on frame type, social distance, and own-reward magnitude is available, that is, at the decision stage. Therefore, it is interesting as well as plausible to speculate that, while TPJ might be involved only at a later stage of the decision process, posterior insula might signal the frame context, and, hence, prompt generosity already at trial onset. Future research needs to clarify the specific functional differentiation of anterior versus posterior insula, as well as TPJ, contributions to social economic decision-making, by using, for instance, finite impulse response (FIR) models or mental chronometry approach (e.g. Menon, 2012; Schilbach et al., 2008).

Our findings expand on previous evidence that preventing harm to others is a great motivator of prosocial performance (Everett et al., 2015; Wang et al., 2017; Xiao et al., 2016; Zhang et al., 2017; Zheng et al., 2010). However, while others have found that harm prevention was particularly pronounced in a public context (Everett et al., 2015) and dependent on social feedback (Sip et al., 2015; Smith et al., 2015), we show that similar cognitive mechanisms can strongly boost generosity even in a private context and in the absence of social feedback, thus independent of reputational concerns, judgment by social peers, or third-party punishment threats. This suggests that other-harm prevention might be an internalized motive that works unconditionally and universally across contexts, regardless of social consequences. In addition, previous experiments on harm prevention did not manipulate, or provide information on, social distance between donor and recipient (Bardsley, 2008; Crockett et al., 2014; Everett et al., 2015; Li et al., 2017; Liu et al., 2020; Xiao et al., 2016). Hence, while the effects of the resource allocation mode on social discounting were elusive so far, our findings imply that it matters: harm-prevention motives in the loss frame were less dependent on social distance than other-regarding considerations in the gain frame, thus resulting in flatter social discounting.

A recent study used a similar framing manipulation and also reported TPJ involvement (Liu et al., 2020). However, their study differed from ours in several important ways. First, the task in Liu et al. (2020) involved trading off own-wealth maximization with avoiding electric shocks to others. However, their task did not involve social distance information about the recipients of shocks. Second, Liu et al. (2020) did not reveal any insula recruitment, or insula-VMPFC connectivity - the core finding in our study - related to generosity or task framing. Most importantly, perhaps, while Liu et al. (2020) identified TPJ-VMPFC connectivity to be relevant for their frame-related increase in costly harm-prevention, we found instead that insula-VMPFC connectivity was associated with the frame-related boost in generosity during social discounting. This suggests that Liu et al. (2020) most likely studied different framing-related cognitive and neural mechanisms than the ones investigated here.

Our results are consistent with the idea that certain costly altruistic behaviors are not motivated by genuinely other-regarding considerations, but instead by compliance to internalized social norms. But what impels participants to comply to social norms? Here, we propose, along with previous evidence (Chang et al., 2011; Spitzer et al., 2007), that compliance to social norms might be linked to anticipated feelings of guilt, shame, and remorse, and accompanied by insula activation (see also Belfi et al., 2015; Sellitto et al., 2016), which ultimately sustain prosocial behavior. According to this view, insula would reflect the negative sentiment associated with social norm transgressions as they occur when being responsible for someone else's loss (i.e. vicarious loss experience). Our data show that this social sentiment and accompanying neural signature can be elicited even when the others' outcomes are merely described as losses, thus, in the absence of real losses to others.

The success of our framing manipulation in increasing generosity came at a methodological cost: because participants rarely made selfish choices in the loss frame, the analyses on selfish choices were underpowered. Hence, results involving selfish decisions, and how they map on insula, TPJ, or VMPFC, have to be interpreted with caution. To shed more light on the neural correlates of selfish choices in the loss frame, future studies should replicate our experiment with a slightly less effective nudge that would allow for more selfish choices.

5. Conclusions

The acceptance and support of the principle of a caring society, and the attitude towards the welfare of socially remote strangers, is central for a civilization to function well. It seems vital for societies to successfully meet current challenges, such as integrating refugees, addressing economic inequality, acceding the trials and promises of a globalized world (Kalenscher, 2014), or managing the public health implications of the current COVID-19 pandemic. Here, we present a simple behavioral framing manipulation that boosts generosity towards socially remote others: framing a selfish choice as a loss to others can motivate prosocial behavior, even if the framing of the choice options is irrelevant for the actual payoff to others. Our neuroimaging data identify insula as the core component in a network associated with this enhanced generosity in the loss frame. Our results imply that prosocial attitudes towards others are highly malleable and strongly depend on the architecture of the decision problem. The insights gained in this study might, thus, help in designing policies aimed at increasing the acceptance and support of the principle of a caring society, and to change the attitude towards the welfare of socially remote strangers.

Credit authorship contribution statement

Manuela Sellitto: Conceptualization, Visualization, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Susanne Neufang:** Writing – original draft, Writing – review & editing. **Adam Schweda:** Visualization, Investigation, Writing – review & editing. **Bernd Weber:** Visualization, Writing – review & editing. **Tobias Kalenscher:** Conceptualization, Writing – original draft, Writing – review & editing.

Acknowledgments

This work was supported by Deutsche Forschungsgemeinschaft (DFG) grant no. KA 2675/4–3 (to T.K.).

Data and code availability statement

Digital Imaging and Communications in Medicine (DICOM) images reported in this paper have been deposited in XNAT Central (https://central.xnat.org/) under the project name ‘Framesocdisc’. Behavioral datasets have been supplied in Figshare (https://figshare.com/) under the doi: 10.6084/m9.figshare.10265309. The code used for fMRI analyses was retrieved from the statistical parametric mapping (SPM) platform.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118211.

References

Archambault, C., Kalenscher, T., De Laat, J., 2019. Generosity and livelihoods: experimental evidence on the multidimensional nature of sharing among the Kenyan Maasai. J. Behav. Decis. Mak. 1–12.

Bardsley, N., 2008. Dictator game giving: altruism or artefact? Exp. Econ. 11, 122–133. doi:10.1007/s10683-007-9172-2.

Bartra, O., Mcguire, J.T., Kable, J.W., 2013. The valuation system : a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage 76, 412–427. doi:10.1016/j.neuroimage.2013.02.063.

Baumeister, R.F., Stillwell, A.M., Heatherton, T.F., 1994. Guilt: An Interpersonal Approach. Psychol. Bulletin. 115, 243–267.

Belfi, A.M., Koscik, T.R., Tranel, D., 2015. Damage to the insula is associated with abnormal interpersonal trust. Neuropsychologia 71, 165–172. doi:10.1016/j.neuropsychologia.2015.04.003.

Bellucci, G., Feng, C., Camilleri, J., Eickhoff, S.B., Krueger, F., 2018. The role of the anterior insula in social norm compliance and enforcement: evidence from coordinate-based and functional connectivity meta-analyses. Neurosci. Biobehav. Rev. 92, 378–389. doi:10.1016/j.neubiorev.2018.06.024.

Canessa, N., Crespi, C., Baud-bovy, G., Dodich, A., Falini, A., Antonellis, G., Cappa, S.F., 2017. Neural markers of loss aversion in resting-state brain activity. Neuroimage 146, 257–265. doi:10.1016/j.neuroimage.2016.11.050.

Canessa, N., Crespi, C., Motterlini, M., Baud-bovy, G., Chierchia, G., Pantaleo, G., Tettamanti, M., Cappa, S.F., 2013. The functional and structural neural basis of individual differences in loss aversion. J. Neurosci. 33, 14307–14317. doi:10.1523/JNEUROSCI.0497-13.2013.

Chang, L.J., Sanfey, A.G., 2013. Great expectations : neural computations underlying the use of social norms in decision-making. Soc. Cognit. Affect. Neurosci. 8, 277–284. doi:10.1093/scan/nsr094.

Chang, L.J., Smith, A., Dufwenberg, M., Sanfey, A.G., 2011. Article triangulating the neural, psychological, and economic bases of guilt aversion. Neuron 70, 560–572. doi:10.1016/j.neuron.2011.02.056.

Cho, Y.T., Fromm, S., Guyer, A.E., Detloff, A., Pine, D.S., Fudge, J.L., Ernst, M., 2013. Nucleus accumbens, thalamus and insula connectivity during incentive anticipation in typical adults and adolescents. Neuroimage 66, 508–521. doi:10.1016/j.neuroimage.2012.10.013.

Civai, C., Crescentini, C., Rustichini, A., Rumiati, R.I., 2012. Equality versus self-interest in the brain: differential roles of anterior insula and medial prefrontal cortex. Neuroimage 62, 102–112. doi:10.1016/j.neuroimage.2012.04.037.

Corradi-Dell’Acqua, C., Civai, C., Rumiati, R.I., Fink, G.R., 2013. Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. Soc. Cognit. Affect. Neurosci. 8, 424–431. doi:10.1093/scan/nss014.

Crockett, M.J., Kurth-nelson, Z., Siegel, J.Z., Dayan, P., Dolan, R.J., Crockett, M.J., Kurth-nelson, Z., Siegel, J.Z., Dayan, P., Dolan, R.J., 2014. Harm to others outweighs harm to self in moral decision making. Proc. Natl. Acad. Sci. 111, 17320–17325. doi:10.1073/pnas.1424572112.

Davis, M.H., 1983. A mulitdimensional approach to individual differences in empathy. J. Pers. Soc. Psychol. 44, 113–126. doi:10.1037/0022-3514.44.1.113.

Dreu, D., 1997. Gain-loss frames and cooperation in two-person social dilemmas: a transformational analysis, 5th, 72. J. Person. Soc. Psychol., pp. 1093–1106.

Droutman, V., Bechara, A., Read, S.J., 2015. Roles of the different sub-regions of the insular cortex in various phases of the decision-making process. Front. Behav. Neurosci. 9, 309. doi:10.3389/fnbeh.2015.00309.

Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.H., Evans, A.C., Zilles, K., Amunts, K., 2007. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage 36, 511–521. doi:10.1016/j.neuroimage.2007.03.060.

Evans, A.M., Beest, I.V., 2017. Gain-loss framing effects in dilemmas of trust and reciprocity. J. Exp. Soc. Psychol. 73, 151–163. doi:10.1016/j.jesp.2017.06.012.

Everett, J.A.C., Faber, N.S., Crockett, M.J., 2015. The influence of social preferences and reputational concerns on intergroup prosocial behaviour in gains and losses contexts. R. Soc. Open Sci. 2, 150546. doi:10.1098/rsos.150546.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302. doi:10.1016/S1053-8119(03)00202-7.

Gu, X., Wang, X., Hula, X.A., Wang, S., Xu, S., Lohrenz, T.M., Knight, R.T., Gao, Z., Dayan, P., Montague, P.R., 2015. Necessary, yet dissociable contributions of the insular and ventromedial prefrontal cortices to norm adaptation: computational and lesion evidence in humans. J. Neurosci. 35, 467–473. doi:10.1523/JNEUROSCI.2906-14.2015.

Hayes, A.F., 2017. Introduction to mediation, moderation, and Conditional Process Analysis: A regression-Based Approach. Guilford Publications, New York.

Huggins, A.A., Belleau, E.L., Miskovich, T.A., Pedersen, W.S., Larson, C.L., 2018. Moderating effects of harm avoidance on resting-state functional connectivity of the anterior insula. Front. Hum. Neurosci. 12, 1–9. doi:10.3389/fnhum.2018.00447.

Hutcherson, C.A., Bushong, B., Rangel, A., 2015. A neurocomputational model of altruistic choice and its implications. Neuron 87, 451–463. doi:10.1016/j.neuron.2015.06.031.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790. doi:10.1016/j.neuroimage.2011.09.015.

Jones, B., Rachlin, H., 2006. Social discounting. Psychol. Sci. 17, 283–286. doi:10.1111/j.1467-9280.2006.01699.x.

Kalenscher, T., 2017. Social psychology: love thy stranger as thyself. Nat. Hum. Behav. 1, 0108. doi:10.1038/s41562-017-0108.

Kalenscher, T., 2014. Attitude toward health insurance in developing countries from a decision-making perspective. J. Neurosci. Psychol. Econ. 7:3, 174–193.

Lallement, J.H., Kuss, K., Trautner, P., Weber, B., Falk, A., Fliessbach, K., 2013. Effort Increases Sensitivity to Reward and Loss Magnitude in the Human Brain, 9. Soc. Cogn. Affect. Neurosci., pp. 342–349.

Lamm, C., Decety, J., Singer, T., 2011. NeuroImage Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 54, 2492–2502. doi:10.1016/j.neuroimage.2010.10.014.

Li, R., Smith, D.V., Clithero, J.A., Venkatraman, V., Carter, R.M., Huettel, S.A., 2017. Reason’s enemy is not emotion: engagement of cognitive control networks explains biases in gain/loss framing. J. Neurosci. 37. doi:10.1523/JNEUROSCI.3486-16.2017, 3486–16.

Li, Z., Yan, C., Xie, W.Z., Li, K., Zeng, Y.W., Jin, Z., Cheung, E.F.C., Chan, R.C.K., 2015. Anticipatory pleasure predicts effective connectivity in the mesolimbic system. Front. Behav. Neurosci. 9, 1–8. doi:10.3389/fnbeh.2015.00217.

List, J., 2007. On the interpretation of giving in dictator games. J. Polit. Econ. 115, 482–493. doi:10.1086/519249.

Liu, J., Gu, R., Liao, C., Lu, J., Fang, Y., Xu, P., Luo, Y., Cui, F., 2020. The neural mechanism of the social framing effect: evidence from fMRI and tDCS studies. J. Neurosci. 40. doi:10.1523/jneurosci.1385-19.2020, JN-RM-1385-19.

Margittai, Z., Strombach, T., van Wingerden, M., Joëls, M., Schwabe, L., Kalenscher, T., 2015. A friend in need: time-dependent effects of stress on social discounting in men. Horm. Behav. 73, 75–82. doi:10.1016/j.yhbeh.2015.05.019.

Margittai, Z., Van Wingerden, M., Schnitzler, A., Joels, M., Kalenscher, T., 2018. Dissociable roles of glucocorticoid and noradrenergic activation on social discounting. Psychoneuroendocrinol. doi:10.1016/j.psyneuen.2018.01.015.

Brett, M., Anton, J.L., Romain Valabregue, J.B.P., 2002. Region of interest analysis using an SPM toolbox. In: Proceedings of the 8th International Conference on Functional Mapping of the Human Brain doi:10.1007/978-3-540-75387-2_17.

Mazaika, P.K., Hoeft, F., Glover, G.H., Reiss, A.L., 2009. Methods and Software for fMRI Analysis for Clinical Subjects, 47:1. NeuroImage, p. S58.

Menon, R.S., 2012. Mental chronometry. Neuroimage 62, 1068–1071. doi:10.1016/j.neuroimage.2012.01.011.

Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A.J., Schweizer, S., Frith, C.D., Dalgleish, T., 2009. A key role for similarity in vicarious reward. Science 324, 900. doi:10.1126/science.1170539.

Montague, P.R., Lohrenz, T., Humphrey, H., 2007. To detect and correct : norm violations and their enforcement. Neuron 14–18. doi:10.1016/j.neuron.2007.09.020.

Neufang, S., Akhrif, A., Herrmann, C.G., Drepper, C., Homola, G.A., Nowak, J., Waider, J., Schmitt, A.G., Lesch, K.P., Romanos, M., 2016. Serotonergic modulation of "waiting impulsivity" is mediated by the impulsivity phenotype in humans. Transl. Psychiatry 6. doi:10.1038/tp.2016.210.

Nowak, M.A., 2006. Five rules for the evolution of cooperation. Science, 314:5805, 1560–1563. 10.1126/science.1133755.

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Lorenzetti, V., 2018. The anticipation and outcome phases of reward and loss processing : a neuroimaging meta-analysis of the monetary incentive delay task. Hum. Brain. Mapp. 3398–3418. doi:10.1002/hbm.24184.

Paulus, C., 2009. Der Saarbrücker Persönlichkeitsfragebogen (SPF-IRI) zur Messung von Empathie. 1–11.

Paulus, M.P., Rogalsky, C., Simmons, A., Feinstein, J.S., Stein, M.B., 2003. Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. NeuroImage 19, 1439–1448. doi:10.1016/S1053-8119(03)00251-9.

Rilling, J.K., Sanfey, A.G., 2011. The Neuroscience of Social Decision-Making, 62. The Annual Review of Psychology, pp. 23–48.

Rolls, E.T., Huang, C.C., Lin, C.P., Feng, J., Joliot, M., 2020. Automated anatomical labelling atlas 3. Neuroimage 206, 116189. doi:10.1016/j.neuroimage.2019.116189.

Samanez-Larkin, G.R., Hollon, N.G., Carstensen, L.L., Knutson, B., 2008. Psychol. Sci. 19, 320–323. doi:10.1111/j.1467-9280.2008.02087.x.Individual.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people the role of the temporo-parietal junction in "theory of mind. NeuroImage 19, 1835–1842. doi:10.1016/S1053-8119(03)00230-1.

Schilbach, L., Eickhoff, S.B., Mojzisch, A., Vogeley, K., 2008. What's in a smile? Neural correlates of facial embodiment during social interaction. Soc. Neurosci. 3, 37–50. doi:10.1080/17470910701563228.

Schweda, A., Margittai, Z., Kalenscher, T., 2020. Acute stress counteracts framing-induced generosity boosts in social discounting in young healthy men. Psychoneuroendocrinol. 121, 104860. doi:10.1016/j.psyneuen.2020.104860.

Seara-Cardoso, A., Sebastian, C.L., Mccrory, E., Foulkes, L., Buon, M., Roiser, J.P., Viding, E., 2016. Anticipation of guilt for everyday moral transgressions : the role of the anterior insula and the influence of interpersonal psychopathic traits. Sci. Rep. 6, 36273. doi:10.1038/srep36273.

Sellitto, M., Ciaramelli, E., Mattioli, F., di Pellegrino, G., 2016. Reduced sensitivity to sooner reward during intertemporal decision-making following insula damage in humans. Front. Behav. Neurosci. 9, 367. doi:10.3389/fnbeh.2015.00367.

Singer, T., Seymour, B., Doherty, J.P.O., Stephan, K.E., Dolan, R.J., Frith, C.D., 2006. Empathic neural responses are modulated by the perceived fairness of others. Nature 439, 10–13. doi:10.1038/nature04271.

Sethi, R., Somanathan, E., E., 1996. The evolution of social norms in common property resource use. Am. Econ. Rev. 86 (4), 766–788 (Sep., 1996)Published by : American Econom 86, 766–788.

Sip, K.E., Smith, D.V., Porcelli, A.J., Kar, K., Delgado, M.R., 2015. Social closeness and feedback modulate susceptibility to the framing effect. Soc. Neurosci. 10, 35–45. doi:10.1080/17470919.2014.944316.

Smith, D.V., Sip, K.E., Delgado, M.R., 2015. Functional connectivity with distinct neural networks tracks fluctuations in gain/loss framing susceptibility. Hum. Brain Mapp. 36, 2743–2755. doi:10.1002/hbm.22804.

Sokol-Hessner, P., Camerer, C.F., Phelps, E.A., 2013. Emotion Regulation Reduces Loss Aversion and Decreases Amygdala Responses to Losses, 8:13. Soc. Cogn. Affect. Neurosci., pp. 341–350.

Sokol-Hessner, P., Rutledge, R.B., 2019. The Psychological and Neural Basis of Loss Aversion, 28:1. Curr. Dir. Psychol. Sci., pp. 20–27.

Soutschek, A., Ruff, C.C., Strombach, T., Kalenscher, T., Tobler, P.N., 2016. Brain stimulation reveals crucial role of overcoming self-centeredness in self-control. Sci. Adv. 2, e1600992. doi:10.1126/sciadv.1600992, –e1600992.

Soutschek, A., Ugazio, G., Crockett, M.J., Ruff, C.C., Kalenscher, T., Tobler, P.N., 2017. Binding oneself to the mast: stimulating frontopolar cortex enhances precommitment. Soc. Cognit. Affect. Neurosci. 1–8. doi:10.1093/scan/nsw176.

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., Fehr, E., 2007. The neural signature of social norm compliance. Neuron 56, 185–196. doi:10.1016/j.neuron.2007.09.011.

Stöber, J., 2001. The Social Desirability Scale-17 (SDS-17): convergent validity, discrimant validity, and relationship with age. Eur. J. Psychol. Assess. 17, 222–232. doi:10.1027//1015-5759.17.3.222.

Strang, S., Gerhardt, H., Marsh, N., Oroz, S., Hu, Y., Hurlemann, R., Park, S.Q., 2017. Short communication a matter of distance — the effect of oxytocin on social discounting is empathy-dependent. Psychoneuroendocrinol. 78, 229–232. doi:10.1016/j.psyneuen.2017.01.031.

Strombach, T., Jin, J., Weber, B., Kenning, P., Shen, Q., Ma, Q., Kalenscher, T., 2014. Charity begins at home: cultural differences in social discounting and generosity. J. Behav. Decis. Mak. 27, 235–245. doi:10.1002/bdm.1802.

Strombach, T., Margittai, Z., Gorczyca, B., Kalenscher, T., 2016. Gender-specific effects of cognitive load on social discounting. PLoS One 11, 1–15. doi:10.1371/journal.pone.0165289.

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I.I., Tobler, P.N., Kalenscher, T., 2015. Social discounting involves modulation of neural value signals by temporoparietal junction. Proc. Natl. Acad. Sci. 112, 1619–1624. doi:10.1073/pnas.1414715112.

Studer, B., Koch, C., Knecht, S., Kalenscher, T., 2019. Conquering the inner couch potato: precommitment is an effective strategy to enhance motivation for effortful actions. Philos. Trans. R. Soc. Lond. B Biol. Sci. 374, 20180131. doi:10.1098/rstb.2018.0131.

Tomasino, B., Lotto, L., Sarlo, M., Civai, C., Rumiati, R., Rumiati, R.I., 2013. Framing the ultimatum game: the contribution of simulation. Front. Hum. Neurosci. 7, 1–16. doi:10.3389/fnhum.2013.00337.

Tusche, X.A., Bo, A., Kanske, X.P., Trautwein, X.F., Singer, T., 2016. Decoding the charitable brain : empathy, perspective taking, and attention shifts differentially predict altruistic giving. J. Neurosci. 36, 4719–4732. doi:10.1523/JNEUROSCI.3392-15.2016.

Vekaria, K.M., Brethel-Haurwitz, K.M., Cardinale, E.M., Stoycos, S.A., Marsh, A.A., 2017. Social discounting and distance perceptions in costly altruism. Nat. Hum. Behav. 1, 0100. doi:10.1038/s41562-017-0100.

Vlaev, I., 2012. How different are real and hypothetical decisions ? Overestimation, contrast and assimilation in social interaction. J. Econ. Psychol. 33, 963–972. doi:10.1016/j.joep.2012.05.005.

Von Siebenthal, Z., Boucher, O., Rouleau, I., Lassonde, M., Lepore, F., Nguyen, D.K., De, C., 2017. Decision-making impairments following insular and medial temporal lobe resection for drug-resistant epilepsy. Soc. Cogn. Affect. Neurosci. 128–137. doi:10.1093/scan/nsw152.

Wagner, U., N'Diaye, K., Ethofer, T., Vuilleumier, P., 2011. Guilt-specific processing in the prefrontal cortex. Cereb. Cortex 21, 2461–2470. doi:10.1093/cercor/bhr016.

Wang, X.T., Rao, L.-.L., Zheng, H., 2017. Neural substrates of framing effects in social contexts: a meta-analytical approach. Soc. Neurosci. 12, 268–279. doi:10.1080/17470919.2016.1165285.

Wendelken, C., O'Hare, E.D., Whitaker, K.J., Ferrer, E., Bunge, S.A., 2011. Increased functional selectivity over development in rostrolateral prefrontal cortex. J. Neurosci. 31, 17260–17268. doi:10.1523/JNEUROSCI.1193-10.2011.

Wiehler, A., Petzschner, F.H., Stephan, K.E., Peters, J., 2017. Episodic Tags Enhance Striatal Valuation Signals during Temporal Discounting in Pathological Gamblers, 4.

Wilson, R.C., Collins, A.G.E., 2019. Ten simple rules for the computational modeling of behavioral data. Elife 8, 1–33. doi:10.7554/eLife.49547.

Xiang, T., Lohrenz, T., Montague, P.R., 2013. Computational substrates of norms and their violations during social exchange. J. Neurosci. 33, 1099–1108. doi:10.1523/JNEUROSCI.1642-12.2013.

Xiao, F., Zheng, Z., Zhang, H., Xin, Z., Chen, Y., Li, Y., 2016. Who are you more likely to help? the effects of expected outcomes and regulatory focus on prosocial performance. PLoS One 11, 1–15. doi:10.1371/journal.pone.0165717.

Ying, X., Luo, J., Chiu, C.yue, Wu, Y., Xu, Y., Fan, J., 2018. Functional dissociation of the posterior and anterior insula in moral disgust. Front. Psychol. 9, 1–10. doi:10.3389/fpsyg.2018.00860.

Yu, H., Hu, J., Hu, L., Zhou, X., 2014. The voice of conscience : neural bases of interpersonal guilt and compensation. Soc. Cognit. Affect. Neurosci. 1150–1158. doi:10.1093/scan/nst090.

Zeidman, P., Jafarian, A., Corbin, N., Seghier, M.L., Razi, A., Price, C.J., Friston, K.J., 2019a. A guide to group effective connectivity analysis, part 1: first level analysis with DCM for fMRI. Neuroimage 200, 174–190. doi:10.1016/j.neuroimage.2019.06.031.

Zeidman, P., Jafarian, A., Seghier, M.L., Litvak, V., Cagnan, H., Price, C.J., Friston, K.J., 2019b. A guide to group effective connectivity analysis, part 2: second level analysis with PEB. Neuroimage 200, 12–25. doi:10.1016/j.neuroimage.2019.06.032.

Zhang, X., Liu, Yi, Chen, X., Shang, X., Liu, Yongfang, 2017. Decisions for others are less risk-averse in the gain frame and less risk-seeking in the loss frame than decisions for the self. Front. Psychol. 8, 1–10. doi:10.3389/fpsyg.2017.01601.

Zhang, Y.Y., Xu, L., Liang, Z.Y., Wang, K., Hou, B., Zhou, Y., Li, S., Jiang, T., 2018. Separate neural networks for gains and losses in intertemporal choice. Neurosci. Bull. 34, 725–735. doi:10.1007/s12264-018-0267-x.

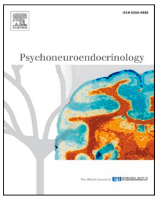
Zheng, H., Wang, X.T., Zhu, L., 2010. Framing effects: behavioral dynamics and neural basis. Neuropsychologia 48, 3198–3204. doi:10.1016/j.neuropsychologia.2010.06.031.

Psychoneuroendocrinology 121 (2020) 104860

Contents lists available at ScienceDirect

Psychoneuroendocrinology

journal homepage: www.elsevier.com/locate/psyneuen



Acute stress counteracts framing-induced generosity boosts in social discounting in young healthy men

A. Schweda*, Z. Margittai, T. Kalenscher

Comparative Psychology, Heinrich-Heine-University Düsseldorf, Germany

ARTICLE INFO

Keywords:
Stress
Acute Stress
Framing
Social Discounting
Prosocial Choice

ABSTRACT

Most individuals are willing to forego resources for the benefit of others, but their willingness to do so typically declines as a function of social distance between the donor and recipient, a phenomenon termed social discounting. We recently showed that participants were more altruistic towards strangers when a costly generous choice was framed as preventing a monetary loss to the other rather than granting them a gain. Here, we asked if acute stress would diminish this frame effect on social discounting. To test this hypothesis, 102 male participants engaged in either the Maastricht Acute Stress Task, or a matched, non-stressful control procedure. They subsequently played a two-frame dictator game version of the social discounting paradigm. Whereas both frame conditions were economically equivalent, in the give frame, participants were asked how much money they would share with other persons on variable social distance levels, and in the take frame, they decided on how much money to take away from the others. While non-stressed control participants showed increased generosity toward strangers in the take compared to the give frame, similar to previous findings of our group, stress attenuated this frame effect on social discounting by reducing generosity toward strangers in the take frame. These findings confirm that stress can corrupt prosocial motives and social norm compliance, diminishing prosocial tendencies toward unfamiliar others.

1. Introduction

Most individuals are willing to forego resources for the benefit of others. However, their willingness to do so typically declines as a function of social distance between donor and recipient, i.e., how much the donor cares about the recipient, a phenomenon termed social discounting (Jones and Rachlin, 2006; Strombach et al., 2015). Social discounting is relatively robust within individuals (Archambault et al., 2019; Kalenscher, 2017; Vekaria et al., 2017), yet it is also malleable. For example, we and others have shown that generosity toward socially close others, e.g., friends and family, can be increased by psychosocial stress (Margittai et al., 2015) or psychopharmacological manipulation (Margittai et al., 2018), and generosity towards strangers can be decreased by transcranial magnetic stimulation (Soutschek et al., 2016; cf. also Gallo et al., 2018). In addition, we recently demonstrated that the way the decision problem was described influenced social discounting (Sellitto et al., 2019): we found that participants were much more altruistic towards others, especially strangers, when the resource allocation problem in the social discounting task was framed as a decision to obtain a personal financial benefit at the other's expense (*take frame*[1]) versus to financially benefit the other at an own personal expense (*give frame*), even when actual economic outcomes were equivalent across frames. For example, people alloted more money to a stranger when being asked how much of a monetary endowment they would take away from that person (*take frame*) compared to when being asked how much of their own endowment they would share (*give frame*). Such framing-induced boosts in generosity towards others is likely to reflect the observation that people are more sensitive to others' losses than gains (De Dreu and McCusker, 1997; Evans and van Beest, 2017; Ishii and Eisen, 2018; Ispano and Schwardmann, 2017; Takahashi, 2013), and they are consequently reluctant to increase their own payoff at the expense of others' welfare (Bardsley, 2008; Baumeister et al., 1994; Chang et al., 2011; Chang and Sanfey, 2013; Crockett et al., 2014). We recently argued (Sellitto et al., 2019; cf. also Decety & Cowell, 2018)

---

* Corresponding author at: Heinrich-Heine-University Düsseldorf, Comparative Psychology, Universitätsstraße 1, 40225 Düsseldorf, Germany.
E-mail address: adam.schweda@uni-duesseldorf.de (A. Schweda).

[1] Note that the terminology used here differs from the one used in Sellitto et al. (2019): the 'give' frame was termed 'gain' frame in Sellitto et al. (2019), and the 'take' frame was termed 'loss' frame. We chose to use a different terminology from Sellitto et al. (2019) to highlight the differences in task details between the two studies, and to better reflect the nature of the frame manipulation used here.

https://doi.org/10.1016/j.psyneuen.2020.104860
Received 5 December 2019; Received in revised form 9 August 2020; Accepted 24 August 2020

that this differential sensitivity to others' gains and losses, and its impact on social discounting is likely to be the result of compliance to the social norm to avoid causing harm to others, which is an important prerequisite for sustainable social relationships.

Generosity, prosocial behavior and social discounting are strongly influenced by acute stress (Margittai et al., 2015, 2018). For example, recent evidence suggests that individuals may react to stress with a "tend-and-befriend"-response (Berger et al., 2016; Margittai et al., 2015, 2018, 2015; Sollberger et al., 2016; Taylor, 2006; Taylor et al., 2000; Von Dawans et al., 2019, 2011) – an increase in costly generosity towards others to mobilize social support in stressful times. Because of their desire to protect, and thus avoid damaging, their social relationships, tend-and-befriend implies that stressed individuals would be even less inclined to cause harm to others. Consequently, stress should amplify the above-mentioned frame effect on social discounting by shifting decisions in the *take* frame away from financially hurting others towards being even more generous toward them.

However, often, the social response to stress is *not* tend-and-befriend, but fight-or-flight (Cannon, 1932; Dedovic et al., 2009; McCarty, 2016; Rodrigues et al., 2009). Fight-or-flight responses involve antagonistic social behaviors aimed at promoting own survival and well-being, potentially at the opponent's expense. This social response to stress has been described almost a century ago (Cannon, 1932), and its discovery has had great impact on the animal and human literature (Haller, 2018; Haller et al., 1998; Jansen et al., 1995; Kruk et al., 2004; McCarty, 2016; Sandi and Haller, 2015; Sgoifo et al., 1996; Sgoifo and Papi, 1995; Terbeck et al., 2016, 2012; White et al., 2019). In humans, antagonistic fight-or-flight-like responses might manifest as higher egocentricity and reduced other-regarding behavior. For example, a recent study found that stress can induce spiteful punishment, weakened trust and reduced reciprocity (Steinbeis et al., 2015). Also, another recent study showed that, under stress, the neural representations of self- and other-regarding values diverged more than in a non-stress condition (Tomova et al., 2020). Furthermore, stress was found to diminish the willingness to share resources (Starcke et al., 2011; Steinbeis et al., 2015; Vinkers et al., 2013) and the stress hormone cortisol has been associated with an increased tendency towards egoistic decision making in everyday moral dilemmas (Starcke et al., 2011). Combined, this suite of evidence tentatively suggests that stress can lead to less other-regarding thinking combined with an erosion of moral and social norms (but see Singer et al., 2017; Nickels et al., 2017). Because, as mentioned above, the frame effect on social discounting likely depends on other-regarding considerations and social norm compliance (Sellitto et al., 2019), stressed individuals reacting in a fight-or-flight-like manner should fail to show frame-dependent differences in generosity toward others. Thus, rather than amplifying it, acute stress would be expected to dampen the frame effect.

Taken together, the frame-amplification hypothesis, inspired by the tend-and-befriend theory, predicts a stress-related amplification of the frame effect on social discounting, i.e., stressed individuals should be even more generous toward others when a resource allocation problem was framed as taking away money from others (*take frame*) versus sharing money with them (*give frame*). By contrast, the frame-attenuation hypothesis, inspired by the fight-or-flight model, states that stressed individuals will show a diminished frame effect on social discounting.

In order to decide between these hypotheses, we pseudo-randomly assigned 102 male participants either to a stress procedure (Smeets et al., 2012) or a non-stress control condition and asked them to complete a social discounting task with randomly interleaved *give* and *take* frame decisions. In each trial, participants decided on the allocation of funds between themselves and another person at a variable social distance level. The *give* frame consisted of a variant of the dictator game (Bolton et al., 1998) where participants were endowed with an amount of money, and decided how much of this endowment they would share with the other person. In the *take* frame, participants were informed that the other person was endowed with an amount of money, and the participant decided how much money to take away from that person. Participants were repeatedly and explicitly instructed that the other persons would only be informed about the outcome of the share, but not about their initial endowment, or the loss of it; all that mattered for the other person was the final payoff. Both frames were economically equivalent in terms of financial outcomes.

In support of the frame-attenuation hypothesis, we found that acute stress diminished the frame effects on social discounting: while non-stressed control participants became more generous towards strangers in the take compared to the give frame, this frame-dependent change in generosity was less pronounced in stressed participants.

2. Methods

2.1. Participants

102 male participants (age M = 22.95, SD = 3.92, range = 18 - 36) were recruited within the University of Düsseldorf. One participant withdrew from participation due to pain during the stress induction. Participants were screened via telephone interview before participation. We considered participants eligible if they were male, between 18 and 40 years old, German speakers, no heavy smokers (< 5 cigarettes/day), no heavy drinkers (< 3 portions/day), and no regular drug users. We excluded individuals diagnosed with current psychiatric, neurological, endocrinal, cardiovascular or urological conditions, users of medication strongly affecting the central nervous system, or cardiovascular or endocrine system. Because recent findings suggest that lesbian, gay or bisexual individuals have an altered physiological CORT response to stress (Hatzenbuehler & McLaughlin, 2014) and an altered diurnal CORT profile (most likely due to minority stress, see Parra et al., 2016), only heterosexual men were included. Also, body weight interacts with CORT baselines and CORT responsiveness to stressors (Herhaus & Petrowski, 2018). Therefore, we only considered participants with a BMI below 30. Furthermore, psychology and economics majors were excluded due to potential prior knowledge about the effects of stress on cognition and economic decision making. The experiment was carried out between 14:00 and 18:00, during the circadian trough of cortisol to minimize the potential moderating role of circadian hormonal fluctuations on stress responsiveness. We asked subjects to abstain from consumption of cigarettes and caffeine for four hours before starting the experiment, food for two hours, and sex, alcohol and medication for 24 hours. We used an exclusively male sample in order to avoid differential HPA-axis activation caused by the intake of oral contraceptives and variations in menstrual cycle (Kirschbaum et al., 1999; Kudielka and Kirschbaum, 2005). All participants gave informed consent, the experiment was approved by the local ethics committee of the Heinrich-Heine-University, and it complied with the regulations of the Declaration of Helsinki.

2.2. Materials

2.2.1. Online Questionnaires

After being pseudo-randomly assigned to either the stress or the control group, but prior to being invited to the laboratory, all participants completed a number of trait questionnaires online, designed to control for potential confounding factors that might interfere with stress reactivity and/or our main outcome measures. We measured impulsivity (Barratt Impulsiveness Scale, BIS-15, Meule et al., 2011), behavioral inhibition and activation (Behavioral Inhibition / Activation Scale, BIS/BAS, Carver and White, 1994), chronotype (reduced version of the Morningness-Eveningness Questionnaire, rMEQ, Randler, 2013), chronic stress (Trier Inventory of Chronic Stress, TICS, Schulz and Schlotz, 1999), social desirability (Social Desirability Scale, SDS-17, Stöber, 2001), trait anxiety (Trait Scale of the State-Trait Anxiety Inventory, STAI, Spielberger, 1983), personality (10-item Big-5 Inventory, BFI-10, Rammstedt, 2007), psychopathy (Levenson Self-Report Psychopathy Scale, LSRP, Levenson et al., 1995), empathy (Saarbruecker Persoenlichkeitsfragebogen, SPF, Paulus, 2009), risk taking and social value orientation (number of socially-oriented decisions in the Triple Dominance Measure, SVO).

2.2.2. Stress Induction: The Maastricht Stress Test

To induce psychosocial stress, we used the Maastricht Acute Stress Test (MAST; Smeets et al., 2012, for instructions see supplemental online material), a hybrid task that combines elements of social-evaluation, physiological stress and uncontrollability. In the stress condition, participants were instructed to alternate between immersing their hand in a 0-2 °C ice water and performing a mental arithmetic task for a purportedly undefined period of time while being socially evaluated and videotaped. The experimenters wore lab coats and behaved in a rigid and non-responsive manner. The actual stress-induction always took 10 minutes, and was preceded by 5 minutes of preparation time. As in the original protocol by Smeets and colleagues (2012), overall hand immersion time summed to 6 minutes.

In the control condition, participants immersed their hand into 36 °C warm water for a total of 6 minutes, no camera recordings were made, and they were asked to count loudly and repeatedly from 1 to 25 upwards. The experimenters behaved friendly and wore no lab coats. Interval durations were equal to the stress condition.

2.2.3. Physiological and subjective stress measures

The neuroendocrine response to stress is complex, non-linear and time-dependent. In brief, organisms respond to acute stress with a rapid release of catecholamines, primarily noradrenaline through the sympathetic nervous system and a slower release of glucocorticoids (mainly cortisol in humans) as the end product of the hypothalamic–pituitary–adrenal (HPA) axis (Joels & Baram, 2009). During and shortly after stress, the physiological effects of cortisol and noradrenaline on social cognition and behavior are characterized by overlapping, combined action of non-genomic cortisol and catecholamines, followed by non-genomic cortisol action alone minutes afterwards, and finally by genomic cortisol effects that develop several hours later (Joels et al., 2011; Hermans et al., 2014).

The MAST has been shown to reliably induce physiological and psychosocial stress in participants, and stimulate the HPA axis as well as the sympathetic nervous system (Smeets et al., 2012). To quantify the intensity of stress-induction in our participants, we collected saliva samples and heart rate measures over the course of the experiment (see Fig. 1).

Saliva samples (Salivette, Sarstedt, Nürnbrecht, Germany) were analyzed for the physiological stress-markers cortisol (CORT) and salivary α-amylase (sAA), an indirect marker of sympathetic activity (Nater and Rohleder, 2009). Participants were asked to place the cotton swab into their mouth for one minute until it soaked with saliva. We collected two baseline samples before the MAST and three samples after the MAST (for timing see Fig. 1). Samples were stored at −26 °C until dispatch, and analyzed with a commercial competitive enzyme immunoassay (cortisol, IBL, Hamburg) or an enzyme liquid-phase assay (sAA). Analyses were performed by LabService Dresden GmbH.

Heart rate (HR) is regarded as a reliable and temporally well-resolved marker of sympathetic activity. We used commercial wristband photoplethysmographs (Polar A370) to make two 3-minute baseline HR recordings before MAST onset, as well as a continuous HR recording during the entire duration of the MAST.

Common stress-induction procedures, such as the MAST, come along with an increase in subjective arousal, tension, and feelings of insecurity. To capture how the MAST (vs. the control procedure) produced such feelings through the experiment, we administered the Positive and Negative Affect Scale (PANAS, Watson & Tellegen, 1989) twice before and twice after the stress-induction procedure. Participants additionally indicated feelings of stress, self-confidence, insecurity and shame on visual analogue scales (VAS, see e.g. Hellhammer and Schubert, 2012) every time the PANAS was delivered.

2.2.4. Social Discounting Task

To elicit social distance representations, we asked participants to imagine 100 people of their social environment on a hypothetical social distance scale, where 1 represents a person they feel closest to and 100 represents a random stranger whom they have never met (Jones and Rachlin, 2006; Margittai et al., 2018, 2015; Strombach et al., 2015). Participants were then asked to write down the names of people who represent social distance levels 1, 2, 3, 5, 10, 20. For social distances 50 and 100, participants were asked to imagine somewhat familiar, or completely unknown strangers, respectively. Also, participants were instructed to not select people they resent.

To assess how the framing of the decision problem moderated social discounting, we used an adapted variant of the dictator game (Archambault et al., 2019; Margittai et al., 2018, 2015) with a give and a take frame condition (Sellitto et al., 2019, see Fig. 2; Bardsley, 2008). In both conditions (see Fig. 2), participants decided how to allocate a monetary endowment between themselves and another person. In each trial in the *give* frame condition, participants received an endowment of either 13EUR, 15EUR or 17EUR, and decided how much to give to one individual they had assigned to a given social distance level. This was repeated for all three endowment levels and all eight social distance levels (1, 2, 3, 5, 10, 20, 50 or 100) in a repeated measures design. In each trial in the *take* frame condition, participants were informed that another individual on social distance level 1, 2, 3, 5, 10, 20, 50 or 100 had received an endowment of 13EUR, 15EUR or 17EUR, and they

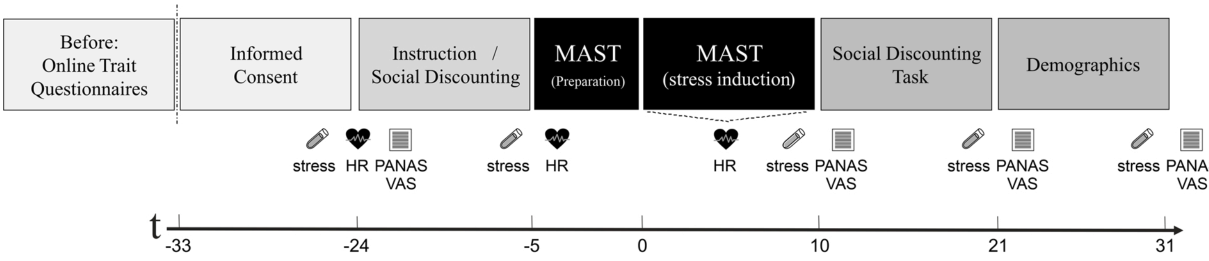
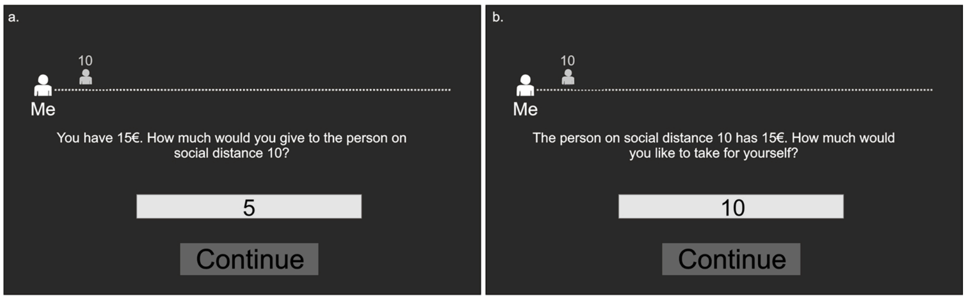


Fig. 1. *Timeline of the experimental procedure.* The x-axis depicts the time of events in minutes, relative to MAST onset. After an initial screening via phone interview, participants completed the online survey. Upon arrival in the laboratory, participants gave informed consent and were introduced to the HR-monitor and the handling of the saliva samples. We then collected the 1st saliva sample and recorded baseline heart rate for three minutes. The instructions for the social discounting task were presented, after which the participants' understanding of the task was tested using a short list of items related to the task's payoff structure. After collecting the 2nd saliva sample and recording another baseline HR signal, a standardized five-minute introduction to the MAST followed. During the MAST, HR was recorded continuously for 10 minutes. Upon completion of the MAST, a 3rd saliva sample was collected, directly segueing into the social discounting task. This was followed by completion of a demographic questionnaire, interspersed with the 4th and 5th saliva sample. All sessions took place between 14:00 and 18:00. The experiment ended with a debriefing. Subjects were paid based on one randomly chosen trial.



You have 15€. How much would you give to the person on social distance 10?

The person on social distance 10 has 15€. How much would you like to take for yourself?

Fig. 2. *Two exemplary trials in the give and the take frame.* Panel a. shows a trial in the give frame. The participant was endowed with 15€ and chose how much to share with another person on the indicated social distance scale. In this example, the participant decided to share 5€ (33.3%) with the person assigned to social distance 10. Panel b. shows a trial in the take frame. The other person on social distance 10 had an endowment of 15€. The participant chose to take 10€ (66.6%) for herself. Give and take frames were identical with respect to final payoff distributions. Participants were explicitly instructed that the other persons would only be informed about the outcome of the share, but not about their initial endowment, or the loss of it.

decided how much money to take away from the other for themselves. To prevent semantically induced choice biases, we explained the payoff contingencies in neutral terms, and strictly avoided negatively connoted terms like “remove”, “withdraw” or “take away” in the instructions and the task itself. Importantly, participants were explicitly informed that the other person was unaware of their initial endowment, and would, hence, not learn about the potential loss of it. Participants were also specifically advised that the other persons had no prior knowledge about the experiment and thus no outcome expectations; all that mattered was the final payoff allocation. All in all, each participant performed 48 trials (8 social distance levels x give/take frame x 3 endowments), presented in a pseudorandom, interleaved order. We used no deception and the task was incentive-compatible: after the experiment, one trial was randomly selected and paid out to the participant (which was added to their show-up fee) and the respective recipient (see Margittai et al. 2015). For social distance levels 50 and 100, which represented unknown to the participant, we gave the respective amount to a random stranger on the university campus.

2.3. Procedure

The procedure is outlined in Fig. 1. Upon arrival, participants signed the informed consent form. The first of two 3-minute baseline HR-recordings was made, followed by the collection of the first saliva sample and PANAS. Hereafter, participants received task instructions via laptop computer and they provided names of individuals in their social environment representing the different social distances. Using a number of comprehension items, we made sure that participants understood the task. Then, we obtained a second HR baseline recording, and took a PANAS mood questionnaire along with another saliva sample. The participant was directed into a different room to perform the MAST followed by the third saliva sample and PANAS. Participants then performed the social discounting framing task, which took no longer than 10 minutes. The fourth saliva sample was collected after task completion. Afterwards, as a manipulation check, participants were asked again about the individuals they had allocated to the different social distance levels to confirm they still remembered, and they finally completed a demographic questionnaire. Then, they were debriefed and received a fixed amount of 15EUR for participation plus the payoff of one randomly chosen trial in the social discounting framing paradigm. The money endowed by the participant to an individual in their social environment was delivered by regular mail, or given to a random person on campus for social distance levels 50 and 100, as described above.

2.4. Design and Statistics

The group or individual social discount functions are often approximated by a hyperbolic model (Jones & Rachlin, 2006; Margittai et al., 2015, 2018; Strombach et al., 2014, 2015; Strang et al., 2017; Vekaria et al., 2017). However, hyperbolic fitting procedures require variance in choices, otherwise any fitting procedure will not converge, or it will yield non-interpretable parameter estimates. This was the case in our results for a large number of choices, where in the take frame, participants often decided not to deduct money from the other person, leading to the exclusion of a large and condition-asymmetric number of participants. To circumvent this problem, we adopted a different analysis approach, resembling the one used by Archambault and colleagues (2019): we linearized the social discount function by rank-transforming social distance levels, i.e., the social distance levels 1, 2, 3, 5, 10, 20, 50 and 100 were replaced by social distance ranks 1 through 8, allowing analysis with a mixed linear model. In order to capture the effects of stress and framing on the discount rates, we regressed the factors *frame* (give vs. take frame), *stress* (stress vs. control), and the *ranked social distance level* (1-8) on trial-by-trial amounts shared with the other individual (i.e., the monetary amount given to the other in the give frame, or the amount left to the other in the loss frame). We allowed intercepts to vary for each *endowment level* (13, 15 and 17 EUR), and for each participant. We furthermore maximized the random effects structure as suggested by Barr, Levy, Scheepers and Tily (2013) and Matuschek, Kliegl, Vasishth, Baayen and Bates (2017), but only adding *frame* as a random coefficient yielded a non-degenerate, non-singular, properly convergent model. We then used backward model selection (using the *step* function from the R package lmerTest by Kuznetsova et al., 2017) to assess if any of the fixed effects, particularly the interaction terms, was redundant, but none of them was eliminated (see supplementary online materials for stepwise model comparisons). We also tested if the same results can be obtained when the ranked social distances were modelled as a categorical predictor (which we confirmed, see supplemental online materials). We predominantly used the R(3.6.1)-Packages *afex* (Singmann, in press) and *emmeans* (Lenth, 2018) for analysis, and we always tested type 3 sum of squares.

To test whether our stress induction was successful, we assessed changes in CORT, sAA, the participants’ HR, as well as participants’ positive and negative affect and subjective stress ratings (measured with PANAS and VAS) over the time course of the experiment. HR recordings were mean-aggregated by participant and recording, resulting in three values per participant. Before analysis, we inspected the distribution of CORT, sAA and HR and assessed normality using qq-plots (see supplemental online materials). Subsequently, we performed (natural-) log-transformations upon CORT, sAA and HR. We used mixed ANOVAs with the within-subject factor *sample* (for saliva) or *time point of measurement* (for HR or PANAS, see Fig. 1 for exact time points) and *group* (stress vs. non-stress group). Significant changes due to stress should primarily occur directly after the MAST, which we tested using simple effects analyses (see supplemental online materials). For the sake of brevity and readability, we only report relevant effects in this article (in these cases, only the interaction terms; we refer to the supplemental online material for all results).

3. Results

3.1. Trait measures

To exclude the possibility that stressed and control participants differed in trait and baseline characteristics that could confound our results, we compared their trait measures using Welch's t-Tests or Mann-Whitney-U-tests, whichever applied. Results can be found in Table 1. Applying a Bonferroni-corrected alpha-level of $\alpha = 0.005$ ($\alpha = 0.05$ divided by 10 comparisons), we find no differences between stress and control group. Note that social desirability scores (SDS) differed significantly between stress and control participants when uncorrected. We therefore repeated our main analyses with the SDS score as covariate. Our results remained robust against inclusion of SDS score, suggesting that our stress effects on framing and social discounting cannot be explained by differences in social desirability (see supplemental materials).

3.2. Physiological and subjective stress measures

3.2.1. Saliva Samples: CORT and sAA

As a manipulation check, we examined group differences on log-transformed values of salivary CORT and sAA concentrations (see Fig. 3). Out of 505 samples in total, we lost 28 CORT samples and 54 sAA-samples, mostly due to insufficient saliva.

A 5 (timepoint) x 2 (stress vs control) mixed ANOVA revealed that the MAST provoked an increase in salivary log(CORT) in the stress group compared to the control group (time point x condition interaction $F(2.03, 171.27) = 29.41$, $p < .001$, $\eta^2 = .099$ see Fig. 1). Simple effect analyses illustrate that log(CORT) was already increased directly after MAST onset in the stress group (stress vs. control group in saliva sample 3 $t(169) = -4.640$, $p < .001$ Cohen's d = -.987, see supplemental online material for full summary).

By contrast, we found no significant increase in log(sAA) levels between the stress and control groups: None of the stress-related factors in a 5 (timepoint) x 2 (stress vs control) mixed ANOVA, and in particular no interaction term, reached significance (time point x condition interaction $F(2.27, 173.17) = .23$, $p = .833$, $\eta^2 < .001$). Note that the lack of an effect of stress on sAA might have resulted from low statistical power due to the high number of lost sAA samples (54 samples; see above). If the analysis is run in a mixed linear model, which is – at least to some extent - capable of handling missing data, the stress group showed significantly higher sAA concentration at timepoint 5 $t(194) = -2.128$, $p = .035$, see supplemental online material for the full analysis). Thus, although caution is required, there is some indication that stress increased sAA in our task.

3.2.2. Heart Rate

A further indicator of sympathetic activation is the heart rate (HR) response (see Fig. 3). A 2 (stress vs control) x 3 (timepoint) mixed ANOVA with individual log-transformed mean HRs for the two baseline recordings and the mean of the log(HR) recording during the MAST shows a stress-related increase in heartbeats per minute for the stress group, but not the control group (recording time point x condition, $F(1.84, 189.5) = 8.61$, $p < .001$, $\eta^2 = .010$ see supplemental online materials for simple effects).

3.2.3. Subjective Stress Measures

Participants in the stress condition indicated stronger feelings of arousal, insecurity, shame and stress after the stress induction. Ratings in both the negative and the positive scale of the PANAS were increased after stress-induction (Positive: time x condition interaction; $F(2.09, 207.33) = 8.49$, $p < .001$, $\eta^2 = .035$; Negative: time x condition interaction $F(2.1, 207.7) = 9.06$, $p < .001$, $\eta^2 = .029$), with the former being most likely mediated by subjective arousal. Moreover, visual analogue scales revealed more feelings of insecurity (interaction time point x condition $F(2.88, 273.83) = 9.75$, $p < .001$, $\eta^2 = .03$), stress (interaction time point x condition $F(2.53, 240.51) = 19.56$, $p < .001$, $\eta^2 = .051$), shame (interaction time point x condition $F(2.19, 208.14) = 1.68$, $p = .187$, $\eta^2 = .008$; condition $F(1, 95) = 5.45$, $p = .022$,

Table 1
Baseline trait measures, age and BMI. To detect differences between stress and control participants, either Welch's t-Tests (t-statistics) or Mann-Whitney-U-Tests (W-statistics) were employed. Normality was examined using Shapiro-Wilk- Tests. Effect sizes are displayed on the right, either applying Cohen's d for t-tests or Cliff's Delta for Mann-Whitney-U tests.

	Stress (n = 50)		Control (n = 51)				
	M	SD	M	SD	Statistic	p-Value	Effect-Size
Impulsivity (BAR)	32.27	5.59	31.86	5.5	W = 1185.5	0.545	δ = -0.104
Behavioral activation (BAS)	40.06	4.32	41.38	4.96	t = 1.425	0.157	d = 0.235
Behavioral inhibition (BIS)	19.47	3.34	18.52	3.3	t = -1.438	0.154	d = -0.25
Social desirability (SDS)	22.41	2.62	21.08	2.51	W = 901.5	0.011	δ = 0.126
Chronic stress (TICS)	142.57	24.09	142.76	32.53	t = 0.034	0.973	d = -0.054
Psychopathy (LSRP)	53.88	10.5	50.68	8.67	t = -1.673	0.1	d = -0.383
Risk taking	2.06	1.1	1.92	1.03	W = 1188	0.534	δ = -0.092
Empathy (SPF-IRI)	40.75	5.71	41.04	5.01	t = 0.277	0.783	d = 0.034
Anxiety (STAI)	53.59	7.39	50.7	8.75	t = -1.791	0.077	d = -0.383
Chronotype (rMEQ)	12.2	4.41	12.02	3.71	W = 1297	0.884	δ = 0.048
Age	22.76	4.09	23.14	3.76	W = 1384	0.355	δ = 0.107
BMI	23.1	2.94	23.59	2.58	t = 0.961	0.339	d = 0.193

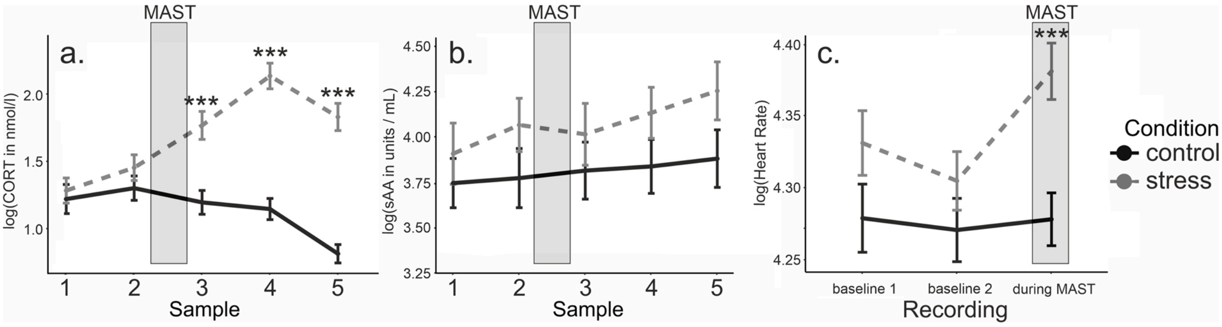


Fig. 3. *Physiological stress markers.* a. Log-transformed salivary CORT concentrations. The MAST procedure (administered as indicated by the grey bar) elicited a CORT-response in the stress, but not the control group. b. There was no significant difference in log-transformed sAA levels between stress- and control-group participants. Note that a significant difference emerged in sample 5 when analyzed using a mixed linear model (see text for details) c. Log-transformed heart rate was significantly higher in the stressed compared to control participants. No differences between the stress and control group were found at baseline-recording 1 and 2 (separate t-test: recording 1 t(97.731) = -1,600 at p = .113, Cohen's d = -.302; recording 2 t(98.723) = -1.085 at p = .280, Cohen's d = -0.216; simple effects: recording 1 t(124) = -1.744, p = .084; recording 2 t(124) = -1.261, p = .21). All error bars indicate ± 1 SEM.

$\eta^2 = .032$, yet, simple effects point to significant differences at timepoint 3) and less feelings of self-confidence (interaction time point x condition $F(2.38,\ 226.56) = 6.08$, $p =\ .001$, $\eta^2 = .013$; see appendix for simple effects) as a result of our stress induction.

In summary, despite the somewhat unclear effects of stress on sAA, all other physiological and psychological measures, including heart rate as a further marker of SAM activity, indicate success of our stress induction.

3.3. Stress diminishes the frame-effect on social discounting

Figs. 4A and 4B show the linearized social discount functions for the stress and control participants in both frames. In line with other studies on social discounting (Jones & Rachlin, 2006; Margittai et al., 2015, 2018; Strombach et al., 2015), the amount shared with others decreased monotonically across social distance in all treatment and frame conditions. In addition, participants overall shared more money in the take than the give frame, suggesting that the frame manipulation worked. More importantly, visual inspection of these figures suggests that there was no clear difference in social discounting between stress and control participants in the give frame. By contrast, in the take frame, the social discount function appeared flatter in control than stress participants, suggesting that non-stressed control participants were more generous in the take frame, in particular toward strangers, than their stressed counterparts.

To quantitatively assess how stress and framing modulated generosity in our social discounting task, we constructed a mixed linear model that regressed the main effects of stress (stress vs. control), frame (give vs. take), the ranked social distances and their interactions on the trial-by-trial amount shared with the other person, similar to the procedure done in Archambault et al. (2019). Furthermore, the model considers individual and item-specific variation by including varying intercepts for each subject and endowment level (13€, 15€, 17€) in the random effect structure. To maximize the random effect structure (Barr et al., 2013; Matuschek et al., 2017), we entered *frame* as a random coefficient varying per subject, which yielded better goodness-of-fit than an intercept-only model ($\chi^2(2) = 1811$, $p < .001$, see supplemental online materials for more information). Regression lines and standard errors are plotted in Fig. 4a and b, and treatment-coded regression estimates are displayed in Table 2. Below in this section, F-tests from an ANOVA-analysis will also be reported. The assumption of normality of residuals is met (see supplemental online materials).

Although main effects should be interpreted with caution in presence of higher order interaction terms, Table 2 shows a significant effect of social distance on amount shared as well as a significant interaction between social distance and frame, suggesting that generosity decreased

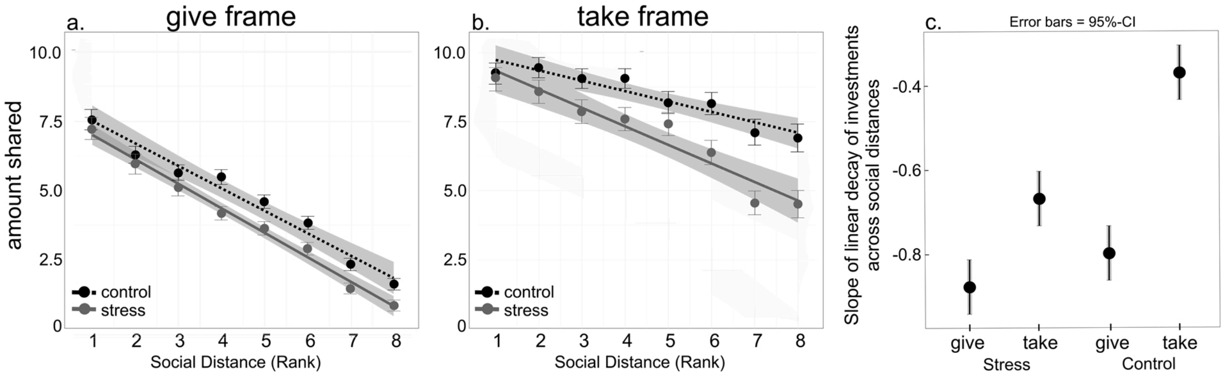


Fig. 4. Results of the Social Discounting Task. The left panel (a.) shows the mean amount shared across ranked social distances in the give frame (±standard error of the mean; SEM), as well as the fitted regression lines. Data are shown for stress and control participants separately. Panel b. shows the mean amount shared (±SEM) and regression lines of stress and control participants in the take frame. While there was no clear difference in social discounting between stressed and non-stressed control participants in the give frame (panel a), control participants showed flatter discounting than stressed participants in the take frame (panel b). This frame-dependent stress effect on generosity was most pronounced at large social distance levels, implying that stressed individuals had a diminished willingness to share with strangers in the take frame. Panel c. displays the slopes of the linearized social discount functions with 95% confidence intervals. There was a distinct difference in the slopes in the control participants, indicating flatter discounting in the take than the give frame, thus reflecting the frame effect on social discounting. This difference in the slope of the social discount function between frames was less pronounced in the stressed participants, confirming that stress diminished the frame effect on social discounting.

Table 2
Regression table for effect of stress, frame and ranked social distance on amounts shared. Effects are treatment-coded.

Regression Table / Main Model			
Predictors	Estimates	CI	p
(Intercept)	10.07	[8.57 : 11.57]	<0.001
Ranked Social Distance	−0.37	[-0.44 : -0.31]	<0.001
Stress	−0.05	[-1.79 : 1.68]	0.953
Frame	−1.77	[-2.95 : -0.58]	0.003
Ranked Social Distance * Stress	−0.3	[-0.39 : -0.21]	<0.001
Ranked Social Distance * Frame	−0.43	[-0.52 : -0.34]	<0.001
Stress * Frame	−0.35	[-2.01 : 1.32]	0.684
3-Way Interaction	0.22	[0.09 : 0.35]	0.001
σ^2	6.9		
$\tau_{00\ \text{subject}}$	18.35		
$\tau_{00\ \text{endowment}}$	0.57		
$\tau_{11\ \text{subject/frame}}$	15.48		
$\rho_{01\ \text{subject}}$	−0.82		
ICC	0.65		

All variables were treatment-Coded. References: Stress = Non-Stress, Frame = Take Frame

across social distance (i.e., social discounting) and that generosity decreased differently across social distance between frame conditions (i.e., frame effects on social discounting).

Most importantly, the results summarized in Table 2 show that the stress effect on social discounting was more pronounced in the take than the give frame; the three-way interaction between stress x frame x social distance on amount shared reached significance ($F(1, 4638.02) = 11.22$, $p < .001$), indicating that stress affected social-distance-dependent generosity differently in the take than the give frame. To break down this three-way interaction effect, we compared the regression slopes, capturing the decline in sharing across social distance, between the give and the take frames (see Fig. 4C); this was done separately for the stress and control participants. We found that the difference in regression slopes between the give and the take frames was more pronounced in non-stressed control participants compared to stressed participants (difference in give/take slopes in the control condition = 0.432, SE = 0.047, $t(4642) = 9.212$, $p < .001$; difference in give/take slopes in the stress condition = 0.211, SE = 0.046, $t(4642) = 4.547$, $p < .001$; overall difference of framing-related slope alterations between stress and control condition =. 0.221, SE = 0.066, $t(4642) = 3.348$, $p < .001$). Hence, while control participants showed flatter social discounting in the take than the give condition, indicating the frame manipulation on social discounting worked, this difference in the steepness of social discounting between frame conditions was less evident in stressed participants. In line with these findings, an analogous mixed linear regression model that included social distance as a categorical variable revealed that the stress x frame interaction only occurred at ranked social distances 7 and 8 (see supplemental online materials).

Further simple slope analyses on the three-way interaction revealed that stress affected the social discount rates primarily in the take frame, and much less so in the give frame (difference in stress/control slopes in the give frame = 0.080, SE = 0.047, $t(4642) = 1.706$, $p = .088$; estimated slope difference in the take frame = 0.300, SE = 0.047, $t(4642) = 6.444$, $p < .001$; difference of stress-related slope between take and give frame = .221 SE = .066, $t(4642) = 3.348$, $p < .001$). Taken together, this analysis supports above mentioned observation that, compared to non-stressed participants, stressed participants were selectively less generous towards strangers, but this effect was found only in the take, not in the give frame condition.

4. Discussion

People are generous, but their generosity typically decreases across social distance to the recipient of help (Jones & Rachlin, 2006; Strombach et al., 2015). We recently showed that framing a financial allocation decision as the prevention of another person's loss, rather than granting them a gain, strongly decreased the social discount rate, implying that the mere description of a decision problem can serve as a nudge to render participants much more generous towards strangers (Sellitto et al., 2019). Here, we asked if an acute social-evaluative and physical hybrid stressor amplifies or attenuates the frame effect on social discounting. We devised a task in which participants decided to share an endowment with other individuals at variable social distances (give frame), or decided to take away money from the endowment of the other individuals (take frame). We, first, replicated our previous finding (Sellitto et al., 2019) that participants exhibited flatter discounting in the take than the give frame, suggesting higher generosity toward strangers in the take frame. Importantly, our stress manipulation revealed support for the frame-attenuation hypothesis: we found that acute stress diminished the frame effect on social discounting and caused stressed participants to be equivalently generous towards others in the take and the give frames. Furthermore, in the take, but not the give frame, stressed individuals were less generous toward strangers than non-stressed controls.

Our finding of a stress-related decrease in generosity towards strangers in the take frame blends with other results demonstrating diminished willingness to share resources under stress (Starcke et al., 2011; Steinbeis et al., 2015; Vinkers et al., 2013), and, in a broader sense, they are consistent with the observation of increased egocentric, antagonistic tendencies under stress (Agnew, 2005; Sandi and Haller, 2015; Silver and Teasdale, 2005), as hypothesized by the fight-or-flight theory almost a century ago (Cannon, 1932). But what causes the stress-related diminution of the frame effect on social discounting? We recently argued that the frame effect on social discounting is the result of people's internalized hesitation to transgress the social norm of preventing harm to others, and the associated feelings of guilt and shame if they do (Sellitto et al., 2019); that is, people generally follow the "do-no-harm principle" (Baron, 1995). This means that, even though it might be socially acceptable to not share money with others in the give frame, social norms strongly prohibit taking away money from others in the take frame. Because stress is known to erode social norm compliance (Starcke et al., 2011; Steinbeis et al., 2015; Vinkers et al., 2013), social decision making of stressed individuals will likely be less influenced by social norms than that of non-stressed people, and they would consequently be less hesitant to cause financial harm to others where social norms would normally forbid doing so. This explanation can account for the fact that the frame effect on sharing behavior was most pronounced when dealing with unknown strangers: social norms guide social behavior especially towards others at larger social distance levels, but they are less relevant for generous behavior towards friends and family where people are often naturally selfless anyway, independent of social norm prescriptions (Rand et al., 2014; Strombach et al., 2015).

Notably, our finding of reduced prosociality after stress in the take frame stands in contrast with results from other experiments, including studies from us, that have shown the opposite pattern of *increased* prosociality after stress (tend-and-befriend; Berger et al., 2016; Margittai et al., 2018, 2015; Singer et al., 2017; Sollberger et al., 2016; Tomova et al., 2014; Von Dawans et al., 2019, 2012; Youssef et al., 2018). Hence, the question remains why our participants did not respond to stress with increased prosociality. It is currently unclear when and why stressed individuals show a stronger or reduced prosocial stress-response. We have recently argued that stress does not provoke *either* fight-or-flight *or* tend-and-befriend (or decreased vs. increased prosociality by proxy), but it may boost *both* tendencies at the same time (Schweda et al., 2019), as outlined in the following. Tend-and-befriend is an alleged coping strategy where stressed individuals invest into their social network in order to receive help and comfort in return (Taylor, 2006; Taylor et al., 2000). Accordingly, we have argued, and shown, that tend-and-befriend behavior is predominantly directed towards socially close others, who, unlike strangers, can potentially provide comfort and support in stressful times (Margittai et al., 2018, 2015). By contrast, fight-or-flight is a strategy primarily aimed at escaping the acute stressor. Acute stress can be expected to originate from socially distant outgroup members who are more likely to present a threat than socially close friends and family, especially at times of conflict. Consequently, aggressive, antagonistic tendencies should be largely aimed at socially distant strangers, but less so at socially close others. Even though evidence for this theory is, admittedly, still scarce (Schweda et al., 2019), our current finding of a stress-related and frame-dependent reduction in generosity towards strangers, but not socially close others, is consistent with this hypothesis.

Further questions remain. We could not replicate our previous findings that socio-evaluative stress (Margittai et al., 2015), or exogenous psychopharmacological challenges aimed at mimicking the natural endocrine response to stress (Margittai et al., 2018), selectively increased generosity towards socially close others. One possible explanation for the discrepancy in findings is the difference in the stressor used between this and our previous (Margittai et al., 2015) study; while, here, we employed the MAST (see methods), we used the group version of the Trier Social Stress Test in our previous experiment. The type of stressor matters as it has been shown to alter social behavior in several experiments. For example, according to von Dawans et al. (2018) while physical and psychosocial stressors alone impair prosocial behavior, the two combined actually restore prosociality. Though this finding is not compatible with our results, as we found less prosociality under a combined physical and psychosocial stress induction procedure, this example nonetheless illustrates the complexity of the relationship between social behavior, social norm compliance and stress. Another possibility to account for the differences in results between the current study and our previous work is the task used to elicit social preferences. The frame version of the social discounting task is more complex and procedurally different to the simple social discounting task used before (Margittai et al., 2018, 2015). Thus, task performance might not be perfectly translatable between tasks, and within-task spill-over effects are to be expected. Whatever the reason for the divergence in results, definitive conclusions about underlying mechanisms of our effects cannot be made with certainty at present. Our results pave the way for future research investigating the frame effect and its interaction with stress in shaping prosocial behaviors.

The current study involved male participants only. Following research needs to clarify whether framing and stress interactions on social discounting occur in women, too. Gender differences in social frame effects have been found before (Chowdhury et al., 2017; Ellingsen et al., 2013; Strombach et al., 2016), and we know the stress response is susceptible to variations in sex hormone concentrations (Kirschbaum et al., 1999; Kudielka and Kirschbaum, 2005). Furthermore, stress is multidimensional (Joëls and Baram, 2009), and we still lack knowledge of which mediators of the stress-response affect social behavior. Pharmacological intervention studies using, for instance, corticosteroids and adrenergic drugs, as well as conditions with time lags between stressor and task would be the optimal choice for future studies (e.g. Margittai et al., 2018, 2015; Vinkers et al., 2013). In the present study, analyses considering the involvement of specific biomarkers have only yielded non-significant results (see supplemental online materials).

In summary, our study replicates our previous findings that participants prefer more generous resource allocations to strangers when donations are framed as preventing financial harm to others. We demonstrate that stress mitigates this frame effect on social discounting, so that stressed participants are less generous towards strangers than non-stressed controls. This finding can be tentatively explained as a stress-induced diminished compliance to the social norm to "do-no-harm".

These findings contribute to our understanding of how acute stress alters decision making, social norm compliance and interpersonal harm avoidance. Thus, our study broadens our understanding of the impact our psychological state has for our everyday moral and social behavior. We believe that this result is not only relevant for cognitive scientists studying the effects of stress on cognition and behavior, it is also important for policy makers and corporate decision makers; knowing under which circumstances stress boosts or corrupts prosociality, especially towards strangers, has practical implications for charity advocacy about the way charity calls or appeals for donations could be worded.

Declaration of Competing Interest

We have no conflict of interest to declare.

Consent for publication

Not applicable.

Financial and Non-Financial Competing interest

We have no conflict of interest to declare.

Statement of responsibility

AS designed the task, ran the data collection, analyzed the data and wrote the paper, TK designed the task, wrote the paper, provided consultation at all stages of the project and funded the project, Z.M edited the paper and provided consultation at all stages of the project.

Ethics approval and consent to participate

The study was approved by the ethics committee of the Medical Faculty of the Heinrich-Heine-University in Dusseldorf. All participants gave their informed consent.

Funding

The study was supported by a grant from the Deutsche Forschungsgemeinschaft (Bonn, Germany) to TK (grant-ID: DFG-KA 2675/4-3).

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.psyneuen.2020.104860.

References

Agnew, R., 2005. Why do criminals offend?: A general theory of crime and delinquency. Roxbury, Los Angeles, CA.

Archambault, C., Kalenscher, T., De Laat, J., 2019. Generosity and livelihoods: Dictator Game evidence on the multidimensional nature of sharing among the Kenyan Maasai. J Behav Dec Makin.

Bardsley, N., 2008. Dictator game giving: altruism or artefact? Experimental Economics 11, 122–133.

Baron, J., 1995. Blind justice: Fairness to groups and the do-no-harm principle. Journal of Behavioral Decision Making 8, 71–83. https://doi.org/10.1002/bdm.3960080202.

Barr, D.J., Levy, R., Scheepers, C., Tily, 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language 68 (3), 255–278.

Baumeister, R.F., Stillwell, A.M., Heatherton, T.F., 1994. Guilt: an interpersonal approach. Psychological bulletin 115, 243.

Berger, J., Heinrichs, M., von Dawans, B., Way, B.M., Chen, F.S., 2016. Cortisol modulates men's affiliative responses to acute social stress. Psychoneuroendocrinology 63, 1–9. https://doi.org/10.1016/j.psyneuen.2015.09.004.

Bolton, G.E., Katok, E., Zwick, R., 1998. Dictator game giving: Rules of fairness versus acts of kindness. Game Theory 27, 269–299. https://doi.org/10.1007/s001820050072.

Cannon, W.B., 1932. Natural Defenses of the Organism. The Wisdom of the Body. W W Norton & Co, New York, NY, US, pp. 213–215.

Carver, C.S., White, T.L., 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. Journal of Personality and Social Psychology 67, 319–333. https://doi.org/10.1037/0022-3514.67.2.319.

Chang, L.J., Sanfey, A.G., 2013. Great expectations: neural computations underlying the use of social norms in decision-making. Soc Cogn Affect Neurosci 8, 277–284. https://doi.org/10.1093/scan/nsr094.

Chang, L.J., Smith, A., Dufwenberg, M., Sanfey, A.G., 2011. Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion. Neuron 70, 560–572. https://doi.org/10.1016/j.neuron.2011.02.056.

Chowdhury, S.M., Jeon, J.Y., Saha, B., 2017. Gender Differences in the Giving and Taking Variants of the Dictator Game. Southern Economic Journal 84, 474–483. https://doi.org/10.1002/soej.12223.

Crockett, M.J., Kurth-Nelson, Z., Siegel, J.Z., Dayan, P., Dolan, R.J., 2014. Harm to others outweighs harm to self in moral decision making. PNAS 111, 17320–17325. https://doi.org/10.1073/pnas.1408988111.

De Dreu, C.K.W., McCusker, C., 1997. Gain–take frames and cooperation in two-person social dilemmas: A transformational analysis. Journal of Personality and Social Psychology 72, 1093–1106. https://doi.org/10.1037/0022-3514.72.5.1093.

Dedovic, K., Duchesne, A., Andrews, J., Engert, V., Pruessner, J.C., 2009. The brain and the stress axis: the neural correlates of cortisol regulation in response to stress. Neuroimage 47, 864–871. https://doi.org/10.1016/j.neuroimage.2009.05.074.

Ellingsen, T., Johannesson, M., Mollerstrom, J., Munkhammar, S., 2013. Gender differences in social framing effects. Economics Letters 118, 470–472. https://doi.org/10.1016/j.econlet.2012.12.010.

Evans, A.M., van Beest, I., 2017. Gain-loss framing effects in dilemmas of trust and reciprocity. Journal of Experimental Social Psychology 73, 151–163. https://doi.org/10.1016/j.jesp.2017.06.012.

Gallo, S., Paracampo, R., Müller-Pinzler, L., Severo, M.C., Blömer, L., Fernandes-Henriques, C., Gazzola, V., 2018. The causal role of the somatosensory cortex in prosocial behavior. Elife 7, e32740.

Haller, J., 2018. The role of central and medial amygdala in normal and abnormal aggression: A review of classical approaches. Neuroscience & Biobehavioral Reviews 85, 34–43.

Haller, J., Makara, G.B., Kruk, M.R., 1998. Catecholaminergic involvement in the control of aggression: hormones, the peripheral sympathetic, and central noradrenergic systems. Neurosci Biobehav Rev 22, 85–97.

Hatzenbuehler, M.L., McLaughlin, K.A., 2014. Structural stigma and hypothalamic–pituitary–adrenocortical axis reactivity in lesbian, gay, and bisexual young adults. Annals of Behavioral Medicine 47 (1), 39–47.

Hellhammer, J., Schubert, M., 2012. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. Psychoneuroendocrinology 37, 119–124. https://doi.org/10.1016/j.psyneuen.2011.05.012.

Herhaus, B., Petrowski, K., 2018. Cortisol stress reactivity to the Trier social stress test in obese adults. Obesity facts 11 (6), 491–500.

Hermans, E.J., Henckens, M.J., Joels, M., Fernández, G., 2014. Dynamic adaptation of large-scale brain networks in response to acute stressors. Trends in neurosciences 37 (6), 304–314.

Ishii, K., Eisen, C., 2018. Cultural Similarities and Differences in Social Discounting: The Mediating Role of Harmony-Seeking. Front Psychol 9. https://doi.org/10.3389/fpsyg.2018.01426.

Ispano, A., Schwardmann, P., 2017. Cooperating over losses and competing over gains: A social dilemma experiment. Games and Economic Behavior 105, 329–348. https://doi.org/10.1016/j.geb.2017.08.008.

Jansen, A.S.P., Nguyen, X.V., Karpitskiy, V., Mettenleiter, T.C., Loewy, A.D., 1995. Central Command Neurons of the Sympathetic Nervous System: Basis of the Fight-or-Flight Response. Science 270, 644–646. https://doi.org/10.1126/science.270.5236.644.

Joels, M., Baram, T.Z., 2009. The neuro-symphony of stress. Nature reviews neuroscience 10, 459.

Joels, M., Fernandez, G., Roozendaal, B., 2011. Stress and emotional memory: a matter of timing. Trends in cognitive sciences 15 (6), 280–288.

Jones, B., Rachlin, H., 2006. Social discounting. Psychol Sci 17, 283–286. https://doi.org/10.1111/j.1467-9280.2006.01699.x.

Kalenscher, T., 2017. Social psychology: Love thy stranger as thyself. Nature Human Behaviour 1, 0108. https://doi.org/10.1038/s41562-017-0108.

Kirschbaum, C., Kudielka, B.M., Gaab, J., Schommer, N.C., Hellhammer, D.H., 1999. Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. Psychosom Med 61, 154–162. https://doi.org/10.1097/00006842-199903000-00006.

Kruk, M.R., Halász, J., Meelis, W., Haller, J., 2004. Fast positive feedback between the adrenocortical stress response and a brain mechanism involved in aggressive behavior. Behav. Neurosci. 118, 1062–1070. https://doi.org/10.1037/0735-7044.118.5.1062.

Kudielka, B.M., Kirschbaum, C., 2005. Sex differences in HPA axis responses to stress: a review. Biological Psychology 69, 113–132. https://doi.org/10.1016/j.biopsycho.2004.11.009.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed models. Journal of statistical software 82 (13).

Levenson, M.R., Kiehl, K.A., Fitzpatrick, C.M., 1995. Assessing psychopathic attributes in a noninstitutionalized population. J Pers Soc Psychol 68, 151–158.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., Bates, D., 2017. Balancing Type I error and power in mixed linear models. Journal of Memory and Language 94, 305–315.

Margittai, Z., Strombach, T., van Wingerden, M., Joëls, M., Schwabe, L., Kalenscher, T., 2015. A friend in need: Time-dependent effects of stress on social discounting in men. Hormones and Behavior 73, 75–82. https://doi.org/10.1016/j.yhbeh.2015.05.019.

Margittai, Z., van Wingerden, M., Schnitzler, A., Joëls, M., Kalenscher, T., 2018. Dissociable roles of glucocorticoid and noradrenergic activation on social discounting. Psychoneuroendocrinology 90, 22–28. https://doi.org/10.1016/j.psyneuen.2018.01.015.

McCarty, R., 2016. The Fight-or-Flight Response: A Cornerstone of Stress Research. In: Fink, George (Ed.), Stress: Concepts, Cognition, Emotion, and Behavior: Handbook of Stress Series. Academic Press.

Meule, A., Vögele, C., Kübler, A., 2011. Psychometrische Evaluation der deutschen Barratt Impulsiveness Scale – Kurzversion (BIS-15). Diagnostica 57, 126–133. https://doi.org/10.1026/0012-1924/a000042.

Nater, U.M., Rohleder, N., 2009. Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: Current state of research. Psychoneuroendocrinology 34, 486–496. https://doi.org/10.1016/j.psyneuen.2009.01.014.

Nickels, N., Kubicki, K., Maestripieri, D., 2017. Sex differences in the effects of psychosocial stress on cooperative and prosocial behavior: evidence of 'fight or flight' in males and 'tend and befriend' in females. Adaptive Human Behavior and Physiology 3 (2), 171–183.

Parra, L.A., Benibgui, M., Helm, J.L., Hastings, P.D., 2016. Minority stress predicts depression in lesbian, gay, and bisexual emerging adults via elevated diurnal cortisol. Emerging Adulthood 4 (5), 365–372.

Paulus, C., 2009. Der Saarbrücker Persönlichkeitsfragebogen SPF(IRI) Zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index.

Rammstedt, B., 2007. The 10-item big five inventory. European Journal of Psychological Assessment 23, 193–201.

Rand, D.G., Peysakhovich, A., Kraft-Todd, G.T., Newman, G.E., Wurzbacher, O., Nowak, M.A., Greene, J.D., 2014. Social heuristics shape intuitive cooperation. Nature Communications 5, 3677. https://doi.org/10.1038/ncomms4677.

Randler, C., 2013. German version of the reduced Morningness–Eveningness Questionnaire (rMEQ). Biological rhythm research 44, 730–736.

Rodrigues, S.M., LeDoux, J.E., Sapolsky, R.M., 2009. The influence of stress hormones on fear circuitry. Annu. Rev. Neurosci. 32, 289–313. https://doi.org/10.1146/annurev.neuro.051508.135620.

Sandi, C., Haller, J., 2015. Stress and the social brain: behavioural effects and neurobiological mechanisms. Nature Reviews Neuroscience 16, 290–304. https://doi.org/10.1038/nrn3918.

Schulz, P., Schlotz, W., 1999. The Trier Inventory for the Assessment of Chronic Stress (TICS). Scale construction, statistical testing, and validation of the scale work overload. Diagnostica 45, 8–19.

Schweda, A., Faber, N.S., Crockett, M.J., Kalenscher, T., 2019. The effects of psychosocial stress on intergroup resource allocation. Scientific Reports. 9 (1), 1–12.

Sellitto, M., Schweda, A., Weber, B., Kalenscher, T., 2019. To give or not to take – insular cortex subserves framing-induced boosts in generosity during social discounting. bioRxiv 841338. https://doi.org/10.1101/841338.

Sgoifo, A., De boer, S.F., Haller, J., Koolhaas, J.M., 1996. Individual Differences in Plasma Catecholamine and Corticosterone Stress Responses of Wild-Type Rats: Relationship With Aggression. Physiology & Behavior 60, 1403–1407. https://doi.org/10.1016/S0031-9384(96)00229-6.

Sgoifo, A., Papi, F., 1995. Effects of social and non-social acute Stressors on plasma levels of cate-cholamines and corticosterone in wild rats. Rend. Fis. Acc. Lincei 6, 289. https://doi.org/10.1007/BF03001690.

Silver, E., Teasdale, B., 2005. Mental Disorder and Violence: An Examination of Stressful Life Events and Impaired Social Support. Soc Probl 52, 62–78. https://doi.org/10.1525/sp.2005.52.1.62.

Singer, N., Sommer, M., Döhnel, K., Zänkert, S., Wüst, S., Kudielka, B.M., 2017. Acute psychosocial stress and everyday moral decision-making in young healthy men: The impact of cortisol. Hormones and Behavior 93, 72–81. https://doi.org/10.1016/j.yhbeh.2017.05.002.

Smeets, T., Cornelisse, S., Quaedflieg, C.W.E.M., Meyer, T., Jelicic, M., Merckelbach, H., 2012. Introducing the Maastricht Acute Stress Test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. Psychoneuroendocrinology 37, 1998–2008. https://doi.org/10.1016/j.psyneuen.2012.04.012.

Sollberger, S., Bernauer, T., Ehlert, U., 2016. Stress influences environmental donation behavior in men. Psychoneuroendocrinology 63, 311–319. https://doi.org/10.1016/j.psyneuen.2015.10.017.

Soutschek, A., Ruff, C.C., Strombach, T., Kalenscher, T., Tobler, P.N., 2016. Brain stimulation reveals crucial role of overcoming self-centeredness in self-control. Science Advances 2, e1600992. https://doi.org/10.1126/sciadv.1600992.

Spielberger, C.D., 1983. Manual for the State-Trait Anxiety Inventory STAI (Form Y) ("Self-Evaluation Questionnaire").

Starcke, K., Polzer, C., Wolf, O.T., Brand, M., 2011. Does stress alter everyday moral decision-making? Psychoneuroendocrinology 36, 210–219. https://doi.org/10.1016/j.psyneuen.2010.07.010.

Steinbeis, N., Engert, V., Linz, R., Singer, T., 2015. The effects of stress and affiliation on social decision-making: Investigating the tend-and-befriend pattern. Psychoneuroendocrinology 62, 138–148. https://doi.org/10.1016/j.psyneuen.2015.08.003.

Stöber, J., 2001. The Social Desirability Scale-17 (SDS-17). European Journal of Psychological Assessment 17, 222–232. https://doi.org/10.1027//1015-5759.17.3.222.

Strang, S., Gerhardt, H., Marsh, N., Artigas, S.O., Hu, Y., Hurlemann, R., Park, S.Q., 2017. A matter of distance—The effect of oxytocin on social discounting is empathy-dependent. Psychoneuroendocrinology 78, 229–232.

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I.I., Tobler, P.N., Kalenscher, T., 2015. Social discounting involves modulation of neural value signals by temporoparietal junction. PNAS 112, 1619–1624. https://doi.org/10.1073/pnas.1414715112.

Strombach, T., Margittai, Z., Gorczyca, B., Kalenscher, T., 2016. Gender-specific effects of cognitive load on social discounting. PloS one 11 (10).

Takahashi, T., 2013. The q-exponential social discounting functions of gain and loss. Applied Mathematics 4, 445.

Taylor, S.E., 2006. Tend and Befriend: Biobehavioral Bases of Affiliation Under Stress. Curr Dir Psychol Sci 15, 273–277. https://doi.org/10.1111/j.1467-8721.2006.00451.x.

Taylor, S.E., Klein, L.C., Lewis, B.P., Gruenewald, T.L., Gurung, R.A.R., Updegraff, J.A., 2000. Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. Psychological Review 107, 411–429. https://doi.org/10.1037/0033-295X.107.3.411.

Terbeck, S., Kahane, G., McTavish, S., Savulescu, J., Cowen, P.J., Hewstone, M., 2012. Propranolol reduces implicit negative racial bias. Psychopharmacology (Berl.) 222, 419–424. https://doi.org/10.1007/s00213-012-2657-5.

Terbeck, S., Savulescu, J., Chesterman, L.P., Cowen, P.J., 2016. Noradrenaline effects on social behaviour, intergroup relations, and moral decisions. Neurosci. Biobehav. Rev 66, 54–60. https://doi.org/10.1016/j.neubiorev.2016.03.031.

Tomova, L., von Dawans, B., Heinrichs, M., Silani, G., Lamm, C., 2014. Is stress affecting our ability to tune into others? Evidence for gender differences in the effects of stress on self-other distinction. Psychoneuroendocrinology 43, 95–104. https://doi.org/10.1016/j.psyneuen.2014.02.006.

Tomova, L., Saxe, M., Klöbl, M., Lanzenberger, R., Lamm, C., 2020. Acute stress alters neural patterns of value representation for others. NeuroImage 209, 116497.

Vekaria, K.M., Brethel-Haurwitz, K.M., Cardinale, E.M., Stoycos, S.A., Marsh, A.A., 2017. Social discounting and distance perceptions in costly altruism. Nature Human Behaviour 1, 0100. https://doi.org/10.1038/s41562-017-0100.

Vinkers, C.H., Zorn, J.V., Cornelisse, S., Koot, S., Houtepen, L.C., Olivier, B., Verster, J.C., Kahn, R.S., Boks, M.P.M., Kalenscher, T., Joëls, M., 2013. Time-dependent changes in altruistic punishment following stress. Psychoneuroendocrinology 38, 1467–1475. https://doi.org/10.1016/j.psyneuen.2012.12.012.

Von Dawans, B., Ditzen, B., Trueg, A., Fischbacher, U., Heinrichs, M., 2019. Effects of acute stress on social behavior in women. Psychoneuroendocrinology 99, 137–144. https://doi.org/10.1016/j.psyneuen.2018.08.031.

Von Dawans, B., Fischbacher, U., Kirschbaum, C., Fehr, E., Heinrichs, M., 2012. The social dimension of stress reactivity: acute stress increases prosocial behavior in humans. Psychological science 23, 651–660.

Von Dawans, B., Kirschbaum, C., Heinrichs, M., 2011. The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. Psychoneuroendocrinology 36, 514–522. https://doi.org/10.1016/j.psyneuen.2010.08.004.

von Dawans, B., Trueg, A., Kirschbaum, C., Fischbacher, U., Heinrichs, M., 2018. Acute social and physical stress interact to influence social behavior: The role of social anxiety. PloS one 13, e0204665.

White, S.F., Lee, Y., Phan, J.M., Moody, S.N., Shirtcliff, E.A., 2019. Putting the flight in "fight-or-flight": Testosterone reactivity to skydiving is modulated by autonomic activation. Biological Psychology 143, 93–102. https://doi.org/10.1016/j.biopsycho.2019.02.012.

Youssef, F.F., Bachew, R., Bissessar, S., Crockett, M.J., Faber, N.S., 2018. Sex differences in the effects of acute stress on behavior in the ultimatum game. Psychoneuroendocrinology 96, 126–131. https://doi.org/10.1016/j.psyneuen.2018.06.012.