Decision Time in Social Dilemmas –

Personality and Situational Factors Moderating Spontaneous Behavior in First and Second Order Public Good Games

Dissertation

for the award of the degree

"Doctor rerum naturalium" (Dr. rer. nat.)

of the Georg-August University Göttingen

within the doctoral program Biology

of the Georg-August University School of Science (GAUSS)

submitted by

Anna Dorothee Mischkowski

from Göttingen

Cologne, 2019

Thesis Committee

Prof. Dr. Andreas Glöckner

Social Psychology, University of Cologne

Until October 2015: Psychological Assessment, Judgment and Decision Making, Georg-August University of Göttingen

Prof. Dr. Stefan Schulz-Hardt

Economic and Social Psychology, Georg-August University of Göttingen

Members of the Examination Board

Reviewer: Prof. Dr. Andreas Glöckner

Social Psychology, University of Cologne

Second Reviewer: Prof. Dr. Stefan Schulz-Hardt

Economic and Social Psychology, Georg-August University of Göttingen

Further members of the Examination Board:

Prof. DDr. Peter Lewisch

Criminal Law and Criminal Procedural Law, University of Vienna

Prof. Dr. Annekathrin Schacht

Affective Neuroscience and Psychophysiology, Georg-August University of Göttingen

Prof. Dr. Hannes Rakoczy

Developmental Psychology, Georg-August University of Göttingen

Prof. Dr. Nivedita Mani

Psychology of Language, Georg-August University of Göttingen

Date of the oral examination: December 12th, 2019

Abstract

There is an ongoing discussion regarding the circumstances under which individuals seek to maximize the welfare of the common good rather than their own benefit – two motives that are contrasted in social dilemmas. In this context, it could be shown that cooperation behavior in social dilemmas decreases with decision time and, thus, represents a spontaneously expressed phenomenon. This finding has triggered substantial debate about the cognitive underpinnings of prosocial behavior in the fields of psychology and behavioral economics to which this thesis contributes with the scope of two articles: First, dispositional pro-sociality (i.e., social value orientation and Honesty-Humility) was identified as a moderator of spontaneous cooperation. Specifically, spontaneous cooperation was shown to be valid only for prosocial individuals – thus offering an explanation for heterogeneous replication results of the spontaneous cooperation effect. In turn, the second article explores whether spontaneous cooperation can be generalized to costly punishment behavior (also known as instrumental cooperation) in social dilemmas. Specifically, negative affect and social value orientation are investigated as potentially underlying motives of spontaneous punishment. Results show that spontaneous punishment – unlike spontaneous cooperation – is not conducted by prosocials but rather retributively displayed by highly upset, above-average contributors. These results of the similar, spontaneously expressed behavior in social dilemmas and the accompanied motivations are critically discussed concerning their added value to the underlying theory of spontaneous cooperation as well as in light of a theory of spontaneous pro-sociality in general.

Zusammenfassung

Ob die Natur des Menschen ausmacht, intuitiv nach dem eigenen Nutzen zu streben oder diesen – als erste und spontane Reaktion – dem Gemeinwohl unterzuordnen, ist Bestandteil einer andauernden, interdisziplinär-wissenschaftlichen Debatte. Die Veröffentlichung des „spontanen Kooperationseffekts“ im Jahr 2012 von Rand, Greene und Nowak hat in der Psychologie wie Verhaltensökonomie eine Vielzahl an Replikations- und Folgestudien angestoßen. Die vorliegende Arbeit ist Bestandteil dieser Debatte und trägt dazu bei, die heterogene Befundlage verschiedenster Replikationsversuche zu erklären und die Generalisierbarkeit spontaner Kooperation zu testen: Es wird gezeigt, dass spontane Kooperation den Entscheidungsdefault für Individuen mit einer prosozialen Persönlichkeit repräsentiert. Weiterführend wird die Allgemeingültigkeit spontaner Prosozialität untersucht und geprüft, ob Bestrafungsverhalten in sozialen Dilemmata als so genannte instrumentelle Kooperation ebenfalls ein spontanes Phänomen darstellt und analog zu spontaner Kooperation demselben Verlauf über die Entscheidungszeit folgt. In diesem Kontext wurden auch die zugrundeliegenden Motive als moderierende Faktoren untersucht und gegen jene kontrastiert, die spontaner Kooperation unterliegen. Hierbei zeigt sich, dass spontanes Bestrafungsverhalten im Unterschied zu spontaner Kooperation kein Akt von dispositional-prosozialen Individuen ist. Vielmehr ist spontane Bestrafung retributiver Natur und wird von den Personen ausgeführt, die überdurchschnittlich viel zum öffentlichen Gut beigetragen haben. Zusammenfassend werden die Ergebnisse spontan- elementaren wie instrumentellen Kooperationsverhaltens kritisch im Licht einer spontanen Prosozialitätstheorie diskutiert.

Acknowledgment

Earning a doctorate degree is a long process that yields more than merely scientific articles. During this period, long-term personal and professional development has taken place. I would like to use this opportunity to thank the most important people who influenced and accompanied this process.

First and foremost, I dedicate this thesis to my family. In particular, I am grateful to my parents and brother for endowing me with my (Nordic) daily-life essentials of faith, trust, optimism, and dedication.

Furthermore, I would like to express my genuine gratitude to my supervisor, Professor Andreas Glöckner, for his guidance and support. I had never imagined such professional development to be possible. Dear Andreas, I owe this to you and am deeply grateful!

Moreover, I would like to thank Professor Stefan Schulz-Hardt for his readiness to co-supervise my thesis and the decisive advice to submit the punishment paper to OBHDP. Dear Professor Schulz-Hardt, at that moment, I did not know how strongly doing research on advice-taking correlates with the quality of advice-giving. Thank you very much!

Last but by far not least, I would like to thank Professor Peter Lewisch for a unique collaboration that is yet unrivaled. Dear Professor Lewisch, for your scholarly and personal support – the latter particularly in the last months – I can only express my deepest gratitude! The conference coffee break in Long Beach, where we outlined the first draft of the punishment project, remains unforgettable!

Next to the professional support during the doctorate, I would like to thank those whom I can rely on every day, both in and beyond the academic context: Antonia Schöning for the most long-lasting and best friendship. Annika Giersiepen and Stella Wanzel – it is one thing to return to one's hometown but quite another to make new friends there. Thank you for the best of times we had in "Good old Gö" (and Hamburg, Römö, Lisbon, Hagen, ...)! My colleagues in Göttingen, Hagen, and Cologne – in particular, Angela Dorrough and Marc Jekel – for being a tower of strength in all stages of the doctorate. Isabel Thielmann, Sophie Scharf, and Sabrina Strang for personifying the privilege in academia to work together with the nicest and most competent colleagues! Jana Reichhold and Janik Goltermann, as representatives of all (former) research assistants, for your endless support in all (and particularly the critical) circumstances!

Table of Contents

Articles

The relevant articles for this thesis are:

1. Mischkowski, D., & Glöckner, A. (2016). Spontaneous cooperation for prosocials, but not for proselfs: Social value orientation moderates spontaneous cooperation behavior. *Scientific Reports*, *6*, 21555. https://doi.org/10.1038/srep21555
2. Mischkowski, D., Glöckner, A., & Lewisch, P. (2018). From spontaneous cooperation to spontaneous punishment. Distinguishing the underlying motives driving spontaneous behavior in first and second order public good games. *Organizational Behavior and Human Decision Processes*, *149*, 59-72. https://doi.org/10.1016/j.obhdp.2018.07.001

Materials, analysis scripts, and data can be found at the Open Science Framework:

First article: https://osf.io/w7hsk/

Second article: https://osf.io/9rpwn/

Additional analyses of this synopsis: https://osf.io/aqbrc/

Additional articles have been published during the doctorate. They are cited when appropriate.

Bassarak, C., Leib, M., Mischkowski, D., Strang, S., Glöckner, A., & Shalvi, S. (2017). What provides justification for cheating – Producing or observing counterfactuals?. *Journal of Behavioral Decision Making*, *30*(4), 964-975. https://doi.org/10.1002/bdm.2013

Bouwmeester, S., Verkoeijen, P. P. J. L., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., . . ., Mischkowski, D., …Wollbrant, W. E. (2017). Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, 12(3), 527-542. https://doi.org/10.1177/1745691617693624

Jekel, M., Fiedler, S., Allstadt-Torras, R., Mischkowski, D., Dorrough, A., & Glöckner, A. (2019). How to teach open science principles in the undergraduate curriculum – the Hagen Cumulative Science Project. *Psychology Learning and Teaching*, 1475725719868149. https://doi.org/10.1177/1475725719868149

Mischkowski, D., Stone, R., & Stremitzer, A. (in press). Promises, expectations, and social cooperation. *Journal of Law and Economics*.

Mischkowski, D., Thielmann, I., & Glöckner, A. (2018). Think it through before making a choice? Processing mode does not influence social mindfulness. *Journal of Experimental Social Psychology, 74*, 85-97. https://doi.org/10.1016/j.jesp.2017.09.001

Mischkowski, D., Thielmann, I., & Glöckner, A. (2019). Minimizing inequality versus maximizing joint gains: On the relation between personality traits and different prosocial motivations. *Journal of Research in Personality, 81*, 153-157. https://doi.org/10.1016/j.jrp.2019.06.001

1. Introduction

The discussion of human nature, whether prosocial motives dominate over egocentricity, is ongoing in many disciplines such as philosophy, biology, psychology, and economics (e.g., Burnstein, Crandall, & Kitayama, 1994; Dawes & Thaler, 1988; Dawkins, 1976; Rand & Nowak, 2013; Warneken & Tomasello, 2006). Cooperation behavior in social dilemmas poses a form of prosocial behavior, as individual and (societal) group interests are conflicting in this context. This thesis sheds light on the underlying decision processes of cooperation behavior by considering the self-paced decision time for (un-)cooperative behavior. Specifically, it investigates for whom cooperation represents a spontaneously emerging phenomenon that is reflected in shorter decision times than defective (i.e., un-cooperative) behavior. Second, it examines to which extent spontaneous cooperation generalizes to costly punishment behavior in social dilemmas. This is accomplished by investigating whether punishment similarly poses a spontaneously expressed phenomenon that is dependent on dispositional pro-sociality or whether the underlying motives differ from those of spontaneous cooperation.

In recent years, interdisciplinary research has started to focus on the cognitive processes that underlie prosocial behavior (e.g., De Dreu et al., 2010; S. Fiedler, Glöckner, Nicklisch, & Dickert, 2013; Fischbacher, Hertwig, & Bruhin, 2013). Rand, Greene, and Nowak (2012) propose a fundamentally positive evaluation of human nature – in that humans behave intuitively prosocial – by showing a “spontaneous cooperation effect”. This effect is characterized by a negative correlation between decision times in single-play (“one-shot”) social dilemmas and cooperation behavior. Furthermore, putting participants under time pressure yields significantly higher cooperation rates in comparison to a time delay manipulation. This not only indicates the causality of reduced decision time leading to higher cooperation behavior but also implies potential intervention opportunities, in that pro-sociality can be promoted even more strongly by inducing an intuitive processing mode.

These findings fundamentally contradict economic rational choice theory (e.g., Becker, 1976; Kreps, Milgrom, Roberts, & Wilson, 1982), which predicts a striving for individual gain maximization in general irrespective of processing mode. Obviously, Rand et al. (2012) were not the first to foil the image of the *homo oeconomicus*. Models of social preferences have acknowledged the utility that some individuals gain when considering other persons' welfare (Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Fehr & Schmidt, 1999; Van Lange, 1999). In addition, Rand et al. (2014) propose the "Social Heuristics Hypothesis" (SHH), which seeks to explain from an evolutionary perspective how cooperation has proven to be advantageous in daily-life and is adapted as the behavioral default even in one-shot laboratory interactions.

A scientific debate about the spontaneous cooperation effect started shortly after the principal publication and is still ongoing (e.g., Kvarven et al., 2019; Montealegre & Jimenez-Leal, 2019). In addition to the heterogeneous results of several replication studies (e.g., Bouwmeester et al., 2017; Camerer et al., 2018; Tinghög et al., 2013), which opened the discussion of its replicability, moderators were identified to better understand the conditional – situational as well as dispositional – factors that underlie spontaneous cooperation (e.g., Kieslich & Hilbig, 2014; Rand et al., 2012; Santa, Exadaktylos, & Soto-Faraco, 2018). The two articles in this thesis contribute to this debate by answering the following questions:

a) Is spontaneous cooperation replicable as a main effect?

b) Does a person-situation interaction explain the heterogeneity of replication results in that spontaneous cooperation is conditional on dispositional pro-sociality?

c) Does spontaneous cooperation generalize to costly punishment behavior in social dilemmas? If so, how comparable are the underlying motives?

In addition to linking and discussing the two articles of this thesis, this synopsis provides two overarching contributions to the literature: First, it focuses on the underlying theory of spontaneous cooperation, namely the SHH (Rand et al., 2014), and evaluates its empirical content (Popper, 1934/2005). Specifically, it is outlined how the empirical content increases when using the results of the articles of this thesis to modify the SHH. Second, the decision conflict hypothesis (A. M. Evans, Dillon, & Rand, 2015; Krajbich, Bartling, Hare, & Fehr, 2015) is investigated in this synopsis in greater detail – which stems from recent debates in which the authors argue that measured (vs. manipulated) decision time cannot be used to infer an intuitive vs. deliberate processing mode. Rather, decision time was shown to represent the degree of decision conflict (A. M. Evans et al., 2015) or the (lacking) strength of preferences towards one of the choice options (Krajbich et al., 2015). Therefore, this synopsis provides additional analyses whether the data of the two articles support the entanglement of decision conflict and intuitive processing in addition to discussing their relation on theoretical grounds.

Outlining the structure of this thesis, I will first sketch the theoretical background behind the spontaneous cooperation effect. To do so, I will first introduce cooperation and punishment behavior in social dilemmas before linking them to social value orientation as a measure of dispositional pro-sociality. Then, I will direct the focus towards the cognitive processes that potentially underlie (non-)cooperative behavior and describe how they have shown to be reflected in decision time duration. Specifically, I will consider arguments from the debate mentioned above and distinguish between measured and manipulated decision time. Delineating the research question of the first paper, the spontaneous cooperation effect and its underlying theory, the SHH, are presented, followed by an overview of the replication attempts and identification of moderators of the spontaneous cooperation effect.

Discussing the first paper subsequently concerning its limitations and impact requires one to analyze whether decision conflict can similarly account for spontaneously expressed behavior (of prosocials) as could an intuitive processing mode. The results of additionally conducted analyses are subsequently discussed concerning their implications for the current work. Transitioning to the second paper, the question will be addressed regarding how generalizable the spontaneous cooperation effect is, as the SHH as the underlying theory of the spontaneous cooperation effect is mute concerning its applicability beyond cooperation behavior. In this context, I will analyze the empirical content of the SHH to assess its general scientific quality, as this allows one to outline the contribution of both articles in this thesis on theoretical grounds. This synopsis concludes by critically discussing the results of the second paper before turning to the implications this thesis has for a modified theory of spontaneous pro-sociality.

1.1 Cooperation as the rational choice

"Let us try to teach generosity and altruism, because we are born selfish."

(Dawkins, 1976, p. 215)

Cooperation behavior in social dilemmas is a necessary prerequisite for a sustainable society (Hardin, 1968). Environmental protection to minimize climate change represents such a social dilemma, exemplifying one of the currently most complex societal challenges (Cramton, MacKay, Ockenfels, & Stoft, 2017; Milinski, Sommerfeld, Krambeck, Reed, & Marotzke, 2008). Here, the dilemma structure and its societal implications are particularly obvious: It is individually costly to engage in environmental protection (e.g., by recycling waste or using public transportation rather than going by car). Therefore, one might rely on the eco-sensitive behavior of other individuals to achieve the goal of reduced carbon dioxide emission. However, when too many people free-ride and refrain from contributing to environmental protection, global warming is inevitable and (negatively) affects free-riders and contributors to the same extent. In more general terms, Dawes (1980) characterized social dilemmas by two features: First, there is a temptation to not cooperate, as defection individually yields a higher outcome than cooperation, independent of other individuals' behavior. Second, and creating the dilemma structure, if all societal members defect (i.e., no one contributes to the public good), the individual's outcome is lower than if all individuals cooperated (for reviews on social dilemmas, see Kollock, 1998; Van Lange, Joireman, Parks, & van Dijk, 2013).[1]

Economic games allow researchers to model the complex structure of real-life social dilemmas in a simplified way and investigate the determinants of cooperation behavior in laboratory settings while assuring internal validity (Camerer, 2003). The prisoner's dilemma (Rapoport & Chammah, 1965; Tucker, 1983) and its extension to multiple players as the public goods game (Head, 1974; Ledyard, 1995; Marwell & Ames, 1979) are two of the most common social dilemma games. In a public goods game, the relevant social dilemma for this thesis, participants face the conflict between either keeping a monetary endowment (i.e., choosing to defect) or contributing (parts of) it to a common pool (i.e., choosing to cooperate) at the risk of being exploited should the others defect. The degree of conflict is mirrored in the marginal per capita return (MPCR), reflecting the relative earning that the individual receives from the public good for every contributed monetary unit (see Isaac, Walker, & Thomas, 1984). Higher MPCR factors reflect increasing returns from the public good. As soon as the return of the public good is larger than the individual contribution (i.e., an MPCR greater than one), the dilemma situation is dissolved, as cooperation becomes individually beneficial.[2]

Turning to the question of whether people should cooperate from a game theoretical perspective – setting individual payoff maximization as the ultimate goal – defection is the dominant strategy for finite interactions, independent of the other actors' behavior (Selten, 1978). This holds not only for one-shot but also for a finite number of multiple interactions

[1] In this context, it is useful to distinguish between social dilemmas, namely social traps (also known as take some dilemmas or the tragedy of the commons, see Hardin, 1968) and social fences, also known as give some dilemmas (Van Lange et al., 2013). The collective risk of climate change consists of a social trap, where it is individually beneficial to exploit currently existing resources to a maximum degree at the risk of creating unforeseeable damage for future generations (Milinski et al., 2008). In turn, in social fences, or give some dilemmas, the resource needs to be built in the first place (e.g., dikes that protect a population from being flooded) before it can provide its benefits. This thesis uses paradigms of the latter dilemma type, where individuals can choose whether to contribute to a resource and potentially benefit in case of its multiplication.

[2] In turn, a social dilemma is present when the individual only gets a fragment of its contribution in return from the public good (i.e., an MPCR smaller than one). Therefore, a cooperative individual is at risk of being exploited as she is dependent on the contribution behavior of the societal group members in order not to lose originally possessed resources.

(Pettit & Sugden, 1989). In this vein, predictions of economic rational choice theory paint a very bleak picture of human cooperativeness in that decisions are only made in line with an individualistic, self-maximizing tendency (Becker, 1976; Kreps et al., 1982) where cooperation occurs only when it benefits the actor (Hechter, 1987) or when there is a risk of detection and punishment (Becker, 1968). The prototype of a rational actor is typically described as the *homo oeconomicus*. Characterizations of the homo oeconomicus are extensive and sketch his rational and egocentric nature as a "player [who] is not indifferent to any possible profit, however small" (Von Neumann & Morgenstern, 1944, p. 228). This perspective on human nature leaves little space for pro-social behavior – be it helping behavior (e.g., donations) or cooperation behavior in social dilemmas.

Fortunately, the majority of people do not behave as rational choice theory suggests. Research investigating when and why human decision making deviates from the rational choice prediction postulated by classical economic theory is summarized under the framework of behavioral economics (for an overview of the historical development, see Camerer & Loewenstein, 2011). Individuals consider the outcome and welfare of others and refrain from maximizing their own payoffs. For instance, people share almost a third of their resources, on average, in dictator games (for a meta-analysis, see Engel, 2011) and they even cooperate in one-shot public good games without any prospect of reciprocity (e.g., Andreoni, 1988; Yamagishi, Terai, Kiyonari, Mifune, & Kanazawa, 2007).

In addition, individuals not only cooperate but also punish norm-violators at their own cost. Punishment enables individuals to restore equality in outcomes by withdrawing resources at the cost of investing (a portion of) their own resources. Important for the generalization from spontaneous cooperation to spontaneous punishment addressed in this thesis, such punishment similarly represents a form of cooperation behavior as it shares the same characteristic of being costly for the individual yet beneficial for the group. Punishment behavior is therefore referred to as a second-order public good (Fehr & Gächter, 2002; Yamagishi, 1986) as "[e]verybody in the group will be better off if free riding is deterred, but nobody has an incentive to punish the free riders" (p. 137). In a similar vein, Yamagishi (1986, 1988) distinguishes between elementary cooperation where individuals contribute to provide the public good and instrumental cooperation (i.e., punishment behavior). The latter term points out that punishment is an effective tool to make defection unattractive in the first place, as the cost of being punished easily exceeds the benefits of defection (Yamagishi, 1986). Corroborating the functionality of punishment, it was robustly shown that punishment is suited for and used to maintain cooperation in repeated interactions (Boyd & Richerson, 1992; Camerer & Fehr, 2006; Fehr & Fischbacher, 2003; Fehr & Gächter, 2000; Fehr & Rockenbach, 2004; Oliver, 1980). Without the threat of being punished, cooperation diminishes over repeated interactions (Fehr & Gächter, 2002).

Similar to cooperation behavior, however, the rational choice is not to invest any resources to punish others. Stated differently, the homo oeconomicus would never punish, much as he would not cooperate in the first place. However, the behavior that individuals actually display once again contradicts the rational choice prediction, similar to the case of cooperation behavior. Punishment has even been observed in one-shot interactions (e.g., Henrich et al., 2006), where punishment is purely retributive and cannot serve (direct) deterrence purposes.

Taken together, there are two stages in which individuals can cooperate, that is contribute to a public good. The first order public good (i.e., elementary cooperation) consists of the contribution behavior to maintain the public good (or, in the case of a take-some dilemma, refrain from exploiting the resource). As a second order public good, norm violators can be punished to restore equality and deter from future free-riding (i.e., instrumental cooperation; Yamagishi, 1986). Contrary to rational choice theory, people engage in both behaviors. One explanation for the discrepancies between normative predictions and the descriptive level of cooperation behavior can be located in the spectrum of social preferences (e.g., social value orientation; Van Lange, 1999) as presented in the following.

Social value orientation

Models of social preferences capture stable individual differences in inequality aversion (Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Fehr & Schmidt, 1999) or, more generally, the consideration of the outcome and welfare of others (Van Lange, 1999). Social value orientation (SVO) is one prominent concept within the framework of social preferences taking into account the notion "that individuals tend to pursue broader goals than self-interest" (Van Lange, 1999, p. 337). Importantly, SVO still allows one to capture the weighting function of a perfectly self-interested (i.e., individualistic) actor and thus does not contradict but rather expands the rational choice prediction (Murphy, Ackermann, & Handgraaf, 2011). Specifically, SVO mirrors the weight one allocates to outcomes of oneself vs. others (Balliet, Parks, & Joireman, 2009) when distributing (monetary) resources in a series of decomposed games in which an individual allocates a fixed sum of money between herself and a recipient (Murphy & Ackermann, 2014). As an example, consider the item of the SVO Slider Measure (Murphy et al., 2011) depicted in Figure 1: Allocating resources between oneself (upper row) and an anonymous other person (lower row) implies weighting one's own against the other person's outcome in absolute terms as well as in terms of the difference between outcomes (i.e., degree of inequality). Formally, the weighing of one's own versus another person's outcome can be expressed in a utility function where $U = w1 \times \text{(own payoff)} + w2 \times \text{(others' payoff)}$ (Liebrand & McClintock, 1988). Different specifications of this function exist – for instance, by adding a $w3$ component that represents the motivation to achieve equality in outcomes (Van Lange, 1999).[3]

[3] The herein used operationalization of SVO with the SVO Slider Measure (Murphy & Ackermann, 2014; Murphy et al., 2011) calculates within the primary six items solely the weight to the own ($w1$) and the other person's outcome ($w2$). Inequality aversion ($w3$) is seen as a prosocial preference (Murphy et al., 2011; Van Lange, 1999) and contrasted against the maximization of joint gains as a different prosocial motivation in nine secondary items of the Slider Measure (Ackermann & Murphy, 2012).

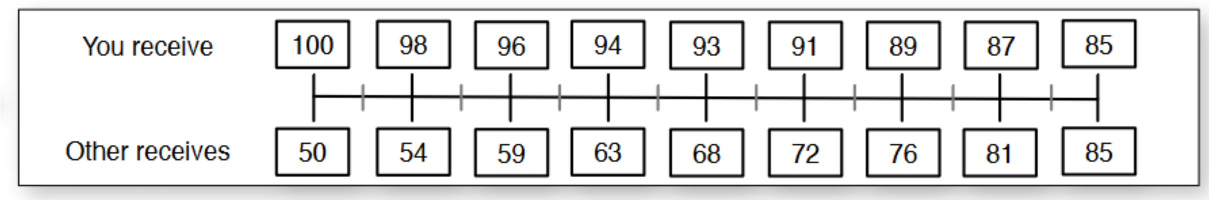

Figure 1. Exemplary item of the SVO Slider Measure (Murphy et al., 2011). Participants allocate monetary resources between themselves (upper row) and an anonymous other person (lower row).

The allocation choices allow for the classification of an individual as one of four SVO types (Liebrand & McClintock, 1988). Figure 2 shows the different categories and corresponding monetary allocations (Murphy et al., 2011): Individualists (i.e., proselfs) only give weight to their own outcome ($w1 = 1$, $w2 = 0$), which is reflected in a corresponding choice of the first option (100 monetary units (MU) for oneself vs. 50 MU for the other person, see also Figure 1 as the corresponding item). Altruists, in contrast, only focus on maximizing the other's outcome at the cost of minimizing their own ($w1 = 0$, $w2 = 1$). In between those two extremes are the prosocials, who give equal weight to both outcomes ($w1 = 1$ and $w2 = 1$) – for instance, by choosing a fair split of 85 MU for both decision makers (see Figure 2). Notably, also competitors consider the other person's outcome, though with a diametrically different aim to maximize inequality to one's benefit ($w1 = 1$, $w2 = -1$). This implies that competitors refrain from maximizing their own payoff for the sake of putting the other person relatively worse off, as for instance reflected in a choice of 85 MU vs. 15 MU instead of equally distributing 85 MU (see Figure 2).

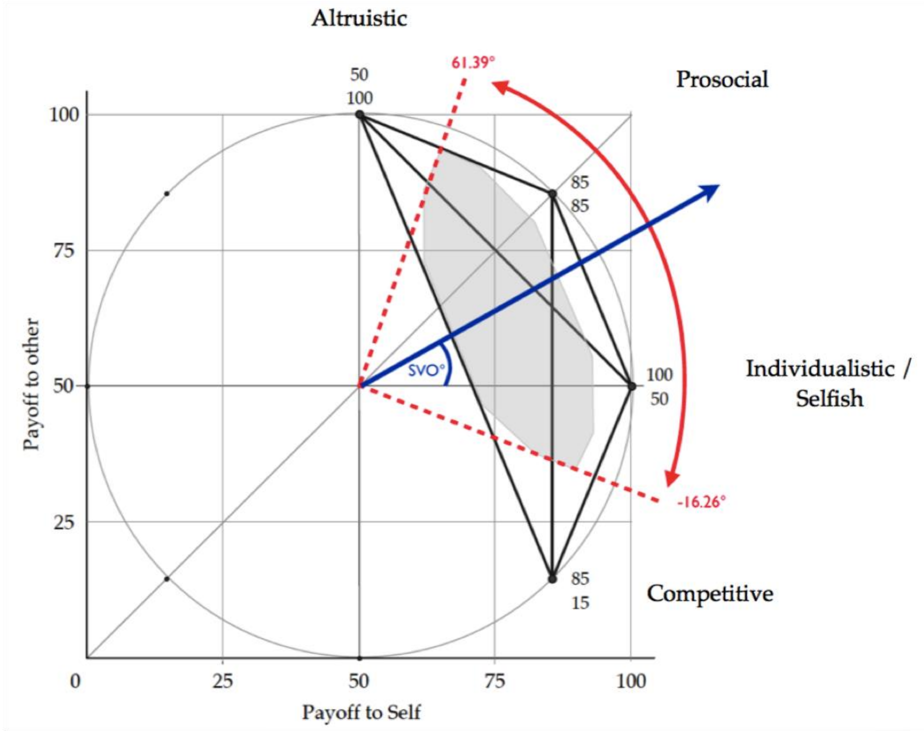

Figure 2. Graphical representation of social value orientation (SVO). Categorizing a person to one of the four categories of social values stems from her weighing her own payoff (*w*1, displayed on the x-axis) against another person's outcome (*w*2, displayed on the y-axis). The solid lines reflect the six primary items of the SVO Slider Measure contrasting the different types against each other. Figure 1 represents the line from an individualistic value orientation (100 MU for oneself vs. 50 MU for the other) to a prosocial value orientation (each gets 85 MU).

Reprinted from "Measuring Social Value Orientation" by K. A. Ackermann, 2011, The 14th international conference on social dilemmas, July 9, Netherlands: Amsterdam. Reprinted with permission.

Initially, SVO was only captured on the category level (e.g., by the Triple Dominance Measure; see Van Lange, Otten, De Bruin, & Joireman, 1997), posing the additional disadvantage of resulting in some unclassifiable individuals when allocation choices were too inconsistent to allow for an unambiguous classification (Murphy & Ackermann, 2014). Historically most influential was the Ring Measure (Liebrand, Jansen, Rijken, & Suhre, 1986; Liebrand & McClintock, 1988), which was the first to conceptualize SVO via the Cartesian SVO framework (as displayed in Figure 2). It not only classifies an individual to one of the four categories, but allows for the calculation of a continuous SVO angle that mirrors the degree of pro-sociality on a more fine-grained (continuous) level (see the exemplary angle delineated in

Figure 2). This gradual differentiation in weights beyond the integer values of -1, 0, and 1 are both conceptually meaningful and empirically supported (S. Fiedler et al., 2013). For instance, individuals often consider the other's outcome but to a lesser extent than their own (e.g., Fischbacher, Gächter, & Fehr, 2001). Subsequent SVO measures such as the SVO Slider Measure (Murphy & Ackermann, 2014; Murphy et al., 2011) build upon the Cartesian SVO framework and provide a more economic measure of SVO by omitting some of the Ring Measure's items that capture empirically less frequently observed phenomena (e.g., individuals choosing negative outcomes for themselves as, for instance, a masochistic individual would do, $w1 = -1$, $w2 = 0$; Murphy & Ackermann, 2014). In addition, it provides very good psychometric properties (e.g., a high test-retest reliability; see Murphy et al., 2011), which is why it is used as the SVO measure in this thesis.

In terms of the relation between SVO, cooperation, and punishment behavior in social dilemmas, two aspects are noteworthy. First, cooperation behavior and SVO are distinct constructs, as SVO does not entail a strategic component affected by expectations regarding other's behavior or a fear of being punished (Murphy & Ackermann, 2014). This independence allows one to unequivocally infer the underlying social preference (e.g., altruism vs. individualism). In contrast, cooperation behavior confounds social preferences and mentioned beliefs about others' behavior (Kelley & Stahelski, 1970). Second – and important when considering in the following SVO as the individual difference when investigating person-situation interactions – SVO consists of a stable individual difference (Camerer & Fehr, 2004; Murphy et al., 2011) that reliably predicts cooperation behavior with a small to medium effect size of approximately $r = .30$ (see Balliet et al., 2009, for a meta-analysis). When correcting for publication bias, the correlation shrinks to $r = .25$ (Renkewitz, Fuchs, & Fiedler, 2011). The relation between SVO and punishment behavior, however, is less clear and empirical evidence is heterogeneous. On theoretical grounds, one would expect less punishment of proselfs who should prioritize the maximization of personal gain over costly restoration of outcome equality.

Additionally, prosocials, in particular, should punish to (re-)establish prosocial norms that protect them from being exploited. In line with this expectation, increased punishment of prosocials was shown by Bieleke, Gollwitzer, Oettingen, and Fischbacher (2016) as well as Haruno, Kimura, and Frith (2014). Rendering the empirical situation less clear, however, several null effects between punishment and SVO were found (Böckler, Tusche, & Singer, 2016; Yamagishi et al., 2012) as well as even the reversed effect of reduced punishment for prosocials (Karagonlar & Kuhlman, 2013).

Taken together, the original idea of only self-interested individuals has been extended within the framework of social preferences; and the idea of human rationality has adapted accordingly (Tetlock & Mellers, 2002). Several models of social preferences (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999; Van Lange, 1999) can account for why cooperation is the rational choice for some individuals – they gain utility not only from considering their own welfare but also from considering the outcome of others, albeit with different aims (e.g., competitors vs. prosocials) and to different degrees (e.g., altruists vs. prosocials). Thus, it is well established that individual differences explain variability in cooperation behavior. However, these cannot explain *how* a decision is made – that is, which cognitive processes are involved that might lead to differences in cooperation behavior. This investigation is outlined in the following, focusing on decision time as an indicator of cognitive processes and its relation with cooperation behavior.

1.2 Cooperation as the default

Research on economic decision making and pro-sociality, in particular, only recently began to investigate the underlying cognitive processes (e.g., De Dreu et al., 2010; S. Fiedler et al., 2013; Fischbacher et al., 2013; Lotito, Migheli, & Ortona, 2013; Rand et al., 2012; Rubinstein, 2007). In addition to physiological data such as eye-tracking (S. Fiedler et al., 2013) and neuroimaging (Smith, Bernheim, Camerer, & Rangel, 2014), decision time represents an efficient and non-obtrusive measure to draw inferences on the underlying cognitive processes (for a review, see Spiliopoulos & Ortmann, 2018). By definition, it refers to the time needed or available to make a decision (Van de Calseyde, Keren, & Zeelenberg, 2014), typically measured from the onset when participants are presented with the screen where they can input their decision until they actually submit it (e.g., via mouse click).

Conceptually, decision time has commonly been understood to allow the differentiation between deliberate and automatic-intuitive processes (Rubinstein, 2007). This differentiation is made in dual process models (for overviews, see J. S. B. Evans, 2007; J. S. B. Evans, 2008; Weber & Johnson, 2009) that "come in many flavors, but all distinguish cognitive operations that are quick and associative from others that are slow and rule-governed" (Kahneman & Frederick, 2002, p. 51). However, there is an ongoing discussion concerning different concepts that decision time may operationalize (A. M. Evans et al., 2015; A. M. Evans & Rand, 2018). Specifically, decision time was shown to increase with decision conflict, showing that long decision times cannot unequivocally be attributed to increased deliberation (A. M. Evans et al., 2015; Krajbich et al., 2015). As this debate is highly relevant for the discussion of both articles (see Chapters 3.2.1 and 4.2.2), both perspectives are introduced in the following.

Decision time in light of dual-process models

Applying a dual process perspective on cooperation behavior in economic games allows one to address the question of which processes underlie cooperative decisions; whether there is "intuitive pro-sociality" (Zaki & Mitchell, 2013) or deliberation is required to behave pro-socially. There are numerous dual process models that differ slightly in their specifications but all share the conceptual duality of the human mind (for an historical overview, see Frankish & Evans, 1983). In contrast to cognitively effortful, rule-based and slow deliberate processes, intuitive processes are assumed to be affect-based, associative, quick, and automatic (e.g., Epstein, 1994; J. S. B. Evans, 1984; Kahneman, 2003; Shiffrin & Schneider, 1977; Sloman, 1996; Strack & Deutsch, 2004). The neutral description for both models consists of the system

1 vs. system 2 classification. System 1 refers to automatic, intuitive processing, whereas system 2 represents the deliberate and rule-based account (Kahneman, 2003). Despite the variety of dual process models, they can be classified into one of two approaches according to how they propose intuition and deliberation to interact (J. S. B. Evans, 2008). The first category takes a default-interventionist perspective where "(…) rapid autonomous processes (…) [system 1] yield default responses unless intervened on by distinctive higher order reasoning processes [system 2]" (J. S. B. Evans & Stanovich, 2013, p. 223). That is, deliberation allows reviewing and, when indicated, overriding initial tendencies, as for instance assumed by Kahneman and Frederick (2002). In contrast, parallel-competitive models (e.g., Sloman, 1996) state that intuitive and deliberate processes are initiated simultaneously; either going hand in hand with one another when leading to the same decision or eliciting decision conflict when provoking contrary responses (J. S. B. Evans, 2007). There is no definitive answer yet as to how the intuitive and deliberate processes interact. However, the present relevant domain of spontaneous pro-sociality has recently been associated with the default-interventionist perspective (Mrkva, 2017).

Based on the distinction between fast and automatic vs. slow and effortful processes, decision time received legitimacy as an operationalization of processing mode. As Alós-Ferrer and Strack (2014) noted "the first, obvious way to distinguish decision processes relies on the measurement of response times" (p. 2).[4] That is, deliberation is assumed to increase with self-paced decision time, whereas quick decisions are associated with an intuitive processing mode (Lotito et al., 2013; Nielsen, Tyran, & Wengström, 2014; Piovesan & Wengström, 2009; Rubinstein, 2007). Moreover, decision time can be experimentally manipulated: To elicit intuitive processing, time pressure is induced and contrasted against time delay or unconstrained decision time to allow for deliberation (Horstmann, Hausmann, & Ryf, 2009).[5]

[4] Note that decision time and response time (RT) are synonymously used in this thesis.
[5] Notably, there are several other options to induce an intuitive vs. deliberate processing mode (e.g., cognitive load). For an overview, see Horstmann et al. (2009).

One criticism that recently gained importance concerns the reverse inference from the speed of decision time to the processing mode when the latter is operationalized as measured rather than manipulated decision time: According to the dual process perspective, intuitive decisions are fast whereas deliberation requires time. However, reversely inferring that quick decisions are necessarily intuitive depicts the classic fallacy of the affirmation of the consequent (see De Neys, Schaeken, & d'Ydewalle, 2005). Stated differently, other factors could account for short decision times above and beyond an intuitive processing mode. Supporting the fallacy on empirical grounds, decision conflict and a lacking strength of preference for a behavioral choice were recently shown to be positively related to decision time (A. M. Evans et al., 2015; Krajbich et al., 2015; Yamagishi et al., 2017).

Decision time as an indicator of decision conflict

The dual-process approach of spontaneous behavior has recently been challenged by showing that decision time reflects the degree of decision conflict (Diederich, 2003) even in the field of economics (for a review, see Clithero, 2018) and, more specifically, pro-sociality (A. M. Evans et al., 2015; Krajbich et al., 2015). Specifically, Krajbich et al. (2015) showed that the (lacking) strength of preference towards a choice option influences decision time of cooperation behavior in a public goods game. The authors varied the return (i.e., the MPCR factor, see Chapter 1.1) from contributing to the public good in order to manipulate the strength of preference towards cooperation. In doing so, they found that decision time is shorter for cooperative choices when a high return favors cooperation. Conversely, given a low return, defection is the faster choice. Complementing the findings of Krajbich et al. (2015), A. M. Evans et al. (2015) showed that self-indicated feelings of decision conflict also prolong decision time when deciding whether to cooperate in a public goods game (A. M. Evans et al., 2015; A. M. Evans & Rand, 2018). These findings are subsumed as the *decision conflict hypothesis*: If low decision conflict rather than intuitive processing is reflected in short decision times, then decision *extremes* (i.e., fully cooperative as well as fully defective behavior such as giving all versus nothing in a public goods game) but not only fully cooperative decisions should be fast. The rationale behind this notion is that cooperation and defection alike can be accompanied by low decision conflict – in that there is a clear strength of preference for either one of the choice options – and can thus both be expressed spontaneously. Solving decision conflict requires time and leads to intermediate levels of cooperativeness (A. M. Evans et al., 2015). The assumed relation when predicting decision time by cooperation behavior takes on an inverted u-shaped pattern with short decision times at both ends of the cooperative spectrum and increased decision time in the case of intermediate cooperativeness. In turn, manipulating an intuitive processing mode has been shown to increase only cooperation behavior, but neither decision extremes nor self-indicated feelings of conflictedness (A. M. Evans et al., 2015).

On theoretical grounds, the decision conflict hypothesis and the resulting inverted u-shaped pattern is reflected in evidence accumulation models (also known as sequential sampling, information accumulation or drift diffusion models; e.g., Ratcliff & Smith, 2004). They posit one process of judgment and decision making rather than differentiating between two systems and thus stand in contrast to the dual process approach (Coricelli, Polonio, & Vostroknutov, forthcoming; Spiliopoulos & Ortmann, 2018). Krajbich et al. (2015) emphasized that "it is critically important to consider the possibility that there may just be a single deliberative process governing choices, and that variations in RT [decision time] are due to the perceived similarity of the choice options and not competing processes" (p. 2). In other words, in contrast to the dual process account, evidence accumulation models assume a single decision process. Predicted decision time varies depending on the amount of information that must be processed in order to exceed a certain decision threshold (Klauer, 2014). The higher the decision conflict, the more information must be acquired – as a result, the decision process takes longer (Krajbich & Rangel, 2011). Taken together, when measuring rather than manipulating decision time – as done in the articles in this thesis – one must take into account the fact that decision time may not only capture an intuitive processing mode. This requires the use of the term *spontaneous* rather than intuitive behavior when referring to short decision times.[6]

However, it would fall short to assign the incorporation of decision conflict only to single-process models. Dual-process models, as well, include decision conflict and its resolution: In their reply to Krajbich et al. (2015), Pennycook, Fugelsang, Koehler, and Thompson (2016) picked up the argument from parallel-competitive models (e.g., Sloman, 1996, 2014) that low decision conflict is a prerequisite for intuitive decision-making. In the case of decision conflict, it is immediately detected (De Neys, 2012, 2014; De Neys & Glumicic, 2008) and causes deliberation (Pennycook, Fugelsang, & Koehler, 2015; Pennycook et al., 2016). Therefore, short decision times as the reflection of the individual default might contain both low decision conflict and intuitive processing. The strength of preference for one option (i.e., the degree of decision conflict) would then determine how quick and intuitive a decision can be made or whether deliberation is required to solve decision conflict.

As a current development, there is the striving to combine the dual process approach with evidence accumulation models (Achtziger & Alós-Ferrer, 2013; Alós-Ferrer, 2018; Chen & Krajbich, 2018), even though the application of the latter on strategic choices is still in its infancy (Coricelli et al., forthcoming). Given the ongoing debate regarding the different interpretations of decision time (see Konovalov & Krajbich, 2019), it is important to clarify that the first article of this thesis (Mischkowski & Glöckner, 2016) consists of a direct replication of the first study of Rand et al. (2012) that measures decision time and is based on the dual process perspective. As the debate regarding the entanglement of intuitive processing and decision conflict gained importance since the publications of A. M. Evans et al. (2015) and Krajbich et al. (2015), the second paper (Mischkowski, Glöckner, & Lewisch, 2018) integrates the decision conflict perspective when reviewing the theoretical background and discussing the

[6] The term “spontaneous” behavior stems from Rand et al. (2012).

results. However, as mentioned above, the influence of decision conflict has not yet been empirically tested for this thesis' articles. Therefore, in Chapters 3.2.1 and 4.2.2 the corresponding analyses are conducted to contribute to the debate concerning whether short decision times are also attributable to low decision conflict.

After discussing the general theoretical framework by introducing the central variables of this thesis – cooperation and punishment behavior, SVO, and decision time – it is time to shed light on their relation. Specifically, the theoretical and methodological details of the spontaneous cooperation effect are introduced in the following before summarizing replication attempts and identifying moderators to outline the research question of the first article.

2. The spontaneous cooperation effect – Replication attempts and identified moderators

In the following, I will introduce the spontaneous cooperation effect and its underlying theory, the Social Heuristics Hypothesis (SHH). I will then summarize replication attempts and identified moderators to outline the scientific debate on the spontaneous cooperation effect that has evolved since the original publication in 2012.

2.1 The Spontaneous Cooperation Effect

In their original publication, Rand et al. (2012) made two major contributions: First, they investigated the underlying processes of cooperation behavior by testing the influence of an intuitive vs. deliberate processing mode. Second, they contributed fundamentally contradicting evidence against the classic economic perspective by showing that cooperation is the first and spontaneously executed response in one-shot interactions.

Overall, the authors conducted ten studies in which they correlatively measured decision time as well as manipulated an intuitive vs. deliberate processing mode in several ways (e.g., via time pressure vs. time delay; priming an intuitive vs. reflective mindset).[7] Their results showed a continuous decrease of cooperation behavior in one-shot social dilemmas (e.g., prisoner's dilemmas and public good games) the longer participants took to make a decision (see Figure 3).

[7] To prime intuitive or deliberate processing, Rand et al. (2012) asked participants to remember either good or bad experiences with decisions made in one of the processing modes. Intuition is primed when participants are asked to remember a positive outcome of a decision that was made intuitively and a negative experience after having decided in a deliberate manner (and vice versa for the prime of deliberation).

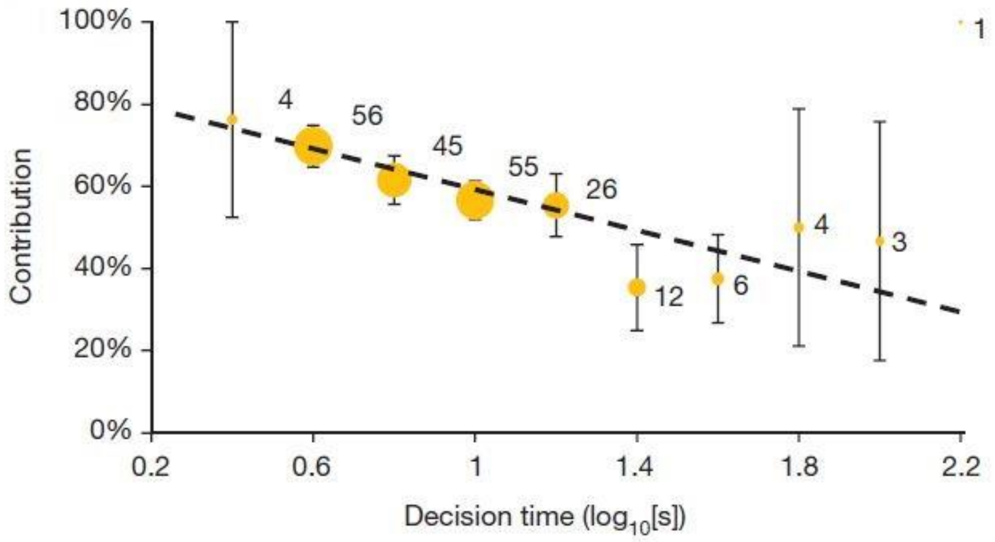

Figure 3. The spontaneous cooperation effect. With increasing, self-paced (i.e., measured) decision time, cooperation behavior operationalized as the percentual contributions in a one-shot public goods game decreases.

Reprinted from "Spontaneous giving and calculated greed" by D. G. Rand, J. D. Greene and M. A. Nowak, 2012, *Nature*, *489*, p. 427.

Additionally, when manipulating processing mode, the results showed statistically significant higher cooperation rates in comparison to a time delay manipulation (see Figure 4). The authors concluded that an intuitive compared to a deliberate processing mode increases cooperation behavior, suggesting an intervention opportunity that even allows one to promote cooperation. This implication is critical, given the outlined importance of social dilemmas in real life (e.g., prevention of climate change; see Chapter 1.1).[8]

[8] Note that a control condition containing unconstrained decision time was taken from the first, correlative study of Rand et al. (2012). It indicated a non-significant difference in contributions to the time pressure condition at the conventional alpha-level of 5% (p = .058; see Figure 4). As Everett, Ingbretsen, Cushman, and Cikara (2017) point out, a control condition is needed to infer whether an intuitive processing mode increases or deliberation decreases cooperative behavior. Their results point even more clearly to the latter, suggesting deliberate defection rather than spontaneous cooperation, thus contradicting the claim of Rand et al. (2012).

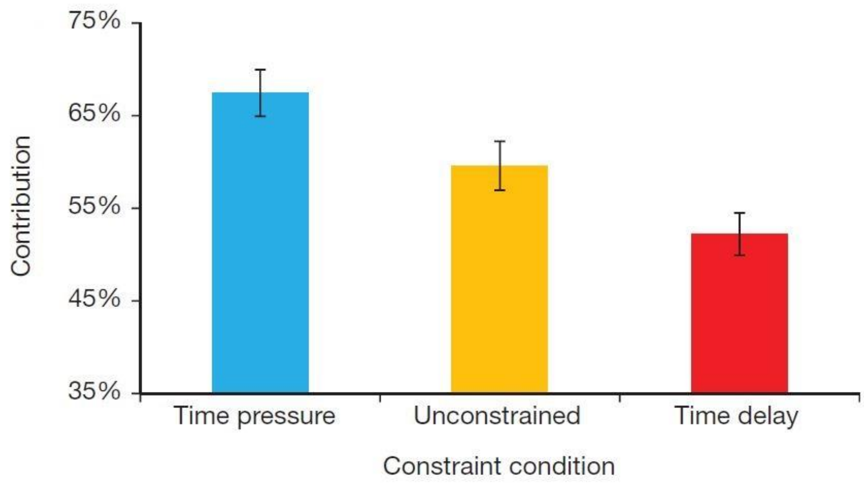

Figure 4. Spontaneous cooperation effect when manipulating decision time. Cooperation behavior (operationalized as percentual contributions) in a public goods game is significantly higher when participants are under time pressure (i.e., forced to decide in less than 10 seconds) in contrast to a time delay condition in which participants were forced to wait at least 10 seconds before entering their contribution decision. The unconstrained condition reflects self-paced (i.e., measured) decision time for contributions.

Reprinted from "Spontaneous giving and calculated greed" by D. G. Rand, J. D. Greene and M. A. Nowak, 2012, *Nature*, *489*, p. 428. Copyright 2019 by Springer Nature. Reprinted with permission.

The Social Heuristics Hypothesis

Based on these results, Rand and colleagues postulated the Social Heuristics Hypothesis (SHH, Bear & Rand, 2016; Rand et al., 2014) to provide a post-hoc explanation of the spontaneous cooperation effect. The SHH states from an evolutionary perspective that cooperation has been internalized as the beneficial strategy in daily-life interactions where one-shot settings are comparably rare. Rather, daily-life interactions mostly consist of repeated, non-anonymous interactions that include a potential threat of being sanctioned and thus elicit spontaneous cooperation. Hence, deliberation is required to adapt to the artificial lab situation that favors defection. To encounter failed replications that were published in the meantime (e.g., Tinghög et al., 2013, see subsequent Chapter 2.2), Rand et al. (2014) acknowledge that spontaneous cooperation consists of an "averaged phenomenon" and postulate that for some individuals defection might be anchored as the first and intuitive response. This could be rooted in prior experience with the laboratory setting of social dilemma experiments (Rand et al., 2012)

– that is, individuals who spontaneously defect internalized defection in contexts that do not support cooperation (Rand et al., 2014). Still, the authors provide a falsifiable theory in that they state that "intuition should never decrease average cooperation relative to reflection in one-shot anonymous social dilemmas" (Rand et al., 2014, p. 2).

From a methodological perspective, the original studies showing spontaneous cooperation suggested an easily replicable effect. The authors built their findings on sufficiently powered lab and online studies with both student and non-student samples (i.e., MTurk workers, see Buhrmester, Kwang, & Gosling, 2011). However, roughly seven years after the original publication, the spontaneous cooperation effect is still highly debated and new evidence – for and against the effect – is constantly added (e.g., Bouwmeester et al., 2017; Everett et al., 2017; Isler, Maule, & Starmer, 2018; Kvarven et al., 2019; Rand, 2016, 2017a, 2017b). In the following, replication attempts are evaluated in more detail before identified moderators are summarized to shed light on the potential sources that may underlie the heterogeneous replication results.

2.2 Replication attempts

The first replication attempt after the original publication was published by Tinghög et al. (2013) who elicited the debate by five studies, including a direct as well as several conceptual replication analyses. The authors consistently found a null effect of manipulated decision time on cooperation behavior. Furthermore, they pointed to a potential selection bias in the original publication, as Rand et al. (2012) only included time compliant participants (i.e., participants who took too long in the time pressure condition or responded too early in the time delay condition were excluded).[9]

[9] Additionally, Tinghög et al. (2013) criticized Rand et al. (2012) for incorrectly controlling for compliance with the time pressure constraint in their analyses after having excluded non-compliant participants. Thereupon, Rand and colleagues replied that the effect persists when including participants who exceeded the time limit (Rand, Greene, & Nowak, 2013). Similarly, they argued that controlling for time limit compliance was not necessary to detect the effect. As subsequent replication studies provided contradictory evidence (Bouwmeester et al., 2017), this debate is still ongoing (Everett et al., 2017; Rand, 2017a) and is evaluated in more detail in the following of this chapter.

Verkoeijen and Bouwmeester (2014) corroborated the doubts regarding the replicability in that they also found no difference in cooperation behavior when manipulating decision time. Capraro and Cococcioni (2016) even showed increased cooperation under time delay when increasing the forced time to deliberate up to 30 seconds instead of 10 seconds. However, some studies also successfully replicated the spontaneous cooperation effect when manipulating an intuitive processing mode: Here, different manipulations were used to induce an intuitive processing mode; Protzko, Ouimette, and Schooler (2016) relied on the classic time pressure vs. time delay induction, whereas Lotz (2015) applied the priming manipulation of Rand et al. (2012) to induce intuitive vs. deliberate processing.

Results of replication attempts that measured decision time are similarly heterogeneous. Some correlative studies successfully replicated spontaneous cooperation behavior in social dilemmas (Lotito et al., 2013; Nielsen et al., 2014), whereas others even identified contrary results, showing increased decision time for cooperative behavior in a public goods game (S. Fiedler et al., 2013; Lohse, Goeschl, & Diederich, 2016).

This opaque situation required meta-analytical clarification and was addressed in two articles in *Psychological Science* that concentrated the debate to its current essence: First, Rand (2016) provided extensive, meta-analytic support for the spontaneous cooperation effect and, thus, evidence for the SHH. However, a many-labs registered replication report (Bouwmeester et al., 2017) conducted a direct replication of the time pressure study by Rand et al. (2012) and refreshed the debate about a potential selection bias (see Tinghög et al., 2013): The effect only appeared when excluding participants who did not obey the time limit constraint (i.e., took too long in the time pressure condition or responded too early in the time delay condition). Even though further analyses in Rand's reply (Rand, 2017a) and a subsequent article by Everett et al. (2017) countered the concern about a potential selection bias, the meta-analysis and registered replication report provide the largest datasets to date to evaluate the replicability of the spontaneous cooperation effect.[10]

Finally, the most recent direct replication attempt was made within a large scale replication project that aimed at replicating all feasible studies of two of the most prestigious interdisciplinary journals, *Science* and *Nature*, published between 2010 and 2015 (Camerer et al., 2018). The spontaneous cooperation effect was among the effects that have been tested for their replicability. Preventing the difficulty of a potential selection bias, the authors refrained from using time pressure to induce intuitive processing and instead relied on the priming manipulation of Rand et al. (2012). Still, the spontaneous cooperation effect was not found. However, Rand (2018) replied that against his advice, the authors did not include prior experience with economic games in their analyses, which he showed to be a moderator of spontaneous cooperation (see Rand et al., 2012). When only including inexperienced participants, there is indeed a similar effect size of spontaneous cooperation comparable to the original paper in 2012 (Rand, 2018).

Summarizing the status quo, replication results are mixed and potential reasons are mainly discussed on methodological rather than theoretical grounds. The largest replication from an independent author group is provided by the many-labs replication project (Bouwmeester et al., 2017), which shows that for a manipulated processing mode, the spontaneous cooperation effect is – if at all – small in size and requires the exclusion of non-compliant (or experienced, see Rand, 2018) participants to appear. Correlative studies measuring decision time as an approximation of the degree of deliberation, in contrast, find evidence of the effect in both directions, suggesting the interplay of moderators. Several studies have investigated interacting state and trait factors of spontaneous cooperation, which are

[10] Rand (2017) tackled in his comment the potential selection bias and analyzed whether there are any relations between additional variables assessed in the replication project (e.g., comprehension of payoff structure, prior experience with economic games) and time limit exceedance without finding any evidence of a selection bias. Everett et al. (2017) succeeded to design an experiment that reduces the percentage of non-compliant participants a priori, replicating reduced cooperation behavior under deliberation.

presented in the following. They lead towards the research question of the first paper, which investigates whether spontaneous cooperation is conditional on dispositional pro-sociality.

2.3 Identified moderators

The identification of moderators helps to unify some, at first glance, contradictory findings as presented in the above discussion of heterogeneous replication results concerning the spontaneous cooperation effect. As already indicated, experience with the laboratory social dilemma setting has been shown to moderate spontaneous cooperation in that only naïve individuals transfer their cooperative default to the lab setting (McAuliffe, Forster, Pedersen, & McCullough, 2018; Rand et al., 2012; Rand et al., 2014). A further situational moderator refers to the understanding of the social dilemma situation (Stromland, Tjotta, & Torsvik, 2018) as a necessary prerequisite for spontaneous cooperation. When individuals do not understand the monetary consequences of their decision, the difference in cooperation behavior between an intuitive vs. deliberate processing mode vanishes.

In addition to the interaction with experience in the laboratory setting, Rand et al. (2012) showed that only individuals with a highly trusting attitude towards the cooperativeness of their daily-life interaction partners cooperate spontaneously. In line with those findings, beliefs about the other players' cooperativeness were identified to moderate spontaneous cooperation: Santa et al. (2018) manipulated expectations and showed that cooperative choices are made faster with increasing (positive) beliefs. For defective choices, the opposite is true in that they become slower with increasing expected cooperativeness. The authors attribute the results to the importance of the social context that determines the speed of cooperation.

Turning to stable individual difference factors, the influence of cultural and demographic factors were tested for their moderating influence on spontaneous cooperation: Nishi, Christakis, and Rand (2017) showed the influence of intercultural differences in that Indians did not cooperate spontaneously and defected overall to a greater degree than US-Americans. The authors only speculate about the underlying reasons and focus on prosocial norms that are particularly developed in countries with high institutional quality (e.g., with a low level of corruption; Peysakhovich & Rand, 2015; Stagnaro, Arechar, & Rand, 2017).

Underlining the conceptual distinctiveness of altruism vs. cooperation behavior (see part of Chapter 1.1 on SVO), Rand, Brescoll, Everett, Capraro, and Barcelo (2016) showed that spontaneous altruism – in contrast to spontaneous cooperation (Rand, 2017b) – is conditional on gender. Specifically, there were gender differences regarding the intuitiveness of dictator game giving in that only females have internalized altruistic giving (Rand et al., 2016). In turn, men and women alike behaved intuitively prosocial in social dilemmas where cooperation increases the joint outcome in contrast to distributing a fixed amount as in dictator games (Rand, 2017b).

Continuing this perspective from whom spontaneous cooperation might be expected, spontaneous cooperation has also been investigated in terms of its dependence on basic personality traits. Specifically, Kieslich and Hilbig (2014) assessed spontaneous cooperation in a mouse-tracking paradigm. The authors showed that the effect is more pronounced for individuals high in honesty-humility, a basic trait of the six factor personality model HEXACO (Ashton & Lee, 2007; Lee & Ashton, 2004) reflecting active pro-sociality (Hilbig, Zettler, Leist, & Heydasch, 2013). In particular, honesty-humility represents the tendency to be fair and genuine in dealing with others, contrasted against greedy, insincere, and manipulative behavioral tendencies (Ashton & Lee, 2008).

To summarize, spontaneous cooperation was shown to be conditional on diverse situational and personality factors. As an attempt to systematize these and find common ground, the situational factors (e.g., positive expectations or trust in others' cooperativeness, existence of prosocial norms) point out that the fear of being exploited must be minimal in order to express spontaneous cooperation behavior. Individual differences (e.g., a dispositional pro-sociality as honesty-humility), in turn, represent person-situation interactions – an approach that the first paper elaborates on by investigating SVO as a moderator of spontaneous cooperation.

3. Unifying heterogeneous replication results – Spontaneous cooperation behavior is conditional on dispositional pro-sociality

Summarizing the vast amount of research on spontaneous cooperation, the reason why the original publication caught that amount of interest might be due to the strong contradiction of rational choice assumptions. Even though social preferences are widely acknowledged, the emphasis on the intuitive component of cooperation behavior contradicts the assumption that prosocial behavior needs effortful persuasion rather than constituting the default (e.g., Chen & Fischbacher, 2016; Piovesan & Wengström, 2009). Additionally, the effect settles in the time of the replication crisis in social psychology (Open Science Collaboration, 2015), where the scientific community speaks out against spectacular, but non-replicable findings.

Identifying person-situation interactions, for whom spontaneous cooperation is valid, allows one to unify heterogeneous replication results and provides the starting point for the research question of the first paper of this thesis (Mischkowski & Glöckner, 2016). Specifically, it is hypothesized that prosocials but not proselfs – as measured by SVO – cooperate spontaneously but override their prosocial default with longer deliberation. This was expected for several reasons: As outlined in Chapter 1.1, SVO (e.g., Van Lange, 1999) represents a disposition of how strongly individuals consider another persons' outcome that is stable over time (Camerer & Fehr, 2004; Murphy et al., 2011) and highly predictive of cooperation behavior (Balliet et al., 2009). Furthermore, SVO is expressed automatically: Cornelissen, Dewitte, and Warlop (2011) showed increased giving in dictator games for prosocials under cognitive load, indicating that even without available cognitive resources, prosocials comply with their social preference.

Additionally, SVO was shown to be related to honesty-humility (Hilbig, Glöckner, & Zettler, 2014; Mischkowski, Thielmann, & Glöckner, 2018, 2019); both concepts capturing dispositional pro-sociality (for the stability of SVO, see Murphy et al., 2011). Given that there is a cooperative default for individuals high in honesty-humility (Kieslich & Hilbig, 2014), it is straightforward to test whether this interaction effect is also valid for prosocials as measured by SVO. Related to the already identified moderators of trust and expectations regarding the cooperativeness of others, prosocials have more heterogeneous but also more positive expectations, on average, regarding other individuals' cooperativeness (Kelley & Stahelski, 1970; Van Lange, 1992). These interaction effects might therefore be rooted in differences in SVO.

To summarize, the aim of the first paper of this thesis was twofold. First, it is investigated whether the spontaneous cooperation effect is conditional on dispositional pro-sociality (i.e., SVO). Second, the three studies consist of a direct replication of the first study by Rand et al. (2012) that was only extended by an assessment of SVO. This made it possible to similarly test for the main effect that might appear given a predominant number of prosocials in the studies. To validate the results of SVO with the related but broader basic personality trait dimension honesty-humility, the spontaneous cooperation effect for individuals high in honesty-humility as shown by Kieslich and Hilbig (2014) was tested for its replicability. To do so, the study with the highest power containing a representative sample for the US and German population also included the HEXACO personality inventory (Ashton & Lee, 2007; Ashton & Lee, 2008). Doing so made it also possible to test whether SVO is predictive of spontaneous cooperation beyond honesty-humility – that is, whether the potential interaction effect of spontaneous cooperation for prosocials persists when controlling for the interaction between honesty-humility and decision time and vice versa.

3.1 1st Article: Spontaneous cooperation for prosocials but not for proselfs: Social value orientation moderates spontaneous cooperation behavior

Materials, analysis scripts, and data can be found at: https://osf.io/w7hsk/

Supplementary information can be found at: https://media.nature.com/original/nature-assets/srep/2016/160215/srep21555/extref/srep21555-s1.pdf

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Received: 26 August 2015
Accepted: 23 December 2015
Published: 15 February 2016

Spontaneous cooperation for prosocials, but not for proselfs: Social value orientation moderates spontaneous cooperation behavior

Dorothee Mischkowski[1] & Andreas Glöckner[2,3]

Cooperation is essential for the success of societies and there is an ongoing debate whether individuals have therefore developed a general spontaneous tendency to cooperate or not. Findings that cooperative behavior is related to shorter decision times provide support for the spontaneous cooperation effect, although contrary results have also been reported. We show that cooperative behavior is better described as person × situation interaction, in that there is a spontaneous cooperation effect for prosocial but not for proself persons. In three studies, one involving population representative samples from the US and Germany, we found that cooperation in a public good game is dependent on an interaction between individuals' social value orientation and decision time. Increasing deliberation about the dilemma situation does not affect persons that are selfish to begin with, but it is related to decreasing cooperation for prosocial persons that gain positive utility from outcomes of others and score high on the related general personality trait honesty/humility. Our results demonstrate that the spontaneous cooperation hypothesis has to be qualified in that it is limited to persons with a specific personality and social values. Furthermore, they allow reconciling conflicting previous findings by identifying an important moderator for the effect.

Cooperation is important for societies to achieve collective goals[1]. In order to be able to foster cooperation a detailed understanding of the processes underlying individuals' decisions to cooperate (or not) is essential and research has increasingly focused on investigating these cognitive processes in more detail[2–5]. One core matter of debate has been whether individuals have a spontaneous tendency to cooperate and findings both in favor[3,4,6] and in conflict[2,7,8] with this hypothesis have been reported. This inconsistent evidence indicates that important moderators that might lie in the situation, the person or their interaction have not been identified yet.

Previous experience with the experimental setting of social dilemma games has been found to reduce the effect[4], and on theoretical[9] and empirical[10,11] grounds it can be expected that other situational factors that influence the activation of social norms such as situation framing affect spontaneous behavior as well. Tests for effects of individual difference moderators have (aside from several null findings) revealed that the spontaneous cooperation effect is observable mainly for persons trusting in the cooperativeness of their daily life interaction partners and it is not observed for persons with experience in related studies[4]. Still a comprehensive analysis of personality moderators is largely missing.

We guide the perspective towards social value orientation (SVO), which is a continuous measure of social preferences and can be expected to be an important moderator for the spontaneous cooperation hypothesis on theoretical grounds[3,12,13]. Specifically, we hypothesize that the spontaneous cooperation effect might be driven by prosocials–having a default willingness to cooperate, which tends to disappear with longer deliberation about the dilemma structure of the situation. No effect of time is expected for proself persons. Based on the findings concerning spontaneous cooperation and taking into account that SVO has been established as a strong predictor of cooperation[14,15], we expect main effects for SVO, decision time and–most importantly–their interaction on cooperation behavior.

[1]University of Göttingen, Department of Psychology, Gosslerstrasse 14, D-37073 Göttingen, Germany. [2]University of Hagen, Department of Psychology, Universitaetsstrasse 27, D-58097 Hagen, Germany. [3]Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn, Germany. Correspondence and requests for materials should be addressed to A.G. (email: andreas.gloeckner@fernuni-hagen.de)

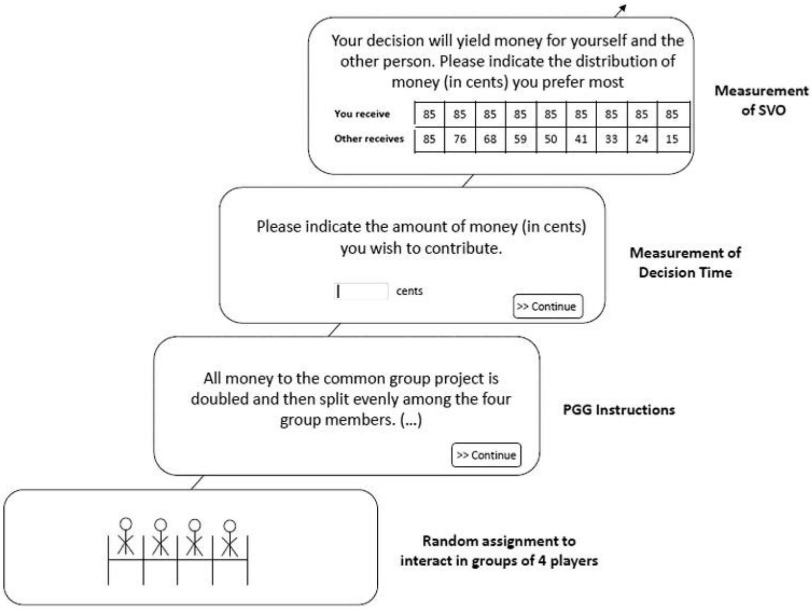

Figure 1. Procedure of the studies. Note: PGG refers to Public Goods Game, SVO stands for Social Value Orientation.

Method

We conducted three studies including one lab and two online experiments. The studies were approved by the local ethics committee of Göttingen University's Psychological Department and conducted in accordance with the approved guidelines. Before conducting the studies, we obtained informed consent from all subjects. With the methods and instructions used, we strictly replicated the previous core investigation by Rand and colleagues[4], except for some important extensions. Participants played one-shot public good games (PGG) in groups of four. In the PGG, independent and anonymous individual contributions of each member were doubled and split evenly among the members of the group.

The basic procedure is described in Fig. 1. After group assignment, participants were presented with an instruction screen describing the structure and the rules of the game. After reading the screen, participants clicked a continue button to proceed to the response screen, which contained the request to indicate the amount of money (in cents) they wish to contribute. Participants indicated their contributions by typing a number in the respective field and confirmed their input with a mouse click on a continue button. The time between onset of the response screen and the confirmation of the contribution by mouse click constituted our main predictor variable decision time. Hence, decision time exclusively captures the time required for generating the response concerning how much to contribute. The reading time of the Public Goods Game instructions was explicitly detached by displaying instruction on a separate screen.

Decision time was measured via the programs used for data collection (i.e.'Unipark' for the online studies and 'Bonn Experiment System' for the lab study). In the first study we analyzed data of 134 Amazon Mechanical Turk (MTurk) workers (51 female, mean age = 30.5 years) half of them from the United States and the remaining from other western countries and India. Second, we elaborated these results in a more controlled setting and conducted a laboratory study with mostly German students (N = 105, 70 female, mean age = 26.97 years). Finally, as a third experiment, we conducted a high-power online study with representative samples concerning age and gender for the US (N = 249) and German populations (N = 255), resulting in a total sample of N = 504 (258 female, mean age = 46.13 years). Data for the third experiment was gathered via a professional panel provider. The endowment for the PGG varied between studies (i.e., USD 0.40 in the first study conducted on MTurk, 2.00€ (approx. USD 2.70) in the lab experiment and USD 1.50 in the third study (Panel)) but the multiplication factor of two for contributed money remained constant.

In line with previous studies we additionally gathered data about beliefs (the expectations about the other players' cooperation behavior) as well as cooperativeness of interaction partners in daily life and previous experience with the experimental setting of social dilemma games. As an important extension, we additionally measured social value orientation using the SVO Slider Measure[16] at the end of the study, which allows to calculate a continuous SVO angle value. An angle of around zero indicates proself persons; these are persons that maximize their own outcomes without considering outcomes of others. Positive values indicate more prosocial behavior in that people gain positive utility from outcomes of others. The SVO measure was incentivized and it was common knowledge that it was determined by a random draw whether the PGG or the SVO were incentivized. In order

Contribution (in %)	Overall Analysis	Study 1 (MTurk)	Study 2 (Lab)	Study 3 (Panel)
Decision Time (DT in log10(sec))	−54.52*** (−3.94)	−25.64 (−1.20)	−42.78 (−1.02)	−64.76** (−3.48)
Social Value Orientation (SVO angle in degree)	3.19*** (11.39)	3.86*** (6.80)	2.71*** (3.94)	3.00*** (8.39)
Interaction of DT * SVO	−2.90** (−3.19)	−2.14 (−1.40)	−4.80+ (−1.77)	−2.52* (−1.98)
Constant	66.37*** (7.59)	42.23*** (5.81)	75.05*** (7.37)	100.74*** (19.25)
Observations	743	134	105	504
Pseudo R2	0.052	0.083	0.041	0.036

Table 1. Tobit regression of decision times and Social Value Orientation on contributions. $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ (two-sided). Note. *t*-values are presented in parentheses; predictors are centered. Overall analysis also includes two study dummies which are not reported.

to check for correlations with broader personality factors, we included the 60 items general personality questionnaire HEXACO at the end of the third study[17]. More details on the procedure and the fully instruction are provided in the supplementary online materials.

Results

In an overall analysis of the three experiments (Table 1), we find support for the spontaneous cooperation hypothesis in that contributions increase with decreasing decision time (see Fig. S1). Increasing decision times from 3.16 sec ($10^{0.5} = 3.16$) to 31.6 sec ($10^{1.5} = 31.6$) was accompanied by a decrease of cooperation by 54.52% according to a Tobit regression and by 19.08% according to an OLS regression (see Table S2). Considering the three studies separately, we replicate the spontaneous cooperation main effect in Study 3 (the representative online sample study) and find trends in the same direction for the other two studies with lower power. As a side effect, one should note that there are roughly 10% differences in the average contributions among studies, being highest in the panel sample (67.10%) and lowest in the MTurk sample (44.59%; one-way ANOVA: $F(2, 740) = 17.21$, $p < 0.001$). This difference was paralleled by a difference in social value orientation (one-way ANOVA: $F(2, 740) = 30.68$, $p < 0.001$) with around 5° of SVO angle differences between studies. The parallel development of contributions and SVO is–as expected–reflected in a strong correlation between both measures ($r = 0.48$, $p < 0.001$). Hence, differences in social values (social preferences) among the different samples used in our studies have most likely driven the level differences in contributions. To correct for these differences, we included study dummies as control variables in the overall analysis of contributions.

Most importantly, in this overall analysis we find the expected interaction of decision time and SVO (Fig. 2). In the separate analyses comparable effects are found for all three studies although in the MTurk study the effect fails to reach conventional levels of significance.

Separate analyses for prosocials (SVO angle > 22.45°) and proselfs (SVO angle < 22.45°) reveal that the spontaneous cooperation effect is only observed for prosocials (Tobit and OLS overall: $p < 0.001$) but does not hold for proselfs (Tobit: $p = 0.836$; OLS: $p = 0.475$). This also holds true in separate analyses for the three studies (see Table S3). Furthermore, the observed interaction also holds when controlling for the previously observed moderators cooperativeness of daily life interaction partners and experience in related studies (Table S4).

To validate our results also on the basis of a broader personality construct, we measured Honesty-Humility as one of the six factor personality scale HEXACO[17,18]. Honesty-Humility represents “the tendency to be fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them without suffering retaliation” (p.156)[18] and has shown to be highly predictive of cooperation behavior[19,20]. Again, we find the same interaction pattern in that cooperation decreases with decision time for people high on Honesty-Humility only but not for people low in this personality trait (Table S5). As in previous studies, SVO and Honesty-Humility are moderately correlated ($r = 0.22$, $p < 0.001$), but the interaction effect persists on a marginal significance level also when controlling for SVO and its interaction with decision time ($p = 0.064$). The observed interaction concerning the specific construct social values therefore also generalizes to the related but broader personality construct Honesty-Humility.

Discussion

In a set of studies we show that the controversially debated spontaneous cooperation effect replicates in population representative samples and beyond. Given its small effect size, large samples are needed in order to find the effect on a conventional significance level. Crucially, however, we show that the spontaneous cooperation effect is also strongly dependent on personal characteristics. For persons that are prosocial and high on the personality factor Honesty-Humility cooperation decreased with increasing decision time. These persons drive the spontaneous cooperation effect reported in some of the previous studies. No relation between decision time and cooperation exists for selfish persons. The demonstration of this moderator allows reconciling previous diverging results.

The finding that only prosocials show the spontaneous cooperation effect is consistent with the explanation that cooperation might be their default response, which is sometimes overwritten by increased deliberation. Given that prosocials are mainly conditional cooperators that are particularly sensitive to changes in beliefs concerning the behavior of other group members[21] deliberate reflection about the game structure might decrease

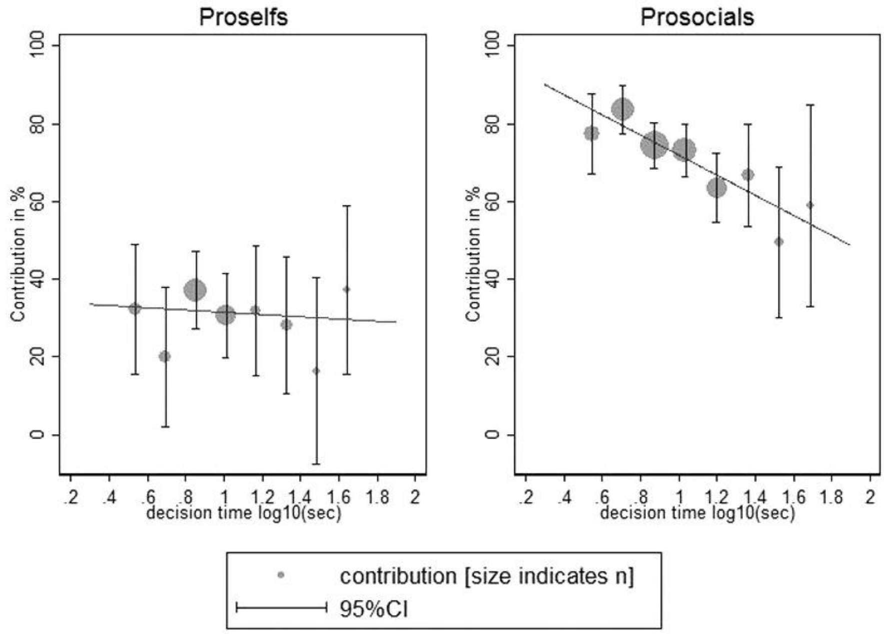

Figure 2. There is a spontaneous cooperation effect for prosocials (SVO angle >22.45°), but not for proself persons (SVO angle <22.45°). Prosocials contribute less the longer the decision time (decision time between 2 and 72 seconds). For proselfs, however, there is no difference in cooperation behavior, independent of decision time they stick to low contributions. Regression lines represent the effect of reaction times on contributions from separate OLS regressions. Whiskers denote the 95% confidence interval; the diameter of the grey dots is proportional to the sample size.

their contributions. No such adjustments can be expected for proself persons. Still, given the correlational design of our research such causal interpretations of processes cannot be directly tested, which is due to further research.

References

1. Hardin, G. The tragedy of the commons. *Science* **162,** 1243–1248, doi: 10.1126/science.162.3859.1243 (1968).
2. Fiedler, S., Glöckner, A., Nicklisch, A. & Dickert, S. Social Value Orientation and information search in social dilemmas: An eye-tracking analysis. *Organ. Behav. Hum. Decis. Process*. **120,** 272–284, doi: 10.1016/j.obhdp.2012.07.002 (2013).
3. Rand, D. G. *et al.* Social heuristics shape intuitive cooperation. *Nat. Commun*. **5,** 1–12, doi: 101038/ncomms4677 (2014).
4. Rand, D. G., Greene, J. D. & Nowak, M. A. Spontaneous giving and calculated greed. *Nature* **489,** 427–430, doi: 10.1038/nature11467 (2012).
5. De Dreu, C. K. W. *et al.* The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* **328,** 1408–1411, doi: 10.1126/science.1189047 (2010).
6. Nielsen, U. H., Tyran, J.-R. & Wengström, E. Second thoughts on free riding. *Econ. Letters* **122,** 136–139, doi: 10.1016/j.econlet.2013.11.021 (2014).
7. Tinghög, G. *et al.* Intuition and cooperation reconsidered. *Nature* **498,** 427–430, doi: 10.1038/nature12194 (2013).
8. Verkoeijen, P. & Bouwmeester, S. Does intuition cause cooperation. *PLoS One* **9,** e96654, doi: 10.1371/journal.pone.0096654 (2014).
9. Dufwenberg, M., Gächter, S. & Hennig-Schmidt, H. The framing of games and the psychology of play. *Games Econ. Behav*. **73,** 459–478, doi: 10.1016/j.geb.2011.02.003 (2011).
10. Engel, C. & Rand, D. G. What does "clean" really mean? The implicit framing of decontextualized experiments. *Econ. Letters* **122** 386–389, doi: 10.1016/j.econlet.2013.12.020 (2014).
11. Liberman, V., Samuels, S. M. & Ross, L. The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Pers. Soc. Psychol. B*. **30** 1175–1185, doi: 10.1177/0146167204264004 (2004).
12. Bogaert, S., Boone, C. & Declerck, C. Social value orientation and cooperation in social dilemmas: A review and conceptual model. *Br. J. Soc. Psychol.* **47,** 453–480, doi: 10.1348/014466607X244970 (2008).
13. Cornelissen, G., Dewitte, S. & Warlop, L. Are social value orientations expressed automatically? Decision making in the dictator game. *Pers. Soc. Psychol. B*. **37,** 1080–1090, doi: 10.1177/0146167211405996 (2011).
14. Balliet, D., Parks, C. & Joireman, J. Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group. Process. Intergroup. Relat*. **12,** 533–547, doi: 10.1177/1368430209105040 (2009).
15. Van Lange, P. The Pursuit of Joint Outcomes and Equality in Outcomes: An Integrative Model of Social Value Orientation. *J. Pers. Soc. Psychol.* **77,** 337–349, doi: 10.1037/0022-3514.77.2.337 (1999).
16. Murphy, R. O., Ackermann, K. A. & Handgraaf, M. J. J. Measuring Social Value Orientation. *Judgm. Decis. Mak*. **6,** 771–781, doi: 10.2139/ssrn.1804189 (2011).
17. Ashton, M. C. & Lee, K. The HEXACO model of personality structure and the importance of the H factor. *Soc. Personal. Psychol. Compass* **2,** 1952–1962, doi: 10.1111/j.1751-9004.2008.00134.x (2008).

18. Ashton, M. C. & Lee, K. Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Pers. Soc. Psychol. Rev.* **11,** 150–166, doi: 10.1177/1088868306294907 (2007).
19. Hilbig, B. E., Glöckner, A. & Zettler, I. Personality and pro-social behavior: Linking basic traits and Social Value Orientations. *J. Pers. Soc. Psychol.* **107,** 529–539, doi: 10.1037/a0036074 (2014).
20. Zettler, I., Hilbig, B. E. & Heydasch, T. Two sides of one coin: Honesty–Humility and situational factors mutually shape social dilemma decision making. *J. Res. Pers.* **47,** 286–295, doi: 10.1016/j.jrp.2013.01.012 (2013).
21. Kelley, H. H. & Stahelski, A. J. Social interaction basis of cooperators' and competitors' beliefs about others. *J. Pers. Soc. Psychol.* **16,** 66–91, doi: 10.1037/h0029849 (1970).

Acknowledgements

We thank the University of Hagen for covering open-access publication fees.

Author Contributions

D.M. and A.G. designed the study and discussed the results. D.M. organized the studies, analyzed the data and wrote a first draft of the paper, which was revised by A.G.

Additional Information

Data Availability: Data and materials are also available online at Open Science Framework (https://osf.io/w7hsk/).

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Mischkowski, D. and Glöckner, A. Spontaneous cooperation for prosocials, but not for proselfs: Social value orientation moderates spontaneous cooperation behavior. *Sci. Rep.* **6**, 21555; doi: 10.1038/srep21555 (2016).

3.2 Discussion

The first paper supports a cooperative default for dispositional prosocials, meaning that these individuals cooperate spontaneously but reduce their cooperation behavior with increasing decision time. Proselfs, in turn, consistently behave in a more selfish manner, irrespective of decision time. As dispositional pro-sociality may vary among different samples used to investigate spontaneous cooperation, some of the failed replications (i.e., null effects) might be the result of a predominantly proself sample.

3.2.1 Differentiating between decision conflict and intuitive processing

It is still unclear which concepts are operationalized by self-paced decision time, namely whether it confounds decision conflict and intuitive processing (see Chapter 1.2). To reiterate, short decision times do not necessarily (only) capture intuitive processing but were also shown to reflect low decision conflict (A. M. Evans et al., 2015; Krajbich et al., 2015). Specifically, A. M. Evans et al. (2015) provide evidence for an inverse u-shaped pattern when predicting decision time with cooperation behavior: Both fully cooperative and defective choices (e.g., contributing everything or nothing to a public good) were made more quickly than partly cooperative choices (e.g., contributing about half of the amount to the public good). This behavioral pattern is predicted by single process models (e.g., evidence accumulation models) that attribute longer decision times to increased decision conflict (e.g., Klauer, 2014) and, therefore, challenge the common practice of inferring how intuitive vs. deliberate individuals decided dependent on their self-paced decision time. Additionally, A. M. Evans et al. (2015) showed that decision time mediates the relation between self-indicated decision conflict and decision extremities in that low conflict decisions were made more quickly and led to more extreme (that is highly cooperative and highly defective) decisions. The conceptual distinction between measured and manipulated decision time was further supported by the fact that manipulating decision time increased cooperation behavior – as suggested by the SHH – but it did not influence decision conflict or decision extremes. Taken together, this evidence supports the so-called *decision conflict hypothesis*, which led to the conclusion that a reverse inference on the processing mode based on decision times is invalid.

The studies conducted in Mischkowski and Glöckner (2016) allow one to test whether the findings by A. M. Evans et al. (2015) are replicable. Therefore, I conducted the corresponding additional analyses, checking for a potential inverse u-shaped pattern of the relation between decision time and cooperation behavior. Specifically, this required two additional analyses. First, I applied the regression model by A. M. Evans et al. (2015), testing whether decision extremes as absolute deviations from average cooperation behavior (i.e., the absolute difference from the descriptive scale mean of contributions) are faster than intermediate contributions. Second, as a robustness check, I tested whether the relation between contributions and decision times is (inversely) u-shaped. Specifically, this was accomplished by including a quadratic term of contributions as an additional predictor in the regression model predicting decision time by contributions.

Descriptive statistics of the relevant variables are comprehensively provided in Table 1. Concerning the first analysis, I relied on the same regression model used by A. M. Evans et al. (2015) to investigate whether extreme (i.e., very high *and* very low) contributions are faster. Therefore, logarithmized decision time was regressed on the contributions as well as on the absolute difference of contributions to the scale mean.[11] Similar to the findings by A. M. Evans et al. (2015), this analysis showed a negative relationship between decision extremes and decision time (β = -.10, p = .009). Note, however, that the negative relation between contributions and decision time (i.e., the spontaneous cooperation effect) was observed, as well (β = -.16, p < .001), even when controlling for decision extremes. This is in line with the explanation that in largely prosocial samples (as it is true for the data of Mischkowski and

[11] Following A. M. Evans et al. (2015), both predictors were centered on the sample mean. Furthermore, consistent with the analyses in Mischkowski and Glöckner (2016), two study dummies are included to control for potential differences among the three studies. In addition, decision time was logarithmized to approximate a normal distribution (Mayerl, 2009).

Glöckner (2016) where 70.9% in the overall analysis were prosocial according to their SVO angle), the relation between decision time and contributions is negative (A. M. Evans et al., 2015).

Table 1. Descriptive statistics for the relevant variables of the additional analyses

	Mean (*SD*)	Min	Max
Contributions (in %)	61.47 (41.71)	0	100
Decision extremes	38.50 (15.99)	1.47	61.47
Decision Time (in sec.)	11.90 (10.52)	2.00	86.00
Decision Time (log 10 sec.)	0.98 (0.27)	0.30	1.93

Note. Decision extremes consist of the absolute difference of contributions to the scale mean.

Even though the operationalization of decision extremes and their negative relation with decision time hints toward an inverse u-shaped pattern between contributions and decision time, it does not explicitly test for the curvilinear trend. The most common procedure to test such a pattern implies including a squared term of the relevant predictor (in this case, contributions) in the regression model (Simonsohn, 2018). However, this procedure has been criticized in the past, arguing that a significant squared term does not provide sufficient evidence to infer a u-shaped pattern and might thus lead to false, alpha-inflated conclusions (Lind & Mehlum, 2010; Simonsohn, 2018). Following the approach of Lind and Mehlum (2010), testing for significance of the squared term and analyzing whether the estimated extremum of the curve is within the data range only represents necessary but not sufficient evidence. To account for the latter, one must additionally test whether the steepness of the curve (i.e., the slope) significantly increases on the left-hand side of the extremum point and decreases on its right-hand side.[12] The corresponding analyses were conducted using the Stata module provided by Lind and Mehlum

[12] Specifically, Lind and Mehlum (2007, 2010) test the slope of the tangent at the lowest and highest value of the predictor (i.e., at the minimum and maximum of contributions).

(2007). The results are in line with the previous finding of decision extremes: When regressing logarithmized decision time on mean-centered contributions and their squared term, there is a small-sized negative relation between the squared contributions and decision time, providing preliminary evidence of an inverse u-shaped pattern (β = -.12, p = .009, see Figure 5).[13]

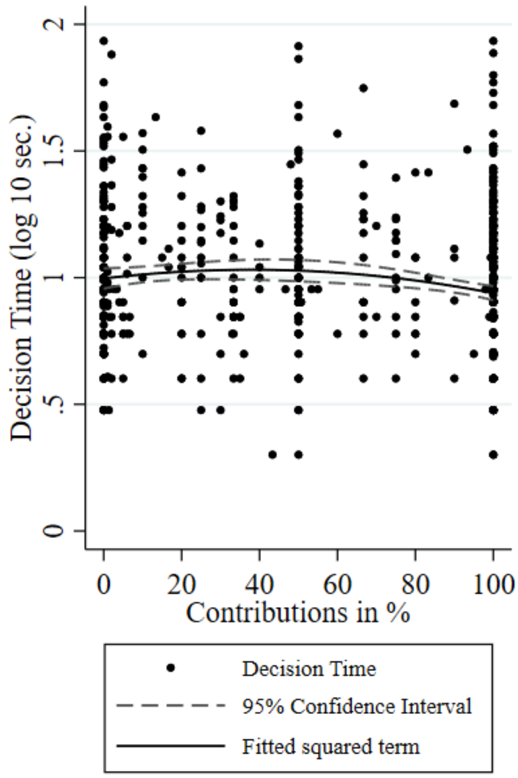

Figure 5. Prediction of decision time as a function of contributions and their squared term. Grey-dashed lines denote the 95% confidence interval.

More specifically, there is a significant positive slope on the left hand-side of the extremum point of cooperative choices predicting decision time (b = .002, p = .03), the extremum being at 38.99% of contributions.[14] Correspondingly, the upper bound indicates a significant decrease in contributions (b = -.003, p < .001, see Figure 5).

[13] Contributions are mean-centered to avoid multicollinearity to their squared term. For illustrative purposes, the graph contains non-centered contributions. The regression model again includes two study dummies. Note, that the effect of contributions on decision time remains significant (β = -.18, p < .001), indicating that the linear relation persists even when controlling for the curvilinear trend.
[14] The reported extremum is corrected for the linear shift of mean-centered contributions.

3.2.2 Implications for spontaneous cooperation (of prosocials)

Summarizing these results in terms of their implications, they support – despite the small effect size – the assumption that decision conflict and intuitive processing mode are entangled when measuring decision time. Not only cooperative decisions but also defective choices as compared to intermediate choices are made faster. Thus, the data of the thesis' first article supports the terminology of *spontaneous* (rather than intuitive) cooperation. However, the inverted u-shaped pattern as an indicator of decision conflict is only one part of the results of A. M. Evans et al. (2015) and does not sufficiently allow for testing the confounding influence of decision conflict. For instance, the data of the first article did not contain any self-rating measures of decision conflict and thus cannot investigate whether these would relate to decision time. In addition, the distinction between decision conflict and intuitive processing also rests on the finding that cooperation, but not decision extremes, increase when individuals are under time pressure (A. M. Evans et al., 2015). However, given that the spontaneous cooperation effect is not reliably found (see Chapter 2.2) and highly debated in terms of a potential selection bias (see Bouwmeester et al., 2017; Tinghög et al., 2013), the possibility remains that decision conflict is also influential when manipulating decision time: Future research could investigate whether mainly those individuals who experience high decision conflict exceed the time pressure limit, as the arbitrary decision time limit of ten seconds chosen by Rand et al. (2012) still offers the potential that individuals may experience some degree of decision conflict and deliberation (Myrseth & Wollbrant, 2017). If so, there might be a correlation between decision conflict and time limit exceedance. Given that the spontaneous cooperation effect is only found for those individuals who comply with the time constraint (Bouwmeester et al., 2017), low decision conflict might be a prerequisite for spontaneous cooperation – be it for self-paced or manipulated decision time.

In addition, marginalizing potential differences between measured vs. manipulated decision time on theoretical grounds, one must keep in mind that manipulating an intuitive processing mode does not shape the default but rather triggers a reliance on it. Stated differently, putting individuals under time pressure does not inherently increase pro-sociality as Rand et al. (2012) suggested, but only causes individuals to rely on their disposition, which can imply any social preference. As outlined in the following, future research should attempt to combine rather than separate investigations on cognitive processes underlying cooperation behavior.

3.2.3 Further development of the spontaneous cooperation effect and the Social Heuristics Hypothesis

Following the results of A. M. Evans et al. (2015) and Krajbich et al. (2015), Rand (2016) considered only manipulated but not measured decision time as a valid operationalization for an intuitive, respectively deliberate processing mode. Given the heterogeneous replication results for the spontaneous cooperation effect – also when manipulating the processing mode – the subsequent discussion in the field of spontaneous cooperation mainly focuses on debating the right (usage of) processing mode manipulation. However, hardly any attention is directed towards the underlying theory. On an empirical level, the focus is directed towards replication studies – which is an important step in itself (see Open Science Collaboration, 2015) – knowing that the SHH will at some point lose its right to exist when its most prominent effect (the effect on which the theory was built post-hoc) is not replicable. However, the pure methodological discussion appears to be insufficient for evaluating and revising the theory. To move the debate forward on theoretical grounds, it is important to assess how to improve (i.e., revise) the theory underlying the spontaneous cooperation effect. To do so, the following chapter deals with the evaluation of the SHH based on scientific-theoretical criteria of the critical rationalism, namely the empirical content as described by Karl Popper (1934/2005). Thereby, the articles of this thesis can be embedded in a more general analysis of what the SHH provides in terms of its theoretical value (i.e., its empirical content) as well as how the results of the first paper and the hypotheses of the second paper can improve it.

4. From spontaneous cooperation to spontaneous punishment

– Towards a more general theory of spontaneous pro-sociality?

From a meta-scientific perspective, the debate about the spontaneous cooperation effect and its corresponding theory, the SHH, demonstrates in a prototypical way the phenomenon in (social) psychological research of focusing on proving effects rather than testing theories (K. Fiedler, 2004; Kruglanski, 2004). As many of these "effects" were not replicable, the question was raised whether psychology was caught in a replication crisis (Pashler & Wagenmakers, 2012; Yong, 2012), initiating diverse replication initiatives not only in the field of social psychology (e.g., Camerer et al., 2018; Open Science Collaboration, 2015). The following chapter reflects on the core of replicability by focusing on the need of theorizing based on the specific examples of the spontaneous cooperation effect and the underlying SHH. Based on an analysis of its original formulation (Rand et al., 2014), the accomplishments of research on spontaneous cooperation in terms of improving its empirical content and explaining what future research must do to develop the theory further become visible. Therefore, this chapter has two aims: First, it evaluates the SHH as underlying theory of the spontaneous cooperation effect according to the criterion of its empirical content (Popper, 1934/2005). Based on this evaluation, it delineates the research question of this thesis' second paper, investigating the generalizability of spontaneous cooperation towards a (potential) theory of spontaneous *pro-sociality*.

Starting with a chronological retrospection, the spontaneous cooperation effect was published first (Rand et al., 2012), before inductively inferring the SHH (Rand et al., 2014). In line with a "Confirm-early/ disconfirm late" strategy (Manktelow, 2012), Rand and colleagues primarily published confirmatory results in the following years, making use of re-analysis of large datasets that provide support for the SHH (e.g., Cone & Rand, 2014; Nishi et al., 2017;

Rand, 2016, 2017b; Rand et al., 2016).[15] Several direct and conceptual replication attempts – both successful and unsuccessful (e.g., Lotz, 2015; Tinghög et al., 2013, see Chapter 2.2) – focused on the effect rather than the underlying theory. However, this effect-focused debate concerning heterogeneous replication results shifted towards a methodological dispute about how to operationalize and manipulate processing mode (Bouwmeester et al., 2017; Rand, 2017a), rather than questioning the quality of the theory. Up until now, the methodological debate about the SHH appears to be circular: Either there is confirmatory support for the theory (e.g., Everett et al., 2017; Isler et al., 2018) or seemingly methodological flaws prevent finding supportive evidence (see Mosleh & Rand, 2018; Rand, 2018). This reasoning, however, is highly dangerous when keeping in mind that science should primarily seek to maximize the gain of knowledge by falsifying wrong theories (Popper, 1934/2005).[16]

The criteria to evaluate the value of a scientific theory – or even the legitimacy to call it scientific – were introduced by Karl Popper (1934/2005). His concept of a theory's empirical content provides the opportunity to exit this circular debate and focus on how modifying the SHH based on previous results increases the theory's empirical content. Specifically, an analysis of a theory's empirical content can be accomplished by considering two criteria: the question about (i) which individuals and situations the theory can be applied to (i.e., its scope conditions or level of universality or generality) and (ii) which behavioral patterns the theory actually forbids in its areas of application (its degree of precision or specificity).

In line with this reasoning, I will analyze the SHH according to its empirical content in the following and outline how identifying social value orientation (SVO) as a moderator of spontaneous cooperation increases its specificity. Transitioning to the second paper, this similarly aims at contributing to the SHH's empirical content, though with a different focus on

[15] This strategy shall not be criticized; it is often observed (Dunbar, 1995) and, more importantly, also rational in terms of confirming a theory's right to exist first before determining its boundaries.

[16] Note that Popper (1934/2005) considers a priori every theory to be wrong, as theories consist of simplified reflections of reality. Falsification thus refers to the aim of disproving one's theory by testing its boundaries – or as put by Platt (1964) by designing an experiment that could disprove one's hypothesis rather than confirm it.

its generalizability (via testing its level of universality). Specifically, it investigates whether the negative relation between decision time and elementary cooperation also applies for instrumental cooperation, that is for punishment behavior.[17]

Starting with the analysis of the SHH's empirical content requires the identification of the theory's original reference, in this case the 2014 article of Rand et al. "Social heuristics shape intuitive cooperation". The description of the SHH consists of a rather short post-hoc explanation about the spontaneous cooperation effect and its underlying moderators identified in 2012 (e.g., trust in cooperativeness of daily life interaction partners and experience with the laboratory setting of social dilemma experiments, see Chapter 2.3 of this thesis). Still, in the following, I attempt to formalize the original version of the theory (Rand et al., 2014) by defining its situational and dispositional antecedents that elicit, when fulfilled, the hypothesized behavioral consequences (see Table 2). In a second step, it is pointed out how the theory develops (i.e., how the antecedents and consequences change), while taking the evidence of the first paper and the hypotheses of the second paper into account.

[17] To clarify, the empirical content of a theory is not empirically tested. However, given that the SHH is vague in its specification and mute in terms of its applicability beyond elementary cooperation behavior, the aim of both articles of this thesis is to consider reasonable specifications and further applicable areas and test these empirically. This allows one to revise the theory on an iterative basis with an increased empirical content.

Table 2. Increase in empirical content when modifying the SHH in accordance with identified and hypothesized moderators

Article	Increase in empirical content	Antecedent	Consequence
Rand et al. (2014)	Reference point: The original SHH as stated by Rand et al. (2014)	If individuals decide in a one-shot, anonymous social dilemma whether to cooperate or not,	then spontaneous decisions are more or similarly cooperative – though never more defective – in comparison to deliberate decisions.
1st article (Mischkowski & Glöckner, 2016)	Increased specificity by identifying SVO as a moderator for spontaneous cooperation	If individuals decide in a one-shot, anonymous social dilemma whether to cooperate or not,	then prosocials are more likely to cooperate spontaneously, whereas proselfs are expected to defect irrespective of decision time.
2nd article (Mischkowski, Glöckner, et al., 2018), 1st hypothesis	Increased level of universality given spontaneous cooperation in first and second order public goods	If any kind of cooperative choice (i.e., elementary cooperation or punishment as instrumental cooperation) is made in a one-shot, anonymous social dilemma,	then spontaneous decisions are more or similarly cooperative – though never more defective – in comparison to deliberate decisions.
2nd article, 2nd hypothesis	Increased level of universality as well as increased specificity given spontaneous cooperation and punishment being only valid for prosocials	If any kind of cooperative choice (i.e., elementary or instrumental cooperation) is made in a one-shot, anonymous social dilemma,	then prosocials are more likely to cooperate spontaneously, whereas proselfs are expected to defect irrespective of decision time.
2nd article, 3rd hypothesis	None	If individuals decide in a one-shot, anonymous social dilemma whether to punish or not,	then punishment decreases with decision time for highly upset individuals, whereas individuals low in negative affect are expected to invest few resources to punish irrespective of decision time.

As Glöckner and Betsch (2011) point out, formalizing a scientific theory according to Karl Popper implies defining its antecedents and consequences. The antecedents refer to the areas of applications, reflecting the theory's generality in terms of the situational and dispositional properties the theory makes predictions about. In turn, a theory's specificity describes what it actually forbids, that is, which (e.g., behavioral) pattern would contradict and thus disprove the theory. According to Popper (1934/2005), scientific theories relate both components in a set of general implications that follow the form where all values of an object *'x'* (e.g., individuals) that satisfy the antecedent function φ(x) also fulfill the statement function of the consequences *f*(x). This means, simply put, that if a theory's area of application is met, its prediction(s) must hold.

Transferring these abstract considerations to the SHH illustrates how useful they are to evaluate theories: Reflecting on the SHH's antecedents (i.e., its generality), it is inherently limited to observations of cooperation behavior in one-shot or the first round of anonymously conducted social dilemmas that generate a clear conflict between self- and collective interest. Even though further pro-sociality measures have been investigated in their dependence on processing mode (e.g., cheating behavior and social mindfulness, Köbis, Verschuere, Bereby-Meyer, Rand, & Shalvi, 2019; Mischkowski, Thielmann, et al., 2018), the SHH in its original description solely refers to (elementary) cooperation behavior as measured in economic games entailing a social dilemma (Rand et al., 2014). Thus, the situations the theory applies to are rather limited, only allowing variation in the operationalization of cooperation behavior. Rand (2016) considers "the four canonical games" (p. 1194) as valid operationalizations of cooperation behavior, such as one-shot prisoner's dilemmas, public good games, trust games, and the proposer's behavior in ultimatum games (Rand, 2016).[18] Selecting these games was a

[18] The trust game consists of a dilemma situation between two players, where the first person (the trustor) sends any amount of her endowment to the second player, the trustee (Berg, Dickhaut, & McCabe, 1995; Kreps, 1990; for a review, see Thielmann & Hilbig, 2015). The amount is multiplied by a certain factor (larger than 1), so that the total amount increases. However, the trustee can decide how much, if any, she wants to send back to the trustor, putting the trustor in a vulnerable position. In turn, the ultimatum game consists of a dyad where one

matter of ensuring a clear conflict between collective and self-interest (not) to cooperate in one-shot vs. repeated interactions (Rand, 2016).[19] Specifically, Rand (2016) argues that it can be beneficial to cooperate in repeated interactions as reward or punishment mechanisms can rationalize cooperation behavior, whereas in one-shot interactions the payoff maximizing choice is always to defect.

In this vein, clear operationalizations of a theory's included concepts are an important part of good theorizing (Glöckner, 2019). A theory that includes unmeasurable concepts is impossible to test. In terms of the SHH, one must acknowledge the fact that it is clear which operationalizations of the main constructs (i.e., processing mode and cooperation behavior) are considered to be valid tests of the theory. This not only refers to cooperation behavior as outlined above, but also to the intuitive vs. deliberate processing mode (even though Rand changed the operationalization of the latter over time): As described in Chapter 2.1, an intuitive processing mode was originally also captured via decision time measurement in addition to induced time constraints (see Rand et al., 2012). In his meta-analysis in 2016, Rand explicitly states that cognitive load, ego depletion, and inducing an intuitive processing mode (e.g., via instructions to decide according to one's gut feeling) are valid operationalizations allowing to test spontaneous cooperation in addition to time constraint manipulations.

Taken together, the original formulation of the SHH provides a limited level of universality, as it only makes predictions about the spontaneity of cooperation, being mute on further pro-sociality measures. Turning towards the degree of precision or level of specificity, Rand et al. (2014) explicitly made the SHH account for both a positive relation between spontaneity/intuition and cooperation as well as null effects stating that "(…) on average,

player is in the role of the proposer, the other in the role of the recipient (Güth, Schmittberger, & Schwarze, 1982). The proposer receives an endowment that she can freely distribute between herself and the receiver. However, the receiver has the option to reject the offer, so that both players receive nothing. As rejecting unfair offers is seen as a form of altruistic punishment (Fehr & Fischbacher, 2003), Rand (2016) excludes the recipients' behavior from the analysis of spontaneous cooperation.

[19] Rand (2016) explicitly excludes giving behavior in dictator games in his meta-analysis as it reflects pure altruism where the conflict between self- and other-interest does not vary between one-shot or repeated interactions (i.e., DG giving is considered neither payoff-maximizing in one-shot nor in repeated interactions).

intuition will favor cooperation" (p. 2). Strictly speaking, if a theory even allows null effects by claiming that these do not necessarily contradict the theory's predictions, one is tempted to question its falsifiability, which is required in order to be considered scientific (Popper, 1934/2005). The SHH is still falsifiable, as Rand et al. (2014) specify that "intuition should never decrease average cooperation relative to reflection in one-shot anonymous social dilemmas" (p. 2). However, it is obvious that a theory allowing both a positive relation between intuitive processing and cooperation as well as a null effect is less specific than a theory predicting either of these. Stated differently, "a theory is more precise than another one if it allows fewer different outcomes" (Glöckner & Betsch, 2011, p. 713). Indeed, Rand et al. (2014) take a first step to further specify the SHH in that they take their identified moderators (trust in the cooperativeness of daily life interaction partners, experience with the laboratory setting of social dilemmas; Rand et al., 2012) into consideration to predict "when intuition will, and will not, have an effect on cooperation" (p. 2). However, in contrast to these post-hoc integrated moderators, SVO explains on theoretical grounds why dispositional pro-sociality is a prerequisite of spontaneous cooperation (see introductory part of Chapter 3). When revising the SHH accordingly, the theory's specificity increases (see Table 2). Importantly, the theory still considers prosocials *and* proselfs in the antecedents and makes predictions for both subgroups. This implies that its level of universality is not reduced. However, this revised version of the SHH specifies the behavioral predictions (i.e., consequences) that a negative relation between decision time and cooperation is expected for prosocials, whereas no relation would be predicted by the revised SHH for proselfs (see Table 2). Stated differently, this specifies the original SHH positing that null effects as well as a negative relation between decision time and cooperation could be expected, relating the different outcomes to the different subgroups.

It goes without saying that the post-hoc identification of moderators cannot substitute a priori theorizing for whom an effect is (not) expected (Glöckner, 2019). Still, specifying a theory on an iterative basis is nevertheless important, as it enables researchers to derive future testable predictions to critically challenge the new version of the theory and consequently allows for easier falsification thereof. As an example, consider the proselfs' limited cooperation behavior irrespective of their decision time (i.e., the null effect of proselfs). It can be tested separately from the spontaneous cooperation effect for prosocials. In consequence, this enables one to better falsify the theory by specifically targeting one of those subgroups.

Taken together, identifying SVO as a moderator allows for the further specification of the SHH, for whom spontaneous cooperation is (not) expected. Turning to the second paper, this targets the applicability of spontaneous cooperation to punishment behavior and thus focuses on the level of universality. To test the increase in empirical content due to the potential generalizability, it is investigated whether spontaneous decisions go in line with increased punishment behavior (see Table 2).[20] Most notably, the second paper differentiates which factors shape a default to punish – that is, which motives underlie spontaneously expressed punishment. In line with the first paper, the first potential moderator under scrutiny was social value orientation. If it moderated spontaneous punishment, it would similarly indicate a default of prosocials not only to cooperate but also to punish norm violators. This finding would emphasize that dispositional pro-sociality is needed for both phenomena to appear. Importantly, this would allow one to not only generalize from spontaneous cooperation to spontaneous punishment – that is, spontaneous cooperation in first and second order public goods – but also provide a first indication towards a theory of spontaneous *pro-sociality*, as both acts are expressed by dispositional prosocials. This relation between punishment and further pro-sociality measures is crucial to be able to consider (second-party) punishment as a prosocial act (Brethel-Haurwitz, Stoycos, Cardinale, Huebner, & Marsh, 2016; Henrich et al., 2006; see also Chapter 4.2.3). Similarly, and not mutually exclusive, a (potential) spontaneous punishment effect could also be motivated by revenge to regulate one's negative affect that stems from

[20] As outlined in Chapter 1.1, costly punishment in social dilemmas consists of a second-order public good (Fehr & Gächter, 2002) and is also called instrumental cooperation (Yamagishi, 1986) as it has the same structure as elementary cooperation of being individually costly but beneficial for the whole group.

being exploited by one's group member(s): Fehr and Gächter (2002) showed that "[e]motions are an important proximate factor behind altruistic punishment" (p. 139). However, if solely negative affect – and not SVO – motivated spontaneous punishment, this would render spontaneous punishment to a purely retributive act and distinguish it from spontaneous cooperation: The latter is inherently independent of any retributive affect directed towards group members, given that only one-shot or the first round of social dilemmas is considered. In consequence, if spontaneous punishment consisted of a retributive act only, without being additionally moderated by SVO, this would not allow one to generalize the SHH to a theory of spontaneous pro-sociality.

Aims and hypotheses of the second article

Taken together, the second paper aimed at investigating whether the spontaneous cooperation effect could be expanded to punishment behavior in that it followed the same behavioral pattern over the course of decision time. However, it was similarly important to scrutinize whether there are different motivations underlying spontaneous behavior in second vs. first order public goods. Therefore, three hypotheses in the second article are tested – the relation between decision time and punishment behavior in general in addition to two potential factors moderating the effect – that is, whether spontaneous punishment is conditional on a prosocial value orientation (second hypothesis) and / or negative affect (third hypothesis).

Beyond testing these hypotheses, implementing a punishment stage holding all other factors of the previous studies' design constant allows one to test whether spontaneous cooperation (for prosocials) replicates with the only modification of an anticipated punishment opportunity. That is, the second paper aimed for a critical test of whether the interaction effect of spontaneous cooperation for prosocials also holds when there is the anticipated threat of being punished. As proselfs are sensitive regarding context cues that might require cooperation (Declerck, Boone, & Emonds, 2013), one might argue that proselfs similarly have a cooperative default when they fear being punished. Further, the choice of a second-party punishment setting, where each group member has the option to punish, allowed for the analysis of how punishment is related to one's own contribution behavior. In terms of how to capture the individual default, measuring decision time for cooperation behavior allowed for a direct replication of the first paper. Even more importantly, choosing to measure decision time for punishment decisions enabled a direct comparison between spontaneous cooperation and punishment during an ongoing discussion about how the processing mode can(not) be captured.

To conclude, the second paper questions whether the results of the first paper generalize to punishment, and if so, whether the underlying motives are comparable. In addition, it allows one to analyze whether the spontaneous cooperation effect (for prosocials) is replicable when there is a threat of being punished.

4.1 2nd Article: Different motives underlie spontaneous behavior in first and second order public good games

Materials, analysis scripts, and data can be found at: https://osf.io/9rpwn/

Organizational Behavior and Human Decision Processes 149 (2018) 59–72

Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

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

From spontaneous cooperation to spontaneous punishment – Distinguishing the underlying motives driving spontaneous behavior in first and second order public good games

Dorothee Mischkowski[a,b,*], Andreas Glöckner[b,c], Peter Lewisch[d]

[a] University of Göttingen, Gosslerstrasse 14, D-37073 Göttingen, Germany
[b] University of Hagen, Universitätsstrasse 27, D-58097 Hagen, Germany
[c] Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn, Germany
[d] University of Vienna, Schenkenstraße 4, A-1010 Vienna, Austria

ARTICLE INFO

Keywords:
Spontaneous cooperation
Spontaneous punishment
Decision time
Second-party punishment
Social Value Orientation
Above-average contributors
Negative affect

ABSTRACT

Recent findings indicate that at least some individuals use prosocial options by default in social dilemmas, known as 'spontaneous cooperation'. In two studies, we test whether this effect generalizes to second order public goods in the form of punishment behavior in one-shot and iterated public good games and investigate the underlying motivations. In line with spontaneous cooperation, punishment decreases with increasing decision time. Negative affect moderates this spontaneous punishment effect in one-shot public good games, in that punishment decisions are made more quickly by persons who are more upset about the contribution behavior of their group members. Unlike spontaneous cooperation, spontaneous punishment is not directly influenced by dispositional pro-sociality but by situationally above-average contributions. An overall analysis indicates a three-way interaction in that the spontaneous punishment effect is mainly valid for above-average, highly upset contributors. Hence, our results highlight the uniqueness of spontaneous punishment as being, in contrast to spontaneous cooperation, an affect-driven phenomenon of above-average contributors.

1. Introduction

Whether humans are intrinsically good or evil is a key question that is discussed in many academic disciplines such as philosophy, economics, and psychology, to name but a few. Research examining situations in which maximizing individual gains conflicts with the community's welfare (e.g., in social dilemmas) has revealed insights into the psychological mechanisms of cooperation behavior (e.g., Hardin, 1968). Recent research that went beyond a mere analysis of choices by including process measures such as decision times and eye-tracking has made a significant contribution to our knowledge concerning the underlying cognitive processes of prosocial behavior (De Dreu et al., 2010; Fiedler, Glöckner, Nicklisch, & Dickert, 2013; Rand, Greene, & Nowak, 2012; Rubinstein, 2007; Fischbacher, Hertwig, & Bruhin, 2013; Lotito et al., 2013). In an influential publication, Rand et al. (2012) suggested that cooperation behavior is more spontaneous than defection and accompanied by shorter decision times and can even be enhanced by manipulation (e.g., time pressure) to impose more intuitive processing. This indicates that the answer to the introductory question could be a positive one, in that humans might indeed be "of a good kind" and that pro-sociality can be promoted even further.

This so-called 'spontaneous cooperation effect' has inspired much subsequent research. On the one hand, many replication studies focused on the causality of the effect by manipulating intuitive processing. Several of those studies successfully replicated the effect (e.g., Rand et al., 2014; Lotz, 2015; Protzko, Ouimette, & Schooler, 2016), whereas others did not (e.g., Tinghög et al., 2013; Verkoeijen & Bouwmeester, 2014). Recently, Rand (2016) published a meta-analysis that considers different experimental manipulations of intuitive processing and shows a strong positive relationship between intuition and cooperation. Importantly, further analyses revealed that there was no statistical evidence of publication bias in the data set. Simultaneously, a many-labs, pre-registered replication project was launched to investigate the effect of time pressure / time delay as one specific manipulation of processing mode more broadly, while preventing publication biases in advance (see Open Science Framework https://osf.io/scu2f/; Bouwmeester et al., 2017). Aggregating over all 21 labs and 3596 participants, the spontaneous cooperation effect was not replicated. Still, the results

* Corresponding author at: FernUniversität in Hagen, Department of Psychology, Universitätsstr. 27, D-58097 Hagen, Germany.
E-mail address: dorothee.mischkowski@fernuni-hagen.de (D. Mischkowski).

https://doi.org/10.1016/j.obhdp.2018.07.001
Received 4 August 2016; Received in revised form 4 June 2018; Accepted 10 July 2018
0749-5978/

painted a heterogeneous picture in that a small-sized effect reappeared when subjects who did not comply with the time pressure/time delay manipulation were excluded – thus opening the discussion for potential selection effects (Bouwmeester et al., 2017; Rand, 2017).

In contrast, several studies applied the methodology of the first, correlative study from the original publication by Rand et al. (2012) by measuring rather than manipulating decision times (e.g., Nielsen, Tyran, & Wengström, 2014). In this paradigm, the negative correlation between decision time and cooperation was repeatedly replicated (e.g., Lotito et al., 2013; Cappelen, Nielsen, Tungodden, Tyran, & Wengström, 2016). Further studies, however, showed that the effect is highly volatile and influenced by multiple moderators, which relate mainly to personality characteristics such as Social Value Orientation (SVO, Mischkowski & Glöckner, 2016), Honesty-Humility (Kieslich & Hilbig, 2014; Mischkowski & Glöckner, 2016) and trust in the cooperativeness of daily-life interaction partners (Rand et al., 2012). These conditional effects indicate that spontaneous cooperation is the default only for persons with a prosocial, honest, or trusting attitude.

Researchers have questioned whether decision time measurement within the correlational paradigm investigating the spontaneous cooperation effect allows for conclusions concerning dual-process models – i.e., whether decisions are made in a more intuitive or deliberate manner (Krajbich, Bartling, Hare, & Fehr, 2015). Decision time measurement is generally evaluated as an important and valuable indicator for the identification of decision processes (e.g., Glöckner & Betsch, 2008a; Fiedler et al., 2013; Heck & Erdfelder, 2017; Spiliopoulos & Ortmann, 2018) that is inherently free of selection bias. However, it is complicated by the fact that it appears to be driven by multiple factors. Deliberation, per definition, should require more time than intuition. In addition, prior evidence in various domains has shown that decision conflict, defined as the subjective discriminability of choice options (i.e., difference in utility between choice options or, more broadly, strength-of-preference or differentiation in the phenomenological field; Cartwright & Festinger, 1943), also determines decision time (e.g., Festinger, 1943; Birnbaum & Jou, 1990). Specifically, the more similar decision options are evaluated to be due to similar aspects or cues, the greater the increase in conflict and, hence, decision time (e.g., Fiedler et al., 2013; Glöckner & Betsch, 2012). In the domain of cooperation behavior, Evans, Dillon, and Rand (2015) also show that decision times reflect the extent of decision conflict (see also Krajbich et al., 2015). Particularly, Evans et al. (2015) found that extreme (i.e., highly cooperative as well as highly defective) decisions are associated with quicker decision times than intermediate decisions, thus providing evidence for an inverse u-shaped pattern of the relation between cooperation behavior and decision time. Decision times not only increased with decision conflict but also mediated the relation between levels of conflict and decision extremity. Finally, manipulated decision time led to higher cooperation behavior but did not influence decision extremity and feelings of conflict. The authors conclude that only manipulated decision time should be used to interpret the degree of intuitiveness as measured decision time always incorporates the level of decision conflict.

Taken these findings together, short decision times might be a reflection of an individual's default manner of decision making in situations with low decision conflict, which are associated with a reduced necessity to invest cognitive resources to inhibit or alter default responses (see also Evans, 2008; Horstmann, Ahlgrimm, & Glöckner, 2009). Low decision conflict can be based on both, individual (i.e., the person generally tends to behave in that manner) as well as situational factors (e.g., all cues of the environment speak for one of the options). In line with this argument, Nishi, Christakis, Evans, O'Malley, and Rand (2016) show that the relationship between measured (endogenous) decision time and cooperation behavior in repeated games depends on both the social environment (i.e., on the cooperation level of the previous partner) and the player's own level of cooperativeness. These findings indicate that individuals' decision behavior is shaped not only by one's personality characteristics (i.e., own cooperativeness) but also by the environment.

This paper approaches the spontaneous cooperation effect from a different perspective, namely by investigating whether it generalizes to punishment decisions (see paragraph below). That is, we tested whether punishment investments involve similar properties concerning cognitive processes as investments in the first order public good (i.e., cooperation behavior). Additionally, we investigated the motivational forces accounting for this (assumed) effect. In doing so, we assessed the possibilities that spontaneous punishment could be based on a revenge-oriented, affect-driven behavior as well as that it could be dependent on traits focusing on equality and fairness, i.e., SVO, in line with spontaneous cooperation for prosocials.

1.1. Punishment behavior in social dilemmas

Punishment in social dilemmas consists of a costly sanction mechanism for norm violators (i.e., distribution norms; Fehr & Fischbacher, 2004; see also Yamagishi, 1986). It constitutes a second order public good in social dilemmas, in that it follows the same incentive structure as cooperation behavior, representing a first order public good. In first order public goods, an individual is confronted with the decision of whether to contribute (monetary) resources to a common pool that benefits the whole group to a larger extent than the individual – with no knowledge of the group members' simultaneous behavior. In contrast, punishment requests an individual's monetary investment to reduce the income of another person, thereby providing the opportunity to punish previous non-cooperative behavior and generate equality among all group members. In anonymous repeated public good situations, individuals usually show substantial cooperation behavior in the beginning that, however, deteriorates over time (Fehr & Gächter, 2002; Egas & Riedl, 2008). The option to punish was introduced as an effective instrument with which to maintain high levels of cooperation (Fehr & Gächter, 2002, Yamagishi, 1986, 1988). Generally, punishment can be conducted by persons that are part of the group (second-party punishment), third-parties who are outside observers of the group setting (altruistic punishment), or fixed punishment mechanisms (Yamagishi, 1986). In the current paper, we consider only the two former cases of punishment. In these types of punishment, after having seen the contributions of other players, participants decide whether or not to invest money to reduce the profit of other players who, for example, have been free-riding on the contributions made by other persons. Thus, punishment represents a second order public good in that "everybody in the group would be better off if free riding is deterred and high levels of cooperation are sustained, but nobody has an incentive to punish the free riders" (Fehr & Gächter, 2002, p. 137) and would preferably leave the costs of punishment to other group members.

To explain in psychological terms why people inflict considerable punishment even given the (second order) public good structure, Fehr and Gächter (2002) suggest that emotions drive altruistic punishment. In one-shot settings, the retributive character of punishing free-riders might be conveyed by anger. In repeated settings, this might be entangled with a future-oriented motivation to deter further free-riding.[1] Importantly, negative affect related to punishment decisions can be further specified. Specifically, researchers have argued that particularly

[1] Empirical results testing the differences between both punishment designs show that second-party punishment is usually conducted more often and more strongly in comparison to third-party punishment (Fehr & Fischbacher, 2004). Several articles shed light on the differences of elicited emotions dependent on whether second- or third-party punishment was conducted (e.g., Fehr & Fischbacher, 2004; Fetchenhauer & Huang, 2004). Consensus is that self-experienced injustice leads to stronger anger whereas altruistic punishment is more closely related to moral indignation (Camerer, 2003).

the highly aroused, negative emotions such as anger or outrage drive punishment (e.g., Stouten, De Cremer, & van Dijk, 2006) in contrast to relatively low activation, negative emotions such as depression. The underlying classification is based on Russel and Caroll (1999), who introduced a two-dimensional system to characterize emotions by the orthogonal dimensions arousal and valence. Following this argument, we understand punishment to be related to negative affect with a high activation – in contrast to low activation (e.g., depressed). Thus, we decided a priori to focus our investigation on the 'upset' cluster of emotions.

As affect is expressed in quick and immediate reactions, time delay manipulations have been shown to be functional in terms of "cooling off" hot emotions. As a result, they reduce rejection rates of unfair offers in Ultimatum Games (e.g., Grimm & Mengel, 2011), at least when the time delay is used for distraction in contrast to rumination (Wang et al., 2011). In terms of the extension of the (causal) spontaneous cooperation effect to punishment behavior, several experimental papers show an effect of time pressure / delay on punishment (e.g., Neo, Yu, Weber, & Gonzalez, 2013; Sutter, Kocher, & Straub, 2003). In addition, other manipulations (e.g., ego depletion) led to increased punishment behavior (e.g., Liu, He, & Dou, 2015). In contrast, a recent field experiment showed that time pressure increases helpful but not punishment behavior (Artavia-Mora, Bedi, & Rieger, 2017). Overall, the findings on punishment behavior from studies that manipulate time pressure and delay are similar to cooperative behavior. However, the mechanisms that lead to these similar effects might be different (e.g., negative affect is relevant only to punishment and not cooperation).

The emotional arousal driving punishment behavior has been shown to be elicited by differences in contribution behavior (Fehr & Gächter, 2002). Individuals who contributed larger amounts than other group members are likely to feel anger (e.g., Fehr & Gächter, 2002) and higher negative arousal about having been exploited (e.g., Fiedler, Glöckner, & Nicklisch, 2012). In line with this argument, primarily above-average contributors engage in altruistic punishment (Fehr & Gächter, 2002). From a trait perspective, cooperation behavior is linked to SVO – also referred to as social preferences – which is the stable concern for the outcomes of others (e.g., Van Lange, 2000) that also involves aspects of fairness and equality. SVO has been found to be an important predictor of cooperation behavior, implying a meta-analytical correlation of $r = .30$ (Balliet, Parks, & Joireman, 2009), or, once correcting for publication bias, an estimated $r = .25$ (Renkewitz, Fuchs, & Fiedler, 2011). More importantly for our current work, however, is research investigating the relation between SVO and punishment, which provides rather heterogeneous findings. Karagonlar and Kuhlman (2013) show that prosocials reject unfair offers in the Ultimatum Game *less* frequently, indicating a reduced tendency for punishment. In contrast, Jordan, Hoffman, Bloom, and Rand (2016) show that trustworthy persons are more likely to engage in 3rd party punishment, indicating the opposite effect. Finally, several investigations indicate no relation between pro-sociality and punishment behavior (e.g., Böckler, Tusche, & Singer, 2016; Peysakhovich, Nowak, & Rand, 2014; Yamagishi et al., 2012).

1.2. Overview of the present research

To summarize, in the current paper we aim to analyze the relation between the self-selected time for making punishment decisions on the magnitude of punishment investments as well as the underlying mechanisms. We employ a second-party punishment design in which participants can retaliate others' (e.g., low) contributions with costly punishment and measure the time they take for this decision and the magnitude of punishment.

We test the following hypotheses:

Similar to the spontaneous cooperation effect, one could expect that quick punishment is higher than slow punishment.

H1: Punishment investments decrease with increasing time for making punishment decisions.

We also refer to this prediction as the spontaneous punishment effect (main effect).

Assuming a dominant prosocial motivation to punish, that is, to restore equality in outcomes, we additionally expect an interaction of decision time with SVO, in that the assumed spontaneous punishment effect should mainly be driven by prosocials.

H2: The spontaneous punishment effect is larger for prosocials than for proselfs, as indicated by an interaction of decision time and SVO in predicting punishment.

Furthermore, in accordance with the argument above that emotions can be expected to cool off over the course of the decision time, an interaction effect with high-arousal, negative affect (i.e., anger, upset) would be expected in line with the retributive character of spontaneous punishment.

H3: The spontaneous punishment effect is larger for more upset compared to less upset individuals, as indicated by an interaction of decision time and upset (as measured by a subscale of negative affect) in predicting punishment.

2. Study 1

Study 1 tested our hypotheses in the context of an iterative public goods game, analyzing only the first round without any spillover effects from previous rounds. We did not apply a between-subject manipulation. However, decision times were measured for contribution and punishment decisions. We expected that higher punishment investments would go in line with shorter decision times, conditional on either negative affect (upset) and/or prosocial disposition.

2.1. Method

2.1.1. Participants and design

Data was collected at the Decision Lab of the Max Planck Institute for Research on Collective Goods in Bonn, Germany, and participants, mainly students of the University of Bonn, were recruited from the participant database using ORSEE (Greiner, 2004). Each session included either 8 or 12 participants. All participants were involved in the same study, without being assigned to any between-subjects manipulation. An a priori power analysis using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) revealed a required sample size of $N = 114$ to detect a small to medium-sized effect ($f^2 = .10$) in a linear multiple regression (3 predictors) with a power of $1 - \beta = .80$. In total, $N = 136$ participants completed the study. However, data from $n = 8$ was lost due to technical failure and $n = 5$ due to lack of game structure comprehension, resulting in a final sample of $N = 123$ (65 female).[2] Participants had a mean age of 22.7 years ($SD = 5.0$; $Min = 17$; $Max = 60$).[3]

2.1.2. Materials and measures

All materials and datasets are available online at https://osf.io/9rpwn/. The study consisted of a lab session that contained the public good interaction and an online part to measure SVO, which was conducted at least 12 h prior to the lab session.

Social value orientation. SVO was measured using the extended 15 items version of the SVO Slider (Murphy, Ackermann, & Handgraaf, 2011). The SVO Slider consists of a series of dictator games including nine money allocation opportunities per item for oneself and an anonymous partner. Dependent on the weight one gives to the outcome

[2] Note that results do not change when including participants demonstrating a lack of game comprehension.

[3] Note that due to the collection of demographics in the online part, we have missing values for participants who could not be matched, resulting in subsample for demographics and SVO of $N = 117$.

of the other player, a continuous SVO angle is determined, in which a higher angle is equivalent to increasing pro-sociality. Aside from measuring SVO, the extended SVO Slider Measure allows for a distinction of pro-sociality in Joint Gain Maximization and Inequality Aversion (Ackermann & Murphy, 2012) that was included for exploratory purposes following the call from Karagonlar and Kuhlman (2013) to apply this recently developed measure to rule out the differences in the broad spectrum of pro-sociality that might lead to differences in punishment behavior. Joint gain maximizers aim to maximize the sum of outcomes among individuals. Inequality averse persons accept lower gains for the sake of more equal distributions between individuals. In particular, the aspiration for joint gain maximization is likely to be incompatible with punishment behavior, since punishment minimizes the total welfare of the group. In contrast, for inequality averse persons it is perfectly effective to punish, as it reduces inequality.

Public goods game (PGG). To measure cooperation and punishment behavior, we conducted an iterative (two rounds) PGG with groups of four players, in which the same groups interacted with each other twice (partner design). Individual identifiers were maintained after the first round to allow for reputation building. Full anonymity was assured and the instructions for the game (including the punishment stage) were public knowledge prior to the start of the game. We used a modified version of the instructions from Mischkowski and Glöckner (2016), which was a translation of the materials used by Rand et al. (2012, Study 1). In the PGG, individuals anonymously contribute any amount of their endowment of 4.00 Euro (approx. USD 4.40) to a public project (in 2 cents units, thereby contributions range between 0 and 400 cents, see Table 1). The contribution to the public project are doubled and distributed equally between the four group members (Marginal per capita return = 0.50 Euro), irrespective of their individual contribution. As an important extension, we added a punishment stage after each contribution stage with a cost-to-impact ratio of 1:4. Specifically, after the contribution stage, participants were informed about the contributions of all other members of their group and could use any amount within the limits of the money they had earned thus far to reduce the payoff of any other group member (second-party punishment). Thus, the damage was four times higher than the invested punishment amount. The second-party punishment mechanism (i.e., punishment by people inside – rather than outside – the group) allowed for a direct comparison of one's own contribution and punishment investments. The partner-matching design (each group plays two rounds in a non-anonymous setting) allowed for direct deterrence and reputation building. As our core predictor, we measured decision time for punishment decisions and – for replication purposes – contribution decisions.

Decision times. In the case of contributions, decision time was measured as the time between the onset of the response screen and the confirmation of the contribution by mouse click. To exclude reading time for PGG instructions, the question of how much to contribute with the subsequent input field was placed on a separate slide. The same procedure was used for the time measurement of the punishment decision, with the exception that participants did not confirm their input via mouse click but instead used either the enter key (if no punishment was made) or the space bar (if punishment was made; see Fig. 1 for an overview of the procedure). Thus, the punishment decision was first made as a binary decision ("Do you want to punish at least one of your team members?") before – in the case of a positive decision – participants defined the amount of punishment investments on a subsequent screen. Our measure of decision time for punishment represents the time for the binary decision measured from the onset of the slide until pressing either the space bar (in order to punish) or the enter key (when deciding not to punish).

Affective states were measured using the German version of the Positive and Negative Affect Schedule (PANAS, Krohne, Egloff, Kohlmann, & Tausch, 1996), in which participants were explicitly instructed to rate their affective state with respect to the contribution behavior of the group members ("How do you currently feel about the contributions of your group members?" e.g., excited) on a five-point scale ranging from "Very little or not at all" to "Very much". As discussed in the introduction, we expect punishment to be based primarily on highly aroused negative affect. We therefore focus the analysis on the subscale "upset" of the negative affect scale, which has been shown in the German PANAS version to be empirically distinct from a second, less aroused negative affect cluster 'afraid' (Janke & Glöckner-Rist, 2014). This subscale consists of three additive items *upset*, *irritable*, and *hostile*. For exploratory purposes, we also included four items capturing

Table 1
Means, standard deviations (in parentheses) and inter-correlations in Study 1.

Variable	Scale	Mean (SD)	Correlations 1.	2.	3.	4.	5.	6.	7.	8.
1. Punishment	0,1	30.08%								
2. Punishment Investments	≥ 0, in Cents	16.71 (35.63)	$.71^{***}$ (123)							
3. Contributions	0–400 Cents	258.98 (152.55)	$.30^{***}$ (123)	$.28^{**}$ (123)						
4. Averaged distance of contributions	± 400 Cents	−0.71 (176.20)	$.46^{***}$ (115)	$.43^{***}$ (115)	$.86^{***}$ (115)					
5. SVO angle	In degree (± 180)	22.97 (13.75)	−.01 (117)	.03 (117)	$.23^{*}$ (117)	.16 (110)				
6. Inequality Aversion	0–1 (0 = IA, 1 = JGM)	.25 (.26)	$-.24^{+}$ (54)	−.16 (54)	−.01 (54)	−.1 (50)	−.02 (54)			
7. Negative Affect (Upset)	0–12	3.61 (3.64)	$.41^{***}$ (123)	$.38^{***}$ (123)	.02 (123)	$.18^{+}$ (115)	$-.20^{*}$ (117)	−.00 (54)	.92 (123)	
8. Decision Time of Punishment	log10(sec.)	.86 (.16)	$-.17^{+}$ (123)	$-.23^{*}$ (123)	−.09 (123)	$-.17^{+}$ (115)	−.02 (117)	.01 (54)	$-.24^{***}$ (123)	

SVO = Social Value Orientation, IA = inequality Aversion, JGM = Joint Gain Maximizer. Note that results only refer to the first round of the PGG. For the binary variable 'Punishment' we report point-biserial correlations. Numbers of observations are given in parentheses. Cronbach's alpha is provided in the diagonal for negative affect (upset).
Note.
$^{+}$ $p < .1$.
* $p < .05$.
** $p < .01$.
*** $p < .001$ (two-sided).

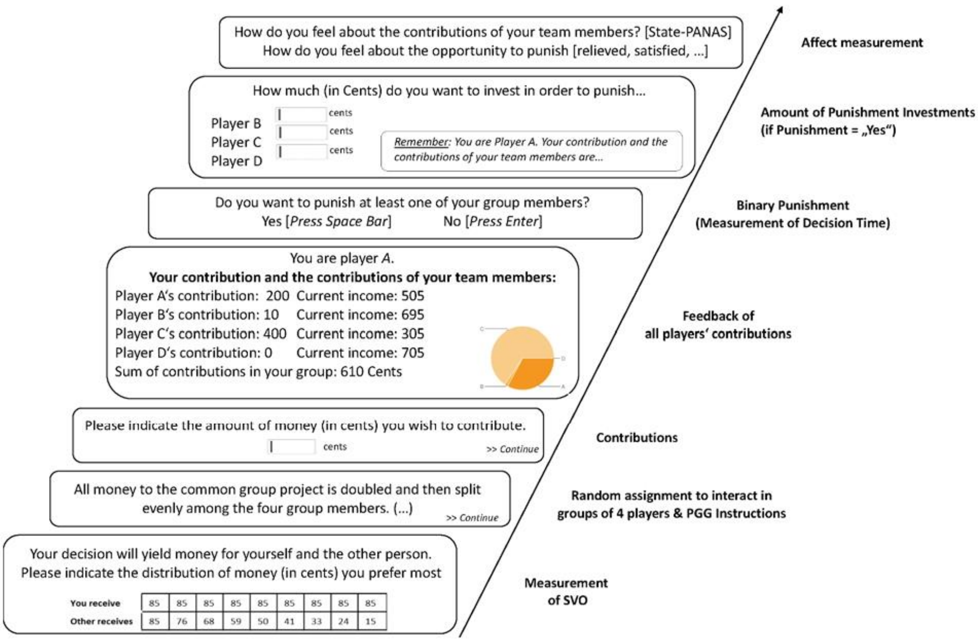

Fig. 1. Overview of both study procedures. Note that in Study 1, the Public Good Game (PGG) consisted of two rounds.

social emotions towards the team members (e.g., gratitude) as well as five items that captured the affect related to the punishment decision (e.g., relief, satisfaction). Additionally, the reasons (not) to punish were surveyed via a free-form field and eight additional items (e.g., how prosocial participants perceived their respective decision to be). In order to not interfere with the game process while investigating cooperation and subsequent punishment behavior, affect was measured only after the punishment stage of the PGG's second round. Specifically, we did not measure affect between contributions and the punishment stage to avoid effects of 'affective labeling' (Lieberman et al., 2007) and 'affective asynchrony' (Peters & Slovic, 2007). Instead, we instructed participants to explicitly consider their feelings towards the group members' contributions when filling out the PANAS.[4]

2.1.3. Procedure

Fig. 1 provides an overview of the general procedure. Participants entered the lab having already completed the SVO Slider Measure and a demographic questionnaire concerning age, gender, and education as an online pre-study up to 12 h before the lab session. In the lab, they first provided informed consent before obtaining hard copies of the PGG instructions and answering several control questions to assess game comprehension (see Rand et al., 2012).[5] Note that instructions entailed information of a subsequent punishment option, so that participants were aware of this opportunity. The PGG was conducted in BoXS (Bonn eXperimental System; Seithe, Morina, & Glöckner, 2015). Participants were matched in anonymous groups of four such that two to three groups were working in parallel. After reading the general instructions and the instructions for the contribution stage, participants made their contribution decisions on a separate screen by entering the amount in cents they want to contribute and confirming the selection by clicking on the continue button (see Fig. 1). Decision time for contributions was measured from the onset of the screen to the self-selected confirmation. Participants then estimated the average contributions of their group members (beliefs). Afterwards, feedback concerning the contributions of all the other group members was provided.

Affective states were measured after the second round of the PGG using the PANAS and nine additional items questioning the affective state in terms of social emotions and emotion regulation of punishment. Then, a free-form field to justify one's (non-) punishment decision was inserted in order to be able to also classify the underlying motivations of punishment on a qualitative level. Afterwards, the motivation and justification of the (non-) punishment decision was assessed (e.g., how prosocial participants perceived their respective decision to be). In line with Rand et al. (2012), trust in the cooperativeness of daily-life interaction partners and the previous experience with the experimental setting of social dilemma games were captured at the end of the study just before participants received feedback about their final outcome and were then paid and debriefed.

2.2. Results and discussion

In order to capture both motivations to punish, deterrence and retribution, we conducted an iterative PGG over two rounds. However, in line with typical investigations of the spontaneous cooperation effect all of our analyses, we focused only on the first round, which precludes any learning effects. We observed 45.53% of maximum contributions (M = 258.98, SD = 152.55). Approximately one third of the participants (30.08%) punished in the first round.

Participants took a median of 6.58 s to make the punishment decision in the first round (second round: median = 3.54 s), which we refer

[4] The former term, affective labeling, indicates that reporting one's affect might reduce the affective state per se, whereas affective asynchrony refers to the phenomenon that deliberation about one's affective state might reduce the relation to its spontaneously associated behavior.

[5] Control questions asked for the amount of contributions that maximize the personal vs. the group's outcome as well as the necessary punishment investment in order to withdraw 1 Euro (approximately USD 1.10) from another group member.

to as decision time for punishment. We observed a standard level of prosociality ($M = 22.97$, $SD = 13.75$) with 57.26% prosocial (SVO angle between 22.45° and 57.15°) and 42.74% individualist (SVO angle between -12.04° and 22.45°) persons. Our affect measure 'upset', i.e., the highly aroused subscale of the PANAS' negative affect (sum score of the three items 'upset', 'irritable' and 'hostile'), was reliable (Cronbach's alpha = .92) and showed considerable variation (the additive scale ranging from 0 to 12; $M = 3.61$, $SD = 3.64$). Our core dependent measure is punishment investment, which participants indicated on the second screen of the punishment phase, except if they indicated that they did not want to punish on the first screen. In this case, the punishment investment was zero. Punishment investments towards all punished group members were summed. The average of the resulting total punishment investment was $M = 16.71$ Cents ($SD = 35.63$, $Min = 0$, $Max = 200$). When considering the cases of punishment only, the average punishment was 55.54 cents ($SD = 45.68$, $Min = 1$, $Max = 200$). An overview of descriptive statistics and bivariate correlations of the main variables can be found in Table 1.[6]

In line with our hypothesis (H1), we found a main effect of spontaneous punishment investments that were lower for longer punishment decision times (OLS regression of log10-transformed decision times on punishment investments: $\beta = -.23$, $p = .01$; here and in the following, all β's refer to standardized coefficients). Participants who decided to punish took, on average, $M = 6.99$ s ($SD = 2.72$), whereas participants who decided against punishment had a decision time of $M = 8.08$ s ($SD = 3.30$).

Unexpectedly, there was no correlation between punishment investments and SVO ($r = .03$, $p = .74$). However, as one would expect on theoretical grounds, participants who are inequality averse tended to punish more frequently than joint gain maximizers (JGM; point-biserial correlation: $r = -.24$, $p = .07$). There was no support for our second hypothesis (H2), in that we did not observe an interaction between decision time to punish and SVO when predicting punishment investments ($\beta = -.09$, $p = .36$).

In line with Fehr and Gächter (2002), we analyzed on an exploratory basis the influence of situational above-average cooperation on punishment. To operationalize above-average cooperation as a continuous variable in relation to the other group members' contributions – being the only accessible standard of comparison for a participant – we averaged the differences in contributions within a PGG group. Differences between a participant and his or her three team members were calculated, with positive values indicating higher contributions than the team member and negative values being equivalent to lower contributions. We calculated the mean of these three differences per participant, resulting in a compensatory aggregate of above-average contribution, with a possible range between -4 Euro indicating maximal defection (all the other players in the group contributed 100%, that is 4 Euro, of their endowment, whereas the player contributed zero) and 4 Euro (vice versa). An average distance of zero indicates either no difference in contributions at all among players in one group or symmetrical differences in contributions from the player in question and that thus compensated each other. We find that situational above-average cooperation had a strong influence on punishment behavior, in that above-average contributors punish more harshly than participants who cooperated below-average ($r = .43$, $p < .001$, see Table 1). Furthermore, the main effect of spontaneous punishment has been shown to be robust and also holds on a marginal significance level when controlling for the averaged distance in contributions within a PGG group ($\beta = -.16$, $p = .057$), indicating a spontaneous punishment tendency even when a measure for the core factor(s) driving the likelihood of punishment is partialled out.

As for the interaction with decision time, there is a significant interaction for the continuous punishment investments ($\beta = -.20$, $p = .018$) in that above-average contributors began by punishing severely but subsequently decreased their punishment behavior over the course of the decision time. Below-average contributors continually invested few resources in punishment. Accordingly, a simple slopes-analysis (Aiken & West, 1991) revealed a significant relation between decision time and punishment investments for above-average contributors (1 *SD* above the mean: $\beta = -.39$, $t(113) = -3.10$, $p = .002$) but no relation for below-average contributors (1 *SD* below the mean: $\beta = .00$, $t(113) = .02$, $p = .99$).

Testing the influence of the negative affect's subscale 'upset' on punishment, there is a distinct correlation indicating increasing punishment investments with negative affect ($r = .38$, $p < .001$).[7] However, the moderating influence of negative affect on spontaneous punishment (H3) in this study is unclear: Regressing punishment investments on logarithmic decision time, upset as the subscale of negative affect (both variables centered on the sample mean), and their interaction did not reveal a conditional effect of negative affect on spontaneous punishment ($\beta = .02$, $p = .86$). Thus, the first study provides no indication that the spontaneous punishment effect is driven by highly upset persons. The question remains whether our affect measurement after the second round might have inflated the results for the first round. As we measured the PANAS after the second PGG round, we may have gained an affect that refers to the more recent interaction, that is, the second round, or at least consists of an affect aggregation of both rounds, implicating also a potential affect regulation of the first round.

To summarize, the first study provides evidence for a spontaneous punishment effect that, unlike the spontaneous cooperation effect, is not moderated by dispositional pro-sociality (i.e., SVO). Rather, the recent contribution behavior in the PGG appears to be more influential in that situational above-average cooperation drives the spontaneous punishment effect. Unexpectedly, though, negative affect (upset) does not moderate this effect. We ran a one-shot PGG in Study 2 to the issue associated with the affect measurement discussed above. In doing so, we are also able to distinguish a retributive motivation to punish from future-oriented deterrence, since the former is the only reasonable motivation in one-shot PGGs.

3. Study 2

Study 2 was designed to test our three hypotheses in a one-shot public goods game, as the iterative PGG in Study 1 does not allow for a differentiation in the motivations of punishment. In an iterative game, the punishment decision in the first round may entail both a retributive character to punish uncooperative behavior as well as a rational component to deter from further freeriding in the second round. Both motivations were potentially present in the first study, as they appeared

[6] To analyze whether participants indeed punished those group members who contributed comparatively less than oneself, we created three variables assessing whether the punishment investments towards each group member separately went in line with a positive difference between participants' own contributions and the punished group members' contributions. In both studies, 100 participants chose to punish at least one of their team members, thus offering in sum 300 potential punishment decisions towards the three group members. Given a positive binary punishment decision, participants punished 148 persons – thus, on average, 1.48 team members. Nine of the 148 punishment decisions (6.08%) were made against group members who contributed more than oneself (and can thus potentially be considered as antisocial punishment). The remaining 139 punishment decisions (93.92%) were directed towards group members who contributed less or the same amount. The results remain qualitatively the same when running the analyses without the subset of antisocial punishment.

[7] Note that when conducting analyses with the total negative affect scale (including the items with low arousal), the bivariate correlation between negative affect and punishment investments is somewhat lower, as expected ($r = .27$, $p = .003$). Results of the interaction, however, remain qualitatively similar.

comparably likely to drive spontaneous punishment. In order to disentangle the influence of the rational deterrence motive from the retributive motive, the second study consisted of a one-shot PGG in which the deterrence motive is excluded.

3.1. Method

3.1.1. Participants and design

Participants were again invited via the ORSEE database (Greiner, 2004) of the Max Planck Institute for Research on Collective Goods in Bonn, Germany. An a priori power analysis using G*Power (Faul et al., 2009) revealed a required sample size of $N = 151$ to detect a small to medium-sized effect ($f^2 = .10$) in a linear multiple regression with an increased number of predictors ($N = 7$) with power ($1 - \beta = .80$) sufficient to test for interaction effects. Again, no between-subject manipulation was administered. We collected data from $N = 168$, but lost data from $n = 13$ due to lack of game comprehension ($n = 7$) or technical failure ($n = 6$). We further excluded one person as an extreme outlier with regard to her punishment decision time (> 3SD above the mean), resulting in a final sample of $N = 154$ (78 females).[8] Participants' mean age is 23.29 years ($SD = 5.20$; $Min = 18$, $Max = 57$).

3.1.2. Materials and procedure

We again used the materials of Study 1 with the exception that we eliminated the second round in the PGG and added one exploratory item at the very end of the study for further clarification of how prosocial participants perceived their decision (not) to punish. The procedure was similar in that the SVO was again measured online up to 12 h before the public goods game (PGG) was played in the lab.

3.2. Results and discussion

The overall level of contributions in this study was comparable to that in the first study. We observed 46.75% of maximum contributions in the one and only round ($M = 262.12$, $SD = 156.14$). We found a slightly higher level of punishment as compared to Study 1. Approximately 40.91% of participants chose to punish at least one of their group members, resulting in a punishment investment level (including also non-punishment investments) of $M = 25.35$ cents ($SD = 47.12$, see Table 2). When considering the cases of punishment only, the average punishment rises to 61.96 cents ($SD = 56.35$, $Min = 0.5$, $Max = 250$). Participants took a median of 7.58 s to make their punishment decisions. We observed a similar level of pro-sociality as in the previous study ($M = 23.81$, $SD = 15.42$) with 55.10% prosocial (SVO angle between 22.45° and 57.15°) and 42.18% individualist (SVO angle between -12.04° and 22.45°) persons. There were also two competitors (SVO angle $< -12.04°$) and two altruists (SVO angle > 57.15°) in the sample. The affect measure 'upset' as the highly aroused subscale of the PANAS' negative affect was again highly reliable (Cronbach's alpha = .92) though lower in intensity ($M = 2.72$, $SD = 3.30$).

In line with our hypothesis (H1), we find the expected main effect of spontaneously conducted, high punishment investments ($\beta = -.19$, $p = .019$), which remains robust when controlling for the average distance between the group members' contributions ($\beta = -.19$, $p = .011$). Similar to Study 1, there was no correlation between SVO and punishment ($r = -.03$, $p = .70$), nor did SVO moderate spontaneous punishment behavior (H2: $\beta = .01$, $p = .93$). The relation between a higher likelihood to punish and Inequality Aversion could not be replicated (point-biserial correlation: $r = .08$, $p = .55$). However, situationally above-average contributions again had a strong influence on punishment behavior, both as a main effect ($r = .36$, $p < .001$, see Table 2) and in interaction with decision time ($\beta = -.23$, $p = .003$). Above-average contributors tended to spontaneously punish severely, yet their willingness to punish decreased with increasing time to make the decision to punish. In contrast, below-average contributors show a minimal, non-changing level of punishment. The simple slopes correspondingly reveal a negative relation for above-average contributors (1 *SD* above the mean: $\beta = -.38$, $t(148) = -3.96$, $p < .001$) but no relation for below-average contributors (1 *SD* below the mean: $\beta = .11$, $t(148) = .89$, $p = .38$).

In line with H3, spontaneous punishment is strongly conditional on negative affect ($\beta = -.43$, $p < .001$), in that there is a spontaneous tendency for highly upset people to punish severely that decreases with increasing decision time.[9] Less upset persons showed a limited willingness to punish – also underlined by the simple slope analyses. This indicates a significantly stronger decrease of decision time for highly upset participants (1 *SD* above the mean: $\beta = -.71$, $t(151) = -5.87$, $p < .001$), in contrast to an even reversed, but small-sized effect of increased punishment investments over the course of the decision time for less upset persons (1 *SD* below the mean: $\beta = .20$, $t(151) = 2.10$, $p = .038$).

4. Overall analysis

Since our studies were very similar, with the only distinction being an iterative vs. one-shot PGG, we also conducted an overall analysis to validate our results with a maximum of power (see Table 3).[10] As the separate analyses in both studies already suggest, there is a decrease of punishment investments with increasing time (H1, $\beta = -.20$, $p = .001$, see Fig. 2). However, unlike spontaneous cooperation, spontaneous punishment is not driven by pro-socials only (H2, $\beta = -.03$, $p = .63$).[11] Interestingly, there is no main effect for pro-sociality (SVO) on punishment investments ($r = -.01$, $p = .91$).[12] However, it is possible that SVO indirectly inserts its influence on punishment through contribution behavior. To test this indirect effect, we applied a structural equation model, in which we simultaneously estimated the total effect of SVO on punishment in addition to the indirect effect of SVO on punishment via contributions. The indirect effect is significant ($coef = .16$, $SE = .06$, $p = .01$), whereas the total effect is non-significant ($coef = -.02$, $SE = .18$, $p = .91$). Hence, we indeed find evidence that SVO does exert its influence on punishment, though indirectly through contributions. In line with Fehr and Gächter (2002), who identified situationally above-average cooperation as one source of punishment, we find that spontaneous punishment is conducted by group members who contributed more in comparison to their team members ($\beta = -.21$, $p < .001$; see Fig. 3). The simple slopes-analysis revealed a significant

[8] Note that results do not change when including the extreme outlier and participants with lack of game comprehension. Demographics were again collected online prior to the lab session. For $N = 7$ persons, data could not be matched between these two sources, so that values for their demographics and SVO are missing.

[9] The interaction also persists when including all items of the negative affect scale ($\beta = -.34$, $p < .001$).

[10] Note that we included a study dummy in all overall regression analyses to control for potential study differences.

[11] One might wonder whether (disappointed) beliefs about the pro-sociality of the group members', rather than individual SVO, might be the motivator to punish spontaneously. Indeed, beliefs are in line with an increased punishment investment ($r = .23$, $p < .001$). However, we did not find evidence for the interaction hypothesis when regressing punishment investments on the interaction term of decision time for punishment and beliefs ($\beta = -.05$, $p = .43$).

[12] We found neither a relation between Inequality Aversion – as a facet of pro-sociality – and punishment behavior ($r = -.04$, $p = .65$) nor a moderator effect of Inequality Aversion on spontaneous punishment ($\beta = -.03$, $p = .76$). Note, however, that the power of this analysis is reduced, as IA can only be meaningful interpreted for pro-socials (Ackermann & Murphy, 2012). In total, we had $N = 114$ pro-socials for testing the moderating influence of IA on spontaneous punishment.

Table 2
Means, standard deviations (in parentheses) and inter-correlations in Study 2.

Variable	Scale	Mean (SD)	Correlations 1.	2.	3.	4.	5.	6.	7.	8.
1. Punishment	0,1	40.91%								
2. Punishment Investments	≥ 0, in Cents	25.35 (47.12)	.65*** (154)							
3. Contributions	0–400 Cents	262.12 (156.14)	.28*** (154)	.26** (154)						
4. Averaged distance of contributions	± 400 Cents	1.85 (180.91)	.35*** (150)	.36*** (150)	.87*** (150)					
5. SVO angle	In degree (± 180)	23.81 (15.42)	−.12 (147)	−.03 (147)	.15^{+} (147)	.07 (143)				
6. Inequality Aversion	0–1 (0 = IA, 1 = JGM)	.30 (.32)	.08 (60)	.02 (60)	.01 (60)	.02 (59)	.30* (60)			
7. Negative Affect (Upset)	0–12	2.72 (3.30)	.20* (153)	.21* (153)	.16^{+} (153)	.29*** (149)	−.14^{+} (146)	.07 (60)	.92 (153)	
8. Decision Time of Punishment	log10(sec.)	.90 (.17)	−.02 (154)	−.19* (154)	.05 (154)	.02 (150)	.04 (147)	.23^{+} (60)	−.13 (153)	

SVO = Social Value Orientation. For the binary variable 'Punishment' we report point-biserial correlations. Numbers of observations are given in parentheses. Cronbach's alpha is provided in the diagonal for negative affect.
Note.
$^{+}$ $p < .1$.
* $p < .05$.
** $p < .01$.
*** $p < .001$ (two-sided).

Table 3
OLS regression for regression models of decision times, negative affect (upset), averaged distance of contributions on punishment investments (overall analysis).

Punishment investments (in Cents)	Model 1 Main effect	Model 2 DT * averaged distance	Model 3 DT * Upset	Model 4 DT * averaged distance * Upset
Decision Time of Punishment (DT in log10(sec))	51.58** (−3.40)	−40.42** (−2.83)	−51.98** (−3.51)	−37.94** (−2.62)
Averaged distance of contributions	–	0.09*** (6.50)	–	0.07*** (5.03)
DT * averaged distance	–	−0.30*** (−3.79)	–	−0.20* (−2.56)
Negative Affect (Upset)	–	–	1.79* (2.40)	.93 (1.26)
DT * upset	–	–	−20.50*** (−4.36)	−14.53** (−3.10)
Averaged distance * upset	–	–	–	0.00 (.60)
DT * averaged distance * upset	–	–	–	−0.05* (−2.29)
Constant	61.00*** (4.50)	14.84*** (4.14)	11.68** (3.23)	11.97** (3.41)
Observations	277	265	276	264
Adjusted R^2	.04	.22	.15	.29

Note. t-values are presented in parentheses; predictors are centered. Overall analyses also include a study dummy which is not reported.
$^{+}$ $p < .1$.
* $p < .05$.
** $p < .01$.
*** $p < .001$ (two-sided).

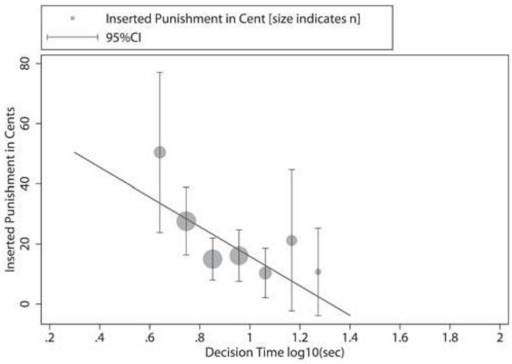

Fig. 2. Overall analysis: The spontaneous punishment effect. Prediction of punishment investments as a function of decision time. Error bars indicate 95% confidence intervals.

relation between decision time and punishment investments for above-average contributors (1 *SD* above the mean: $\beta = -.37$, $t(263) = -4.98$, $p < .001$) but no correlation for below-average contributors (1 *SD* below the mean: $\beta = .05$, $t(263) = .64$, $p = .52$).

Overall, our data suggests that spontaneous punishment decisions are accompanied by highly aroused negative affect such as upset ($\beta = -.26$, $p < .001$; see Fig. 4).[13] When comparing the simple slopes of highly upset (1 *SD* above the mean) against less upset persons (1 *SD* below the mean), a strong relation between decision time and punishment investments becomes visible for the former ($\beta = -.46$, $t(274) = -4.85$, $p < .001$) but not for the latter ($\beta = .08$, $t(274) = 1.07$, $p = .29$).

When exploring the relation between negative affect, above-average contributions, and decision time, we found a significant three-way interaction, in that spontaneous punishment is only valid for above-average, highly upset contributors ($\beta = -.14$, $p = .02$, see Fig. 5). Correspondingly, the simple slopes are only significant for the combination of highly upset and above-average contributors (both variables 1 *SD* above the mean: $\beta = -.59$, $t(260) = -5.83$, $p < .001$) in contrast

[13] Results remain similar when including all items of the negative affect scale ($\beta = -.22$, $p < .001$).

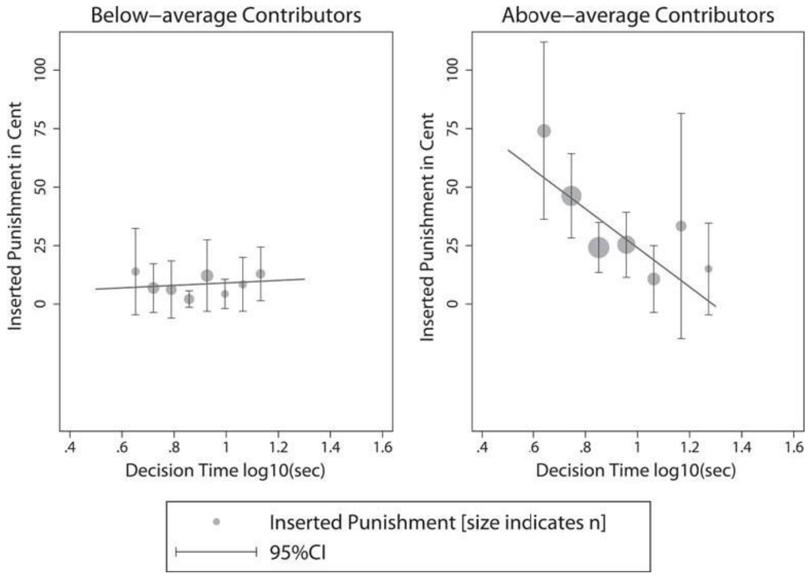

Fig. 3. Overall analysis: Spontaneous punishment for above-average contributors. Prediction of punishment investments as a function of decision time and above- vs. below-average contributions. Error bars indicate 95% confidence intervals.

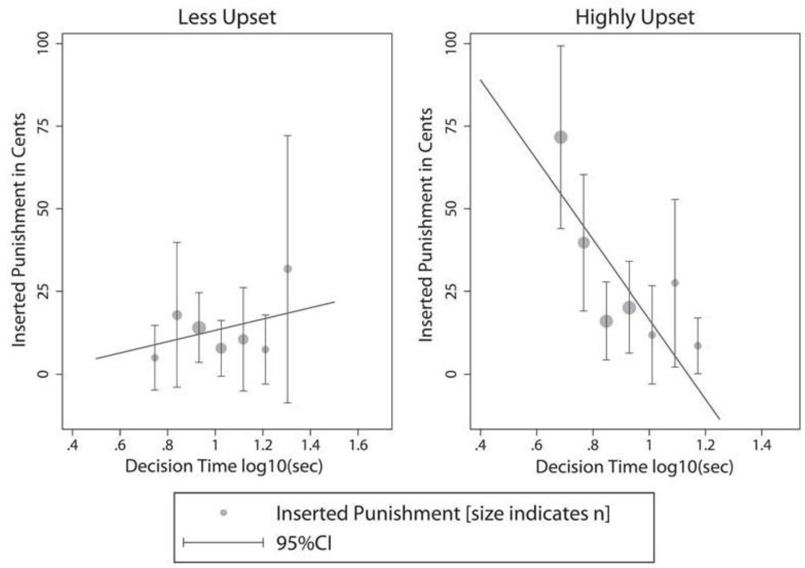

Fig. 4. Overall analysis: Spontaneous punishment for individuals that are highly upset. Prediction of punishment investments as a function of decision time and highly vs. less upset individuals. Error bars indicate 95% confidence intervals.

D. Mischkowski et al. Organizational Behavior and Human Decision Processes 149 (2018) 59–72

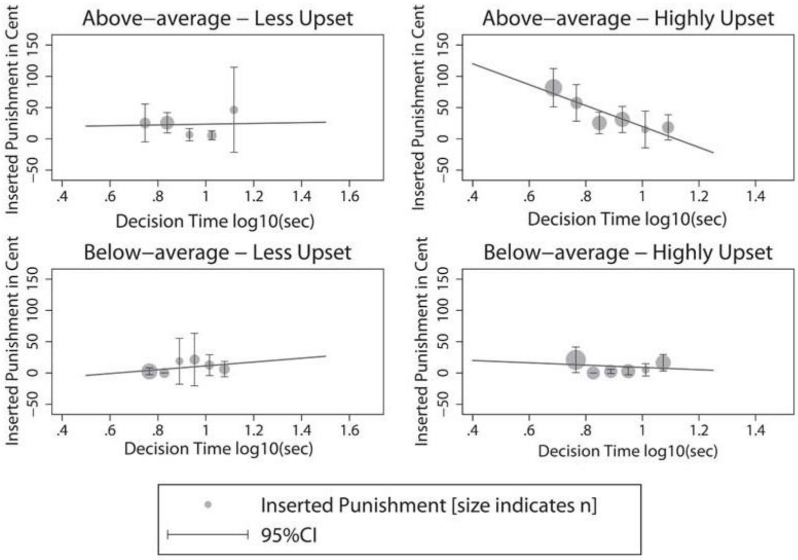

Fig. 5. Overall analysis: Spontaneous punishment for above-average contributors that are highly upset. Prediction of punishment investments as a function of decision time, above- vs. below-average contributions and highly vs. less upset individuals. Error bars indicate 95% confidence intervals.

to the other three combinations (highly upset/below-average contributors, less upset/above-average contributors, less upset/below-average contributors: all $p > .51$).[14]

4.1. Additional exploratory analyses

To further advance our understanding of the mechanisms underlying spontaneous punishment, we conducted exploratory analyses by taking into account additional available measures. Specifically, we analyzed items capturing the social emotions towards the other group members as well as the regulatory function of punishment (e.g., how relieved participants felt). Moreover, we examined the answers in a free form field in which individuals indicated their reasons (not) to punish.

Social emotions. The study included four items that measured the affective reactions towards the group members. These items reliably captured how angry and disappointed vs. grateful and appreciative participants felt regarding group members' perceived fairness (Cronbach's alpha = .84). Punishment investments increased with anger ($r = .21$, $p < .001$) and disappointment ($r = .23$; $p < .001$). Similarly, perceived fairness ($r = -.20$, $p < .001$) and gratitude towards one's group ($r = -.14$, $p = .02$) correlated negatively with punishment investments. In line with the results from the high-arousal, negative affect scale of the PANAS, negative emotions of anger ($r = -.17$, $p = .01$) and disappointment ($r = -.19$, $p = .001$) moderately correlated with a reduced decision time for punishment. More importantly, however, anger and disappointment moderated the spontaneous punishment effect (interaction of decision time x disappointment: $\beta = -.20$, $p = .001$; interaction of decision time x anger: $\beta = -.24$, $p < .001$). In contrast, the positive social emotions fairness and gratitude neither correlate nor interact with decision time for punishment (all $p > .11$).

Emotion regulation through punishment. The five items concerning the emotion regulation function of punishment (e.g., how relieved and satisfied one felt after punishing) reliably measured the emotional regulation construct (Cronbach's alpha = .90). In addition, the overall score correlated strongly with punishment investments ($r = .40$, $p < .001$). This indicates that punishment indeed fulfilled its function of emotion regulation. It is important to mention, though, that all affective responses were measured comprehensively after the punishment decision. Thus, results might be biased due to the fact that individuals aimed to use them as an opportunity to justify their decision and appear consistent in their behavior.

Self-reported reasons (not) to punish. The responses in the free form field in which participants were asked to list their reasons (not) to punish were qualitatively analyzed, in that they were categorized in different classes. In total, participants stated 168 reasons not to punish, which we summarized into seven qualitative categories. Most frequently, rational cost-benefit analyses were mentioned, in that participants might have liked to punish but considered it to be too costly (41.07%). In a similar vein, the contributions of the group members were perceived to be sufficiently high, thus providing no need for punishment (20.24%). Moreover, participants refused the destruction of resources on both ends (13.10%) or even referred to a general (deontological) reluctance to punish (2.98%). Furthermore, the integrity argument was mentioned, indicating that one had no right to punish when one has contributed equivalently or even fewer resources (8.33%). A few notable comments referred on a philosophical level to the free will of humans – that is, the freedom to decide whether one will cooperate or not (8.33%) – as well as to the categorical imperative (1.79%). A few unclassifiable arguments remained (4.17%).

In terms of what justified an exerted punishment, 109 justifications were classified into five categories. Fairness aspects and social norms were most commonly mentioned (68.81%). In addition, participants mentioned educational and deterrence motives (17.43%). In line with

[14] When considering the decision process throughout the game procedure, one might think of a serial mediation in which negative affect is not an independent source but, rather, is elicited by the differences in contributions, which in turn leads to shorter response times for punishment decisions. We explicitly tested this path in a structural equation model. However, results reveal an unsatisfactory model fit on absolute terms (root-mean-square error of approximation (RMSEA) > .1), which is why we refrain from promoting affect as a mediator.

the assumption that inequality averse persons are more likely to punish, several participants reported that their punishment decisions were motivated by an aim to reduce inequality (4.59%). Some participants also frankly admitted that their decision was affect-driven, in that they were angry about their team members' contributions (3.67%). In line with the very few antisocial punishment decisions, several participants explicitly sought to weaken their team members so as to be the winner of the group with the highest outcome in the end (1.83%). Again, a small number of arguments could not be classified (3.67%).[15]

4.2. Replication results of spontaneous cooperation for pro-socials

Our design allows for a direct replication analysis of the spontaneous cooperation effect for pro-socials of Mischkowski and Glöckner (2016), with the only design modification being an anticipated punishment opportunity afterwards. Overall, we replicate the interaction of SVO and decision time on cooperation behavior ($\beta = -.15$, $p = .013$) and also find a main effect of spontaneous cooperation ($r = -.12$, $p = .041$). Generally, in our studies contributions are related to dispositional SVO, thus replicating previous findings ($r = .19$, $p = .002$), although the correlations are lower than previously reported. The interaction of decision time and SVO follows the presented pattern in that spontaneous cooperation behavior is only valid for pro-socials but not for pro-selfs. Independent of decision time, pro-selfs' cooperation behavior is comparably low.

Using the extended SVO Slider Measure (Ackermann & Murphy, 2012), we were able to further differentiate between Joint Gain Maximizers (JGM) and inequality averse (IA) persons and found an overall significant interaction of decision time and JGM (vs. IA) on cooperation behavior ($\beta = -.21$, $p = .028$). With increasing JGM, the decrease of cooperation behavior over time is even stronger. These results need further exploration without anticipated punishment but may provide a starting point to further investigate differences in spontaneous cooperation among pro-socials.

Trust and experience. Rand et al. (2012) identified trust in the cooperativeness of daily life interaction partners and previous experience with the experimental setting of social dilemma games as two moderators of the spontaneous cooperation effect. In doing so, they demonstrated that only naïve subjects and those with a trusting attitude, respectively, cooperate spontaneously. In contrast to previous findings (Rand et al., 2012), we did not replicate these interaction effects in that neither cooperativeness of daily life interaction partners ($\beta = -.02$, $p = .70$) nor experience with the experimental setting of social dilemma games ($\beta = .05$, $p = .44$) moderated spontaneous cooperation. Interestingly, however, we found a general tendency that experience with the experimental setting of social dilemma games is in line with slightly higher punishment investments ($r = .16$, $p = .01$), which contrasts the assumption of increasingly rational (i.e., individually payoff maximizing) behavior with experience.

5. General discussion

The spontaneous cooperation effect (Rand et al., 2012) has inspired much subsequent work – discussions concerning its existence and moderators are still ongoing (Rand, 2016; Bouwmeester et al., 2017; Rand, 2017). In two studies, we show that the often observed finding of a negative correlation between measured decision time and contributions generalizes to second order public goods – i.e., punishment decisions. Hence, we demonstrate a spontaneous punishment effect in that quicker decisions regarding whether to punish or not are accompanied by higher punishment investments, thus supporting our first hypothesis.

There are (at least) two distinct but not mutually exclusive reasons that might contribute to explaining this effect, which we further aimed to investigate in our studies. First, punishment decisions constitute a second order public good that could arguably be based on the same processes as first order public goods underlying the investigations of spontaneous cooperation. If this is the case, one would expect spontaneous punishment to be mainly shown by pro-socials, as is the case for spontaneous cooperation. According to this argument, the spontaneously activated default option for pro-socials is to contribute, which could be generalized to the contribution of one's own resources to reduce others' income when distribution norms are violated. We, however, consistently observed no moderating influence of social preferences on spontaneous punishment (i.e., no interaction between decision time and SVO or inequality aversion, respectively), yielding no evidence for our second hypothesis. However, we did find that SVO has an indirect influence on punishment via contributions, underlining the notion that the moderating effect of above-average cooperation on spontaneous punishment is, at least to some extent, attributable to social preferences.

Second, punishment is arguably related to highly aroused, negative emotions, which are likely to cool down over time and elicit behavior in full intensity only spontaneously and less so if tamed by (more) rational deliberate processes. If this emotion and retaliation-based reason is the main driver of spontaneous punishment, one would expect spontaneous punishment to be moderated by highly activated negative emotions such as upset or anger. In line with this prediction and our third hypothesis, we observed that both upset and anger strongly moderate the spontaneous punishment effect (i.e., an interaction between decision time and upset / anger). Hence, the mechanisms underlying the spontaneous punishment effect appear to be more associated with negative, highly aroused emotions and affect.

5.1. Related findings

Our findings are in line with an intuitive striving for retaliation that has recently been shown in vignette studies (Gollwitzer, Braun, Funk, & Süßenbach, 2016). Furthermore, interesting effects have been shown for incidental (i.e., manipulated) emotions in that they require attentional resources leading to increased third-party punishment (Gummerum, Van Dillen, Van Dijk, & López-Pérez, 2016).

In contrast to our findings, Bieleke, Gollwitzer, Oettingen, and Fischbacher (2017) identify SVO as a moderator of intuitive punishment: They show that prosocials intuitively reject unfair offers in the Ultimatum Game but were more likely to accept such offers in a reflective condition. Similarly, Pfattheicher, Keller, and Knezevic (2017) find an intuitive striving for antisocial punishment conditional on sadistic tendencies. In addition to the difference that we relied on decision time measurement to capture one's individual default rather than manipulating the processing mode, we see a possible reason for our null finding in the second-party setting of the PGG that implies – unlike the Ultimatum Game – own contributions before punishing. These contributions, made immediately before the punishment decision, are most decisive in terms of whether or not to punish. One must not ignore, however, the fact that SVO exerts its influence in this regard, despite the fact that it does so indirectly by influencing contribution behavior.

Our results, however, are in line with the conclusion of Van Dijk, Wit, Wilke, and De Kwaadsteniet (2010), who point out that clear fairness criteria in the division of resources undermine the influence of SVO on individuals' fairness judgments. Furthermore, SVO has been shown to be highly related to Honesty-Humility (HH) as a basic trait in the HEXACO personality model (Lee & Ashton, 2004) that refers to striving to be "fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them without suffering retaliation" (Ashton & Lee, 2007, p. 156). Similar to our

[15] Following the free form field, we measured the relevance of specific motives (e.g., deterrence) and indicators for punishment (e.g., focusing on one's own income vs. on an equal distribution of resources for the whole group) with eight items. The results revealed no additional insights and are thus not reported in detail.

findings regarding SVO and punishment behavior, HH does not predict punishment behavior (Hilbig, Zettler, Leyb, Heydasch, 2013). Further research might examine the interplay of spontaneous punishment and the basic traits of personality.

5.2. Limitations and future directions

One must generally contest a causal interpretation of the spontaneous punishment effect implying that shorter decision times lead to stronger punishment. This would depict an oversimplified, mono-causal misinterpretation that must be avoided, particularly when taking into account the intensely discussed influence of decision conflict on decision time (Evans, Dillon, & Rand, 2015; Krajbich, Bartling, Hare, & Fehr, 2015). Indeed, it seems plausible to assume that individuals who were greatly exploited by (one of) their group members – i.e., above-average contributors – are likely to perceive low decision conflict and thus punish both spontaneously and harshly. Hence, the observed correlation between decision time and punishment might be driven by a third factor such as how clear the signals are to punish (vs. not punish) for the respective person. These signals represent the differentiation in the phenomenological field (Cartwright & Festinger, 1943); that is the degree of decision conflict according to our definition in the introduction.

Relatedly, conclusions concerning a causal role of negative effect driving punishment should be avoided for a similar reason. Both negative affect and punishment could be driven by how severe the norm violation (i.e., the lower contribution behavior of the other group members) is perceived to be, denoting again the (lack of) decision conflict. Overall, we thus explicitly acknowledge decision conflict (and differentiation between options as its complement) as a variable entangled with decision time. Likewise, again, we do not claim any causality regarding the effect such as that shorter decision times would elicit higher punishment decisions.

Rather, we argue that in situations with low decision conflict (e.g., clear and aligned signals whether or not to punish from the environment and the persons general preference) individuals intuitively apply their defaults, which are reflected in shorter decision times and entail intuitive processing. These findings can also be related to the debate between various kinds of dual process models mentioned in the introduction. Thus, decision conflict between various pieces of information appears to be instantly monitored, as assumed and quantitatively described in process models for intuition based on coherence construction (e.g., Glöckner & Betsch, 2008b; Glöckner, Hilbig & Jekel, 2014; Holyoak & Simon, 1999). Hence, our article strengthens the decision conflict perspective on dual process models by showing that the conflict between situational factors and personality characteristics (i.e., social preferences) appear to influence the tendency (not) to punish and (not) follow an individuals' intuitive default.

One question that remains is what motivates on an individual level longer deliberation even though one is upset, given we have no manipulation of time delay. Answering this question is another task for future research, which might focus on manipulating properties and the structure of the game. Assuming that using a punishment opportunity might be considered by a person, conflicting motives (e.g., aiming, on the one hand, to fulfil potential retributive motives and reduce inequality vs. trying to avoid personal costs) will generate a decision conflict. Manipulating the leverage (i.e., the cost to impact ratio of punishment), the number of repetitions of the game played with the same person or even the entire game structure (e.g., using a different kind of game) would allow for testing this alternative hypothesis in more detail.

Our finding that social preferences do not moderate spontaneous punishment in PGG with second-party (or peer) punishment stands in contrast with findings in settings of altruistic (or third-party) punishment, in which an observer that did not contribute can punish in a costly manner. We would argue that situations of altruistic punishment offer a more suitable setting for social preferences to exert direct influence. First, this might be due to the fact that in altruistic punishment decisions there is no diffusion of responsibility but rather one person in the distinguished position to establish fairness without being part of the interacting group. This person knows about his responsibility to watch the overall group's welfare; a perspective that is more likely influenced by general fairness concerns in comparison to an involved individual who is trying to end up best. In this context, moral values might be more strongly activated in third-party punishment situations, which might increase the relation with inequality aversion. Previous research in second-party settings, such as rejection behavior in the Ultimatum Game, has shown that punishment decisions are driven more strongly by impulsive, self-centered emotions such as greed, vengeance, and spite (e.g., Crockett, Clark, Lieberman, et al., 2010; Pillutla & Murnighan, 1996). Still, in third-party punishment, as well, (negative) affect and above-average contributions are important factors that drive punishment (Fehr & Fischbacher, 2004; Fehr et al., 2002; Fehr & Gächter, 2002). We replicate these results for second-party punishment decisions and – even more importantly – show their influence over the course of decision time.

6. Conclusion

Similar to a spontaneous cooperation effect, we find that the invested resources to punish decrease with increasing decision time. Even though the mere behavior follows a pattern similar to that of spontaneous cooperation, the different moderators underlying both effects indicate the different motivations. Spontaneous cooperation appears to be driven by dispositional pro-sociality, that is, SVO, Honesty-Humility, and trust in the cooperativeness of daily-life interaction partners. Spontaneous punishment, in contrast, appears instead to be driven by the recent group interaction and to be used in an act of revenge in which situational high contributors retaliate (subjectively perceived) insufficient cooperation behavior.

7. Additional information

Data Availability: Data and materials are also available online at Open Science Framework (https://osf.io/9rpwn/).

8. Conflict of interest

The authors declare no competing conflict of interest.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors thank Neil Feigenson, Isabel Thielmann, and Mario Gollwitzer for their helpful and constructive comments on previous drafts of this paper.

References

Ackermann, K. A., & Murphy, R. O. (2012). Tutorial on how to evaluate the SVO Slider Measure's secondary items. Retrieved from < http://vlab.ethz.ch/svo/SVO_Slider/SVO_Slider_scoring.html > .

Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.

Artavia-Mora, L., Bedi, A. S., & Rieger, M. (2017). Intuitive help and punishment in the field. *European Economic Review, 92*, 133–145.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150–166.

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes and Intergroup Relaxations, 12*(4), 533–547.

Bieleke, M., Gollwitzer, P. M., Oettingen, G., & Fischbacher, U. (2017). Social value orientation moderates the effects of intuition versus reflection on responses to unfair ultimatum offers. *Journal of Behavioral Decision Making, 30*, 569–581.

Birnbaum, M. H., & Jou, J.-W. (1990). A theory of comparative response times and "difference" judgments. *Cognitive Psychology, 22*(2), 184–210.

Böckler, A., Tusche, A., & Singer, T. (2016). The structure of human pro-sociality: Differentiating altruistically motivated, norm motivated, strategically motivated, and self-reported prosocial behavior. *Social Psychological and Personality Science, 7*(6), 530–541.

Bouwmeester, S., Verkoeijen, P. P. J. L., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., ... Wollbrant, W. E. (2017). Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science, 12*(3), 527–542.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction.* Princeton University Press.

Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J. R., & Wengström, E. (2016). Fairness is intuitive. *Experimental Economics, 19*(4), 727–740.

Cartwright, D., & Festinger, L. (1943). A quantitative theory of decision. *Psychological Review, 50*(6), 595–621.

Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G., & Robbins, T. W. (2010). Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. *Emotion, 10*(6), 855–862.

De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., ... Feith, S. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science, 328*(5984), 1408–1411.

Egas, M., & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society of London B: Biological Sciences, 275*(1637), 871–878. http://rspb.royalsocietypublishing.org/content/275/1637/871.short.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278.

Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General, 144*(5), 951–966.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63–87.

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

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

Festinger, L. (1943). Studies in decision. II. An empirical test of a quantitative theory of decision. *Journal of Experimental Psychology, 32*(5), 411–423.

Fetchenhauer, D., & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences, 36*(5), 1015–1029.

Fiedler, S., Glöckner, A., & Nicklisch, A. (2012). The influence of social value orientation on information processing in repeated voluntary contribution mechanism games: An eye-tracking analysis. In A. Innocenti, & A. Sirigu (Vol. Eds.), *Neuroscience and the Economics of Decision Making: Vol. 5*, (pp. 21–53). New York: Routledge.

Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social value orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes, 120*(2), 272–284.

Fischbacher, U., Hertwig, R., & Bruhin, A. (2013). How to model heterogeneity in costly punishment: Insights from responders' response times. *Journal of Behavioral Decision Making, 26*(5), 462–476.

Glöckner, A., & Betsch, T. (2008b). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making, 3*(3), 215–228.

Glöckner, A., & Betsch, T. (2008a). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1055–1075.

Glöckner, A., & Betsch, T. (2012). Decisions beyond boundaries: When more information is processed faster than less. *Acta Psychologica, 139*, 532–542.

Glöckner, A., Hilbig, B. E., & Jekel, M. (2014). What is adaptive about adaptive decision making? A parallel constraint satisfaction account. *Cognition, 133*, 641–666.

Gollwitzer, M., Braun, J., Funk, F., & Süssenbach, P. (2016). People as intuitive retaliators spontaneous and deliberate reactions to observed retaliation. *Social Psychological and Personality Science, 7*(6), 521–529.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer, & V. Macho (Eds.). *Forschung und wissenschaftliches Rechnen 2003* (pp. 79–93). Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Grimm, V., & Mengel, F. (2011). Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters, 111*(2), 113–115.

Gummerum, M., Van Dillen, L. F., Van Dijk, E., & López-Pérez, B. (2016). Costly third-party interventions: The role of incidental anger and attention focus in punishment of the perpetrator and compensation of the victim. *Journal of Experimental Social Psychology, 65*, 94–104.

Hardin, G. (1968). The tragedy of the commons. *Science, 162*(3859), 1243–1248.

Heck, D. W., & Erdfelder, E. (2017). Linking process and measurement models of recognition-based decisions. *Psychological Review, 124*(4), 442–471.

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty-Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences, 54*(5), 598–603.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General, 128*(1), 3–31.

Horstmann, N., Ahlgrimm, A., & Glöckner, A. (2009). How distinct are intuition and deliberation? An eye-tracking analysis of instruction-induced decision modes. *Judgment and Decision Making, 4*(5), 335–354.

Janke, S., & Glöckner-Rist, A. (2014). Deutsche Version der Positive and Negative Affect Schedule (PANAS). In D. Danner & A. Glöckner-Rist (Eds.), Zusammenstellung sozialwissenschaftlicher Items und Skalen.

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.

Karagonlar, G., & Kuhlman, D. M. (2013). The role of social value orientation in response to an unfair offer in the ultimatum game. *Organizational Behavior and Human Decision Processes, 120*, 228–239.

Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making, 9*(6), 510–522.

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications, 6*(7455).

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS). *Diagnostica, 42*(2), 139–156.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*(2), 329–358.

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science, 18*(5), 421–428.

Liu, Y., He, N., & Dou, K. (2015). Ego-depletion promotes altruistic punishment. *Open Journal of Social Sciences, 3*(11), 62–69.

Lotito, G., Migheli, M., & Ortona, G. (2013). Is cooperation instinctive? Evidence from the response times in a public goods game. *Journal of Bioeconomics, 15*(2), 123–133.

Lotz, S. (2015). Spontaneous giving under structural inequality: Intuition promotes cooperation in asymmetric social dilemmas. *PLoS ONE, 10*(7), e0131562.

Mischkowski, D., & Glöckner, A. (2016). Spontaneous cooperation for prosocials, but not for proselfs: Social value orientation moderates spontaneous cooperation behavior. *Scientific Reports, 6*(21555), 1–5.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making, 6*(8), 771–781.

Neo, W. S., Yu, M., Weber, R. A., & Gonzalez, C. (2013). The effects of time delay in reciprocity games. *Journal of Economic Psychology, 34*, 20–35.

Nielsen, U. H., Tyran, J.-R., & Wengström, E. (2014). Second thoughts on free riding. *Economics Letters, 122*(2), 136–139.

Nishi, A., Christakis, N. A., Evans, A. M., O'Malley, A. J., & Rand, D. G. (2016). Social environment shapes the speed of cooperation. *Scientific Reports, 6*(29622), 1–10.

Peters, E., & Slovic, P. (2007). Affective asynchrony and the measurement of the affective attitude component. *Cognition and Emotion, 21*(2), 300–329.

Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications, 5*(5939), 1–8.

Pfattheicher, S., Keller, J., & Knezevic, G. (2017). Sadism, the intuitive system, and antisocial punishment in the public goods game. *Personality and Social Psychology Bulletin, 43*(3), 337–346.

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes, 68*, 208–224.

Protzko, J., Ouimette, B., & Schooler, J. (2016). Believing there is no free will corrupts intuitive cooperation. *Cognition, 151*, 6–9.

Rand, D. G. (2016). Cooperation, fast and slow. *Psychological Science, 27*(9), 1192–1206.

Rand, D. G. (2017). Reflections on the time-pressure cooperation registered replication report. *Perspectives on Psychological Science, 12*(3), 543–547.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*, 427–430.

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*, 1–12.

Renkewitz, F., Fuchs, H. M., & Fiedler, S. (2011). Is there evidence of publication biases in JDM research? *Judgment and Decision Making, 6*(8), 870–888.

Rubinstein, A. (2007). Instinctive and cognitive reasoning: A study of response times. *The Economic Journal, 117*(523), 1243–1259.

Russel, J. A., & Caroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin, 125*(1), 3–30.

Seithe, M., Morina, J., & Glöckner, A. (2015). Bonn eXperimental System (BoXS): An open-source platform for interactive experiments in psychology and economics. *Behavior Research Methods*, 1–22.

Spiliopoulos, L., & Ortmann, A. (2018). The BCD of response time analysis in experimental economics. *Experimental Economics, 21*(2), 383–433.

Stouten, J., De Cremer, D., & van Dijk, E. (2006). Violating equality in social dilemmas: Emotional and retributive reactions as a function of trust, attribution, and honesty. *Personality and Social Psychology Bulletin, 32*(7), 894–906.

Sutter, M., Kocher, M., & Strauß, S. (2003). Bargaining under time pressure in an experimental ultimatum game. *Economics Letters, 81*(3), 341–347.

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., ... Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature, 498*, 427–430.

Van Dijk, E., Wit, A. P., Wilke, H. A. M., & De Kwaadsteniet, E. W. (2010). On the importance of equality in social dilemmas. In R. M. Kramer, A. E. Tenbrunsel, & M. H.

Bazerman (Eds.). *Social decision making: Social dilemmas, social values, and ethical judgments* (pp. 44–69). New York: Routledge.
Van Lange, P. A. (2000). Beyond self-interest: A set of propositions relevant to interpersonal orientations. *European Review of Social Psychology, 11*(1), 297–331.
Verkoeijen, P., & Bouwmeester, S. (2014). Does intuition cause cooperation. *PloS One, 9*(5), e96654.
Wang, C. S., Sivanathan, N., Narayanan, J., Ganegoda, D. B., Bauer, M., Bodenhausen, G. V., & Murnighan, K. (2011). Retribution and emotional regulation: The effects of time delay in angry economic interactions. *Organizational Behavior and Human Decision Processes, 116*(1), 46–54.
Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*(1), 110.
Yamagishi, T. (1988). Seriousness of social dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly,* 32–42.
Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., ... Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences, 109*(50), 20364–20368.

4.2 Discussion

Similar to spontaneous cooperation, the invested resources to punish also decrease over the course of decision time, thereby providing evidence for a spontaneous punishment effect. In contrast to spontaneous cooperation, however, spontaneous punishment does not represent the default of prosocials and instead is only valid for highly upset, above-average contributors. Still, the spontaneous cooperation effect for prosocials replicates: Prosocials and proselfs differ in their spontaneously expressed cooperativeness, even when there is the threat of being punished. Discussing these results of this thesis' second paper firstly deals with its limitations and tasks for future research. Specifically, the point in time of affect measurement as well as the already debated entanglement of intuition and decision conflict when measuring decision time are taken into account. Finally, the question is discussed whether (second-party) punishment can be interpreted as a prosocial act at all that would allow proposing a more general theory of spontaneous *pro-sociality*. In this vein, related pro-sociality measures are discussed concerning their relation with decision time and their dependence on the processing mode to shed further light on the SHH's level of universality.

4.2.1 Affect measurement

Measuring affect via rating scales after individuals have made their decision is a common procedure (e.g., Bosman & Van Winden, 2002; Cubitt, Drouvelis, & Gächter, 2011). However, a non-obtrusive acquisition of affect in the moment of the punishment decision (or starting even earlier at the evaluation phase of the other players' contributions) would be preferable in future research. Particularly the inclusion of eye-tracking (Partala & Surakka, 2003) or measurement of skin conductance could provide more objective, on-time information about individuals' physiological arousal (see also S. Fiedler, Glöckner, & Nicklisch, 2012), still avoiding affective asynchrony and affective labeling – the two main reasons why affect was assessed after the punishment decision.[21] This would specifically allow for investigating whether the interaction of spontaneous punishment for highly upset individuals in the first study would appear when affect is measured right at the time of decision making. To briefly reiterate, in the first study of Mischkowski, Glöckner, et al. (2018) the public goods game consisted of two rounds and affect was measured comprehensively afterwards. The negative affect of the first round relevant for the analyses was thus confounded as participants reported their negative affect across both interactions. In addition, it cannot be ruled out that additional emotional factors (e.g., the fear of counter-punishment in the subsequent round, see Denant-Boemont, Masclet, & Noussair, 2007; Nikiforakis, 2008; Nikiforakis & Engelmann, 2011) divert the outrage in the first round resulting from defective behavior of other group members. The second study, containing only a one-shot public goods game, could rule out both limitations. Still, future research might benefit from non-obtrusive affect measures.

4.2.2 Differentiating (again) between decision conflict and intuitive processing

The correlative design of the studies reported in Mischkowski, Glöckner, et al. (2018) allows once more to analyze the replicability, respectively robustness of the decision conflict hypothesis, attributing short decision times to low decision conflict rather than intuitive processing. Specifically, it regards decision time as an indicator of an individual’s strength of preference (that is equivalent to low decision conflict) and predicts that not only highly cooperative but also defective choices in the public goods game are made faster than intermediately cooperative decisions (A. M. Evans et al., 2015). In addition to the analysis of elementary cooperation behavior, the second paper also allows for the exploratory analysis of a potentially inverse u-shaped pattern for punishment investments when predicting decision times. Sticking to the procedure of the additional analyses of the first paper, I will again follow the approach of A. M. Evans et al. (2015), analyzing the relation of decision time and

[21] Affective asynchrony refers to reduced affect with increasing reflection about it (Peters & Slovic, 2007); affective labeling represents the phenomenon to reduce the affective state by expressing it (Lieberman et al., 2007).

cooperation extremes in addition to explicitly testing for a non-linear (i.e., inverse u-shaped) relation between cooperation behavior and decision time by adding a squared term of contributions in the regression model. In a second step, the same analyses are conducted for punishment investments, investigating whether they similarly follow an inverse u-shaped pattern in their relation to decision time.

Differentiation for cooperation behavior

To begin with the analysis of cooperation behavior, descriptive statistics are comprehensively provided in Table 3. Similar as in the additional analyses of this thesis' first paper, the results by A. M. Evans et al. (2015) were again replicated in that decision extremes (i.e., highly cooperative and defective choices) are faster in comparison to intermediate contributions (i.e., a negative relation between decision times and deviation in contributions from the scale mean; β = -.28, $p < .001$). Note that again, the negative relation between decision time and contributions remains significant (i.e., the spontaneous cooperation effect; β = -.26, $p < .001$).[22]

Table 3. Descriptive statistics for the relevant variables of the additional analyses

	Mean (*SD*)	Min	Max
Contributions (in %)	65.18 (38.57)	0	100
Decision extremes	34.74 (16.64)	2.32	65.18
Decision Time (in sec.)	17.39 (12.44)	5.51	113.89
Decision Time (log 10 sec.)	1.17 (0.24)	0.74	2.06

Note. Decision extremes consist of the absolute difference of contributions to the scale mean.

[22] Note that a study dummy was again included in this and all subsequent analyses to control for study differences. In line with A. M. Evans et al. (2015) decision extremes and contributions were mean-centered.

Secondly, there is once more a significant quadratic term for contributions when regressing response time on mean centered contributions and their squared term (β = -.35, p < .001) across both studies. The linear effect remains significant (β = -.35, p < .001), showing that even when controlling for the curvilinear trend, the linear relation still holds. As the first necessary condition for an inverse u-shaped pattern is fulfilled by the negative and significant squared term, I followed the approach of Lind and Mehlum (2010) and additionally examined whether the extremum point is in the value range of contributions. Indeed, the extremum is located at 48.49% of contributions, accompanied by a significant positive slope to its left-hand side (b = .006, p < .001) and a negative significant slope to its right-hand side of comparable magnitude (b = -.007, p < .001), unequivocally supporting the inverted u-shaped pattern (see Figure 6).[23]

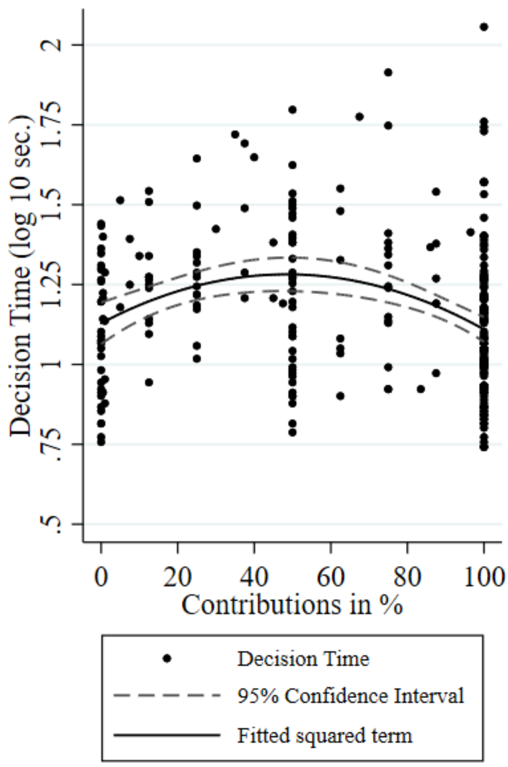

Figure 6. Prediction of decision time as a function of contributions and their squared term. Grey-dashed lines denote the 95% confidence interval.

[23] For illustrative purposes, the graph contains non-centered contributions, the reported extremum correcting for the linear shift of mean-centered contributions and their squared term used in the regression model.

Summarizing the results of the additional analyses on the spontaneity of cooperation behavior of both articles (Mischkowski & Glöckner, 2016; Mischkowski, Glöckner, et al., 2018), there is consistent evidence that decision extremes of cooperation behavior are made faster than intermediately cooperative choices, be it tested as the absolute deviance from the scale mean or by including a quadratic term of contributions. This supports the assumption that the strength of preference or a lack of decision conflict influence decision time, resulting in the spontaneity of both decision extremes – be it cooperation or defection. Thus, our results support the claim that reversely inferring an intuitive processing mode from short decision times neglects additional concepts related to decision time (A. M. Evans et al., 2015; A. M. Evans & Rand, 2018; Krajbich et al., 2015).

What is not targeted by A. M. Evans et al. (2015) shows the incremental value of both papers of this thesis (Mischkowski & Glöckner, 2016; Mischkowski, Glöckner, et al., 2018): Taking person-situation interactions into account allows one to consider the origin of an individual's strength of preference and – in turn – predict her spontaneously expressed response. In other words, considering SVO as the moderator of spontaneous cooperation points out that decision times for cooperation are rooted in social preferences and correspond to different spontaneously expressed extremes: The inverse u-shaped pattern of cooperation behavior and decision time reflects the default of both groups of social preferences, prosocials and proselfs – an advantage of curvilinear testing that is suppressed when only considering the average linear relation between cooperation and decision time that varies in its direction dependent on the proportion of prosocials and proselfs (for similar results, see also Yamagishi et al., 2017).

Future research should investigate whether self-paced short decision time – implying low decision conflict – might not also entail intuitive processing. Upon closer examination, decision conflict as a cause for deliberation is also incorporated by dual process models (e.g., Pennycook et al., 2015). Thus, it might very well be that prosocials decide *intuitively* to cooperate (Chen & Fischbacher, 2019), as they have a clear preference to do so and thus their decision conflict is a priori low. Yamagishi et al. (2017) investigated the decision conflict of prosocials in more detail – specifically, why prosocials prolong their decision time and simultaneously reduce their cooperation. The authors showed that prosocials' decision times increase with a fear of being exploited. This decision conflict might similarly go in line with an increased (necessity for) deliberation about whether to comply with one's default (i.e., cooperate) or override it (i.e., defect).

Referring again to the SHH, excluding measured decision time (see Rand, 2016) might be a practical approach but is certainly insufficient on theoretical grounds. It suggests that intuitive processing is inherently independent of decision conflict and requires the specification of one's understanding of intuition in the first place. It is a demanding task on its own to grasp a common and still specific definition of intuition (see Glöckner & Witteman, 2010). Yet it again serves the purpose of good theorizing not only to provide labels of constructs and corresponding operationalizations but also to provide a deeper comprehension of the construct that in turn allows one to evaluate the suitability of certain operationalizations. Thus, one has to go beyond the descriptive labels of intuitive processing (e.g., fast, effortless, affective; see Chapter 1.2) to define what intuition is and is not. This would greatly facilitate the understanding regarding the variability of findings concerning the cognitive underpinnings of cooperation behavior and find theory-wise common ground, even though this yields the risk of being obliged to modify or even reject the SHH.

Differentiation for punishment behavior

Testing for the inverse u-shaped pattern of decision time for punishment investments sheds additional light on the comparability of cooperation behavior and punishment in their relation to decision time. Thus, I will test in the following whether the same pattern of decision time also holds for punishment investments – that is, whether similarly low as well as high punishment investments are made faster than an intermediate level of punishment. When transferring the analyses from cooperation to punishment, one has to keep in mind that the situation is a priori different in comparison to analyzing the extremities of cooperation behavior. This is because the available resources to punish differ for each player, as these are conditional on the first stage of the game (i.e., the contribution behavior of the individual and one's group members determine the earnings before a potential punishment). Thus, punishment investments of the same size differ in terms of their relative severity, depending on the outcome of the first stage. Stated differently, decision extremities for punishment are difficult to define without having earnings of the first stage of the public goods game as a reference point. This is why punishment investments are rescaled as the percentage of the individual outcome in the first stage of the game. But even when doing so, the distribution of punishment investments is highly skewed to the right (see Figure 7): Almost two thirds of the overall sample chose not to punish at all (n = 177; 63,90%), which is also reflected in a very low, average investment of 4.28% (SD = 9.39) of the earnings of the first stage to punish (Min = 0%, Max = 50.00%). Decision time for punishment varies between 3.04 and 34.40 seconds (M = 8.22, SD = 3.83), being again logarithmized (M = 0.88, SD = 0.17) to approximate a normal distribution (Mayerl, 2009).

The highly skewed distribution of punishment decisions causes a multicollinearity problem between punishment investments and punishment extremes (i.e., the absolute difference between punishment investments relative to one's earnings and the scale mean), even when mean-centering both variables (r = .93, p < .001). In consequence, the accuracy of the corresponding regression model predicting logarithmized decision time is highly reduced as both predictors are nearly redundant. Thus, neither relative punishment investments (β = -.14, p = .386) nor punishment extremes (β = -.09, p = .553) are predictive of decision time in this model.

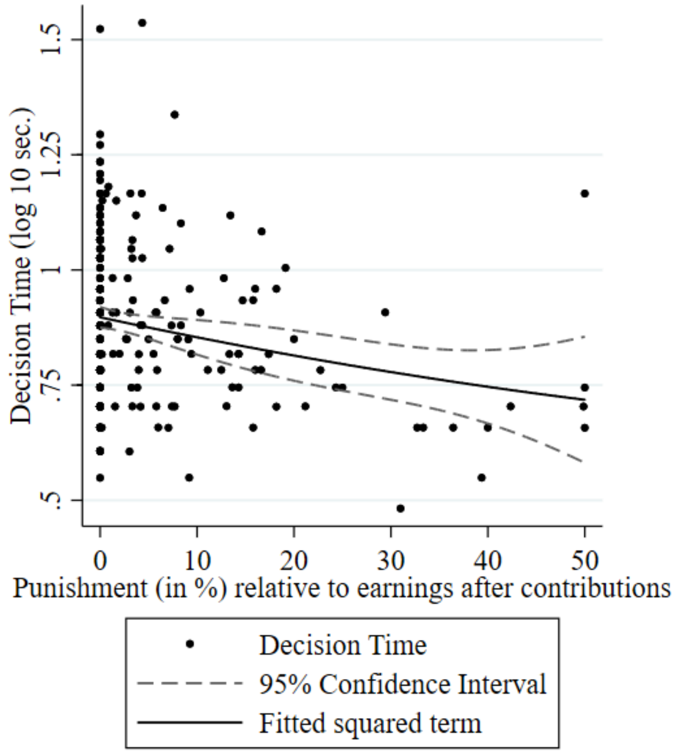

Figure 7. Prediction of decision time as a function of punishment investments, percentual to the earnings after the first stage of the public goods game, and their squared term. Grey-dashed lines denote the 95% confidence interval.

As the graphical inspection already reflects the absence of an inverse u-shaped pattern, the further analysis (Lind & Mehlum, 2007, 2010) consequently supports this null result when including a quadratic term in the prediction of decision time by mean-centered punishment investments (β = .05, p = .715, see Figure 7).[24] Given that the necessary condition of a significant squared term is not fulfilled, it is obsolete to test for the sufficient conditions (i.e., testing whether the extremum is within the data range and a corresponding significant decrease of the curve to both sides of the extremum).

Summarizing these results, the lack of a u-shaped pattern for punishment investments further underlines the distinctiveness of cooperation and punishment behavior despite their conceptual similarities as contributions to first and second order public goods (see Chapter 1.1).

[24] For illustrative purposes, the graph contains non-centered punishment investments.

The following chapter discusses more extensively whether (second-party) punishment represents an act of pro-sociality at all which is crucial to evaluate whether spontaneous punishment (does not) allow for generalizing spontaneous cooperation towards a theory of spontaneous pro-sociality.

4.2.3 Punishment as a prosocial act?

The second article shows that spontaneous cooperation differs in comparison to spontaneous punishment in terms of the underlying motives, even though both cooperation and punishment show a similar pattern over the course of decision time. When only considering the similar relations of cooperation and punishment behavior with decision time, one might argue for summarizing the results under a broader framework of spontaneous pro-sociality. However, when focusing on the underlying motives, our findings suggest that affective rather than prosocial motives influence spontaneous punishment: People invest resources to achieve their goal of retaliation (Camerer & Fehr, 2004). Thus, it is questionable whether the underlying pro-sociality is sufficiently captured in a second-party punishment setting where individuals need to decide whether to cooperate themselves. Typically, third-party punishment is considered altruistic and thus as a form of pro-sociality (Fehr & Fischbacher, 2003). Future research could investigate whether SVO gains importance as a moderator for spontaneous – altruistic – punishment, when the punisher is not a contributing group member, but is responsible to watch over the welfare of the whole group.

However, independent of whether a second- or third-party punishment setting is applied, the general question has been raised whether so-called altruistic punishment is actually related to additional altruistic measures, challenging the pro-sociality of punishment *in general*. Here, contradictory evidence has been found (e.g., Brethel-Haurwitz et al., 2016; Henrich et al., 2006). Evidence linking punishment behavior to dispositional pro-sociality is similarly heterogeneous; for instance, there is no relation with honesty-humility (Zettler, Hilbig, & Heydasch, 2013). As outlined in Chapter 1.1, results for a relation with SVO have been identified in both directions (e.g., Bieleke et al., 2016; Karagonlar & Kuhlman, 2013); the results of Mischkowski, Glöckner, et al. (2018) overall provide no evidence for a relation between SVO and punishment (for similar findings, see also Böckler et al., 2016; Peysakhovich, Nowak, & Rand, 2014; Yamagishi et al., 2012).

Taken together, evidence on the pro-sociality of punishment is too heterogeneous to globally infer its degree of pro-sociality. To further elaborate on the generality of the SHH as a theory of spontaneous pro-sociality, I will summarize evidence for related pro-sociality measures and their dependence on the processing mode, respectively decision time, in the following.

4.2.4 Evidence for related pro-sociality measures

In sum, three additional pro-sociality measures were investigated in their dependence to the processing mode, respectively their relation with decision time. The closest to cooperation behavior being altruism as measured by dictator game giving, next to social mindfulness as a form of low-cost cooperation behavior (Lemmers-Jansen et al., 2018; Van Doesum, Van Lange, & Van Lange, 2013). Finally, there was a recent attempt to apply the SHH to (un-)ethical behavior in that the intuitiveness of cheating was meta-analytically investigated (Köbis et al., 2019).

Starting with altruistic preferences as measured by dictator game giving, these did not provide clear evidence in their relation with decision time in that both spontaneous egocentrism was found (e.g., S. Fiedler et al., 2013; Piovesan & Wengström, 2009) as well as spontaneous fairness as indicated by shorter decision time for fair in comparison to selfish allocations (e.g., Cappelen, Nielsen, Tungodden, Tyran, & Wengström, 2016). A similarly heterogeneous pattern holds when the processing mode was manipulated (e.g., via time pressure or cognitive load): Achtziger, Alós-Ferrer, and Wagner (2015) showed less giving under cognitive load in comparison to a no load condition. Schulz, Fischbacher, Thöni, and Utikal (2014) in turn showed an intuitive striving for fair in comparison to payoff maximizing allocations. Several null effects were similarly found (e.g., Hauge, Brekke, Johansson, Johansson-Stenman, & Svedsäter, 2016; Tinghög et al., 2016). As dictator game giving is closely related to cooperation behavior (see Chapter 1.1), these results were linked to the SHH (e.g., Cappelen et al., 2016), adding to the opaque pattern as outlined in Chapter 2.2. As outlined, Rand (2016) however explicitly excludes dictator games from testing the SHH as these consist only of unilateral giving, which is individually always non-beneficial, be it in one-shot or in iterative interactions.

As a measure of low-cost cooperation, social mindfulness contrasts whether the freedom of choice of another person is left or limited in interdependent relationships (i.e., by leaving or limiting the choice options between two different kinds of objects for a subsequent chooser; Van Lange & Van Doesum, 2015). Specifically, it involves both the skill and the will to act in a socially mindful manner that ensures another individual's control over outcomes (Van Doesum et al., 2013). Based on its comparability to cooperation behavior (Lemmers-Jansen et al., 2018), social mindfulness was strongly suggested to vary dependent on the processing mode (Van Lange & Van Doesum, 2015). However, Mischkowski, Thielmann, et al. (2018) consistently found no effect when testing across several studies and processing mode manipulations.

Finally, a recent meta-analysis on the intuitiveness of cheating behavior was conducted to provide a highly powered test of the question whether there is intuitive (dis-)honesty (Köbis et al., 2019). Similar to the debate on spontaneous cooperation behavior, contradictory findings elicited the debate. To exemplify, it was shown, on the one hand, that 'honesty requires time' (Shalvi, Eldar, & Bereby-Meyer, 2012), whereas, on the other hand, Van't Veer, Stel, and van Beest (2014) showed an intuitively honest attitude of participants. Köbis et al. (2019) meta-analytically confirmed that dishonesty (i.e., cheating behavior) is the intuitive response that is overruled with increasing deliberation, thus fundamentally contradicting a more general approach of spontaneous ethicality. However, the effect reverses when considering a specific moderator, namely whether cheating behavior targets a directly harmed person (e.g., a group member in a two-person cheating paradigm) in contrast to an anonymous experimenter. Köbis et al. (2019) attribute the results in line with the SHH: "In accordance with previous theorizing, particularly the social heuristics hypothesis (Bear & Rand, 2016; Rand, 2016), our data are in line with the idea that salient consequences for others have a substantial impact on people's intuitive decisions." (p. 15).

This broad interpretation of the SHH's application areas – when referring to salient consequences for others as the only perimeter – deserves further investigation, as it has fundamental implications on the SHH's level of universality and would require a reformulation of the SHH's application areas beyond cooperation behavior. Stated differently, these findings are only useful if they are taken to predict future behavior. As is, the SHH is only taken as a post-hoc explanation, mentioning that findings are 'in line with' the theory, but not as a critical test against it (see Platt, 1964).

Taken together, the empirical pattern of conceptual replications or – more broadly – investigation with related pro-sociality measures does not suggest generalizing the SHH to a more universal theory of spontaneous pro-sociality. However, when being limited in its application areas (i.e., its antecedents, in this case cooperation behavior), one cannot claim an intuitiveness of pro-sociality *in general*. Therefore, summarizing the results of Rand and others under the framework of 'intuitive pro-sociality' (Zaki & Mitchell, 2013) was premature. In the following, I will draw a general summary and discuss how future research could most effectively contribute to further develop the SHH.

5. General discussion and outlook

The spontaneous cooperation effect has received enormous attention and correspondingly encouraged a large number of follow-up studies showing how context- and personality dependent the effect is (see Chapter 2.2 and 2.3). As a result of several failed replications, the effect has lost its original impact and spirit. Originally, the effect emphasized cooperation as the initial, ‘spontaneous’ response in social dilemma contexts, underlining the notion that humans are prosocial in the first place and that only increased cognitive effort leads to ‘calculated greed’ – that is, the unintended side effect of promoting selfishness (Rand et al., 2012). More than six years after the original publication, the remaining key message is rather defensive, insisting that when repeated interactions are internalized or expected, spontaneous cooperation also evolves in one-shot social dilemmas (Everett et al., 2017; Isler et al., 2018). Due to this confirmatory evidence, the SHH has not yet been rejected. The reluctance to falsify a theory is a well-known phenomenon (e.g., Wason, 1960), but one might emphasize in terms of the advance of knowledge that “[there are] benefits of being wrong” (Glöckner & Betsch, 2011, p. 711). These benefits – in the Popperian sense of approaching truth by falsification – have so far been neglected with regard to the SHH.

This thesis offers two main contributions to the discussion regarding spontaneous cooperation, both of which are based on theoretical grounds. First, it specifies for whom – that is, which individuals – the prediction of spontaneous cooperation holds and thereby consolidates heterogeneous replication results of the effect: Depending on the average dispositional pro-sociality in a sample, spontaneous cooperation may or may not occur. Second, the thesis tested whether the effect generalizes to second order public goods – that is, punishment decisions – which would increase the theory’s level of universality and thus its empirical content. However, the generalizability is limited: When investigating spontaneous behavior in second order public goods, a similar main effect of spontaneous punishment indicates at first glance that punishment and cooperation are comparable in their course over decision time. When considering the underlying motives, however, clear differences turn visible: The influencing role of negative affect and the only indirect influence of SVO (i.e., via contributions) on spontaneous punishment limit the inference to a more general hypothesis of spontaneous pro-sociality. In this vein, the comparability of elementary and instrumental cooperation in terms of their level of pro-sociality has recently been called into question (see Brethel-Haurwitz et al., 2016).

From a methodological point of view, the two papers of this thesis comply with the standards of reproducibility and open science by providing several replication studies and making all data and materials available. The design of the second paper, in particular, allowed for testing whether both the main effect of spontaneous cooperation and the interaction with SVO replicate when the threat of being punished is anticipated. Admittedly, the results are limited to self-paced (i.e., measured) decision time (see sections 3.2.1 and 4.2.2). The additional analyses provided here show that there is an inverse u-shaped pattern indicating low decision time for decision extremes in elementary cooperation (A. M. Evans et al., 2015). This pattern is attributed to single process models (e.g., Klauer, 2014), showing the importance of decision conflict as a predictor of decision time. Thus, the results provided here corroborate previous findings that short decision times cannot only be ascribed to intuitive processing. However, some dual process models also incorporate (low) decision conflict as a prerequisite for intuitive decision making (De Neys & Glumicic, 2008; Pennycook et al., 2015). Future research is needed to disentangle when long decision times correspond to deliberation and when to a lacking strength of preference (Konovalov & Krajbich, 2019).

To have a more constructive scientific debate about the spontaneous cooperation effect that goes beyond discussing how to best operationalize processing mode and which individuals (not) to exclude (see Bouwmeester et al., 2017; Rand, 2017a, 2018), the corresponding theory needs to be challenged by competing theoretical explanations. In other words, the challenge the SHH must accomplish is to go one step further from post-hoc explanation of identified effects towards predicting behavior a priori. Similar to the aim of the second paper of this thesis, it is an important step to explore the level of universality of a theory, but the next step is clearly in order: Results must be used to propose a modified version of the SHH that can then be critically tested (e.g., testing when deliberation promotes pro-sociality) rather than attributing mixed results only to methodological reasons.

5.1 Future directions for research on spontaneous cooperation

The immunity to critical testing, at least to a certain degree, is a common phenomenon among theories in judgment and decision making (Glöckner & Betsch, 2011). However, it should not be accepted without question. The current lack of empirical content (Popper, 1934/2005) impedes the development of an experimental design that aims at *disproving* the SHH, intending to show its boundaries of validity where cooperation is the deliberate yet non-beneficial choice. In contrast to purely effect-based research, revising and developing the theory is certainly the more difficult endeavor.

However, there is a starting point for future research, as results do exist that show the opposite pattern of increased cooperation under deliberation (even though this evidence has not yet been used to revise the theory). For instance, Yamagishi et al. (2017) show a behavioral pattern of increased cooperation behavior under deliberation for proselfs. Does this falsify the SHH? In line with the current methodological dispute, one could either argue that measuring decision time as Yamagishi et al. (2017) did cannot unequivocally operationalize processing mode and thus may not be used to test the SHH. An approach that focuses on the accumulation of evidence might refer to the different pattern presented in both articles of this thesis, where cooperation remains stable across decision time for proselfs. In general, one cannot consider a theory as falsified based on one or two occasional findings (Glöckner & Betsch, 2011). Thus, it is legitimate to ask for clarification in a meta-analysis, as has been actually called for (JDM-mailing list call of Andrighetto, Bruner, Steinmo, Szekely, & Capraro, 07.02.2018). However, as the SHH forbids increased reflective cooperation, the finding of Yamagishi et al. (2017)

should be taken as a starting point to explore the SHH's boundaries, when deliberation rather than intuition favors cooperation. Metaphorically speaking: Adding further evidence showing spontaneous cooperation is similar to presenting a white swan to support the hypothesis that all swans are white (see J. S. B. Evans, 1982). To gain knowledge, the search for a black swan (that is, replicable evidence that shows the boundaries of spontaneous cooperation) is needed. Strictly speaking, future research has few worthwhile alternatives to this path.

5.2 Conclusion

The replication history of the spontaneous cooperation effect and the underlying SHH contained at first glance all important aspects that heterogeneous findings ask for: Several replication analyses, a meta-analysis, or a registered replication report – most of said endeavors being conducted under the principles of open science, providing preregistrations, open materials, and data. What is lacking is a focus on the SHH as the underlying *theory* and what it forbids rather than attempting (and partially failing) to add confirmatory evidence. The articles of this thesis contribute to specify the SHH, for whom spontaneous cooperation holds, in addition to a test of its (limited) generalizability. Developing alternative theoretical explications that also account for increased cooperation with decision time (see Yamagishi et al., 2017) depending on individuals' social preferences is the next important step this research line must take, as it counteracts the only specification of the original SHH, stating that cooperation should never increase with deliberation (Rand et al., 2014).

From a meta-scientific perspective, existing research on spontaneous cooperation has received outstanding attention (e.g., one meta-analysis and one registered replication report in *Psychological Science* in less than a year). Therefore, contributing researchers have – and should make use of – the rare chance to lead by example in terms of how to proceed when an effect is under dispute. It is important to move the field from a methodological, effect-based debate towards a theory-driven discussion. The spontaneous cooperation effect and its underlying theory provide the opportunity to do so. Taking advantage of this opportunity consists of a cooperative act to the public good of accumulating knowledge itself – be it spontaneous or deliberate.

References

Achtziger, A., & Alós-Ferrer, C. (2013). Fast or rational? A response-times study of Bayesian updating. *Management Science, 60*(4), 923-938. https://doi.org/10.1287/mnsc.2013.1793

Achtziger, A., Alós-Ferrer, C., & Wagner, A. K. (2015). Money, depletion, and prosociality in the dictator game. *Journal of Neuroscience, Psychology, and Economics, 8*(1), 1-14. https://doi.org/10.1037/npe0000031

Ackermann, K. A. (2011, July). *Measuring Social Value Orientation.* Paper presented at the 14th international conference on social dilemmas, Netherlands: Amsterdam.

Ackermann, K. A., & Murphy, R. O. (2012). Tutorial on how to evaluate the SVO Slider Measure's secondary items. Retrieved from http://ryanomurphy.com/resources/SVO_second_item_tutorial.pdf

Alós-Ferrer, C., & Strack, F. (2014). From dual processes to multiple selves: Implications for economic behavior. *Journal of Economic Psychology, 41*, 1-11. https://doi.org/10.1016/j.joep.2013.12.005

Alós-Ferrer, C. (2018). A Dual-Process Diffusion Model. *Journal of Behavioral Decision Making, 31*(2), 203-218. https://doi.org/10.1002/bdm.1960

Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics, 37*(3), 291-304. https://doi.org/10.1016/0047-2727(88)90043-6

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150-166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2008). The HEXACO model of personality structure and the importance of the H factor. *Social and Personality Psychology Compass, 2*, 1952–1962. https://doi.org/10.1111/j.1751-9004.2008.00134.x

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes and Intergroup Relatations, 12*(4), 533-547. https://doi.org/10.1177/1368430209105040

Bear, A., & Rand, D. G. (2016). Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences, 113*(4), 936-941. https://doi.org/10.1073/pnas.1517780113

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*(2), 169-217. https://doi.org/10.1086/259394

Becker, G. S. (1976). *The economic approach to human behavior*. Chicago: University of Chicago Press.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122-142. https://doi.org/https://doi.org/10.1006/game.1995.1027

Bieleke, M., Gollwitzer, P. M., Oettingen, G., & Fischbacher, U. (2016). Social value orientation moderates the effects of intuition versus reflection on responses to unfair ultimatum offers. *Journal of Behavioral Decision Making, 30*(2), 569–581 https://doi.org/10.1002/bdm.1975

Böckler, A., Tusche, A., & Singer, T. (2016). The structure of human prosociality: Differentiating altruistically motivated, norm motivated, strategically motivated, and self-reported prosocial behavior. *Social Psychological and Personality Science, 7*(6), 530-541. https://doi.org/10.1177/1948550616639650

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 166-193. https://doi.org/10.1257/aer.90.1.166

Bosman, R., & Van Winden, F. (2002). Emotional hazard in a power-to-take experiment. *The Economic Journal, 112*(476), 147-169. https://doi.org/10.1111/1468-0297.0j677

Bouwmeester, S., Verkoeijen, P. P. J. L., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., . . . Wollbrant, W. E. (2017). Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science, 12*(3), 527-542. https://doi.org/10.1177/1745691617693624

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13*(3), 171-195. https://doi.org/10.1016/0162-3095(92)90032-Y

Brethel-Haurwitz, K. M., Stoycos, S. A., Cardinale, E. M., Huebner, B., & Marsh, A. A. (2016). Is costly punishment altruistic? Exploring rejection of unfair offers in the Ultimatum Game in real-world altruists. *Scientific Reports, 6*, 18974. https://doi.org/10.1038/srep18974

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

Burnstein, E., Crandall, C., & Kitayama, S. (1994). Some neo-Darwinian decision rules for altruism: Weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology, 67*(5), 773-789. https://doi.org/10.1037/0022-3514.67.5.773

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., . . . Pfeiffer, T. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behavior, 2*(9), 637-644. https://doi.org/10.1038/s41562-018-0399-z

Camerer, C. F., & Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. In J. Henrich, R. Boyd, S. Bowles,

C. F. Camerer, E. Fehr, & H. Gintis (Eds.), *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies* (pp. 55-95). Oxford: Oxford University Press.

Camerer, C. F., & Fehr, E. (2006). When does "economic man" dominate social behavior? *Science, 311*(5757), 47-52. https://doi.org/10.1126/science.1110600

Camerer, C. F., & Loewenstein, G. (2011). Behavioral economics: Past, present, future. In C. F. Camerer, G. Loewenstein, & M. Rabin (Eds.), *Advances in behavioral economics* (pp. 3-52). Princeton, New Jersey: Princeton University Press.

Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., & Wengström, E. (2016). Fairness is intuitive. *Experimental Economics, 19*(4), 727-740. https://doi.org/10.1007/s10683-015-9463-y

Capraro, V., & Cococcioni, G. (2016). Rethinking spontaneous giving: Extreme time pressure and ego-depletion favor self-regarding reactions. *Scientific Reports, 6*, 27219. https://doi.org/10.1038/srep27219

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics, 117*(3), 817-869. https://doi.org/10.1162/003355302760193904

Chen, F., & Fischbacher, U. (2016). Response time and click position: Cheap indicators of preferences. *Journal of the Economic Science Association, 2*(2), 109-126. https://doi.org/10.1007/s40881-016-0026-6

Chen, F., & Fischbacher, U. (2019). Cognitive processes underlying distributional preferences: A response time study. *Experimental Economics*. https://doi.org/10.1007/s10683-019-09618-x

Chen, F., & Krajbich, I. (2018). Biased sequential sampling underlies the effects of time pressure and delay in social decision making. *Nature Communications, 9*, 3557. https://doi.org/10.1038/s41467-018-05994-9

Clithero, J. A. (2018). Response times in economics: Looking through the lens of sequential sampling models. *Journal of Economic Psychology*, *69*, 61–86. https://doi.org/10.1016/j.joep.2018.09.008

Cone, J., & Rand, D. G. (2014). Time pressure increases cooperation in competitively framed social dilemmas. *PloS One*, *9*(12), e115756. https://doi.org/10.1371/journal.pone.0115756

Coricelli, G., Polonio, L., & Vostroknutov, A. (forthcoming). The process of choice in games. In M. Capra, R. Croson, T. Rosenblatt, & R. M. (Eds.), *Handbook of experimental game theory*. Cheltenham: Edward Elgar Publishing.

Cornelissen, G., Dewitte, S., & Warlop, L. (2011). Are social value orientations expressed automatically? Decision making in the dictator game. *Personality and Social Psychology Bulletin, 37*(8), 1080-1090. https://doi.org/10.1177/0146167211405996

Cramton, P., MacKay, D. J., Ockenfels, A., & Stoft, S. (2017). *Global carbon pricing: The path to climate cooperation*. Cambridge, Massachusetts: The MIT Press.

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

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*(1), 169-193. https://doi.org/10.1146/annurev.ps.31.020180.001125

Dawes, R. M., & Thaler, R. H. (1988). Anomalies: Cooperation. *Journal of Economic Perspectives, 2*(3), 187-197. https://doi.org/10.1257/jep.2.3.187

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., . . . Feith, S. 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 Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science, 7*(1), 28-38. https://doi.org/10.1177/1745691611429354

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning, 20*(2), 169-187. https://doi.org/10.1080/13546783.2013.854725

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*(3), 1248-1299. https://doi.org/10.1016/j.cognition.2007.06.002

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning, 11*(4), 349-381. https://doi.org/10.1080/13546780442000222

Declerck, C. H., Boone, C., & Emonds, G. (2013). When do people cooperate? The neuroeconomics of prosocial decision making. *Brain and Cognition, 81*(1), 95-117. https://doi.org/10.1016/j.bandc.2012.09.009

Denant-Boemont, L., Masclet, D., & Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory, 33*(1), 145-167. https://doi.org/10.1007/s00199-007-0212-0

Diederich, A. (2003). Decision making under conflict: Decision time as a measure of conflict strength. *Psychonomic Bulletin & Review, 10*(1), 167-176. https://doi.org/10.3758/BF03196481

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.), *Mechanisms of insight*. Cambridge, MA: MIT Press.

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

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*(8), 709-724. https://doi.org/10.1037/0003-066X.49.8.709

Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General, 144,* 951-966. https://doi.org/10.1037/xge0000107

Evans, A. M., & Rand, D. G. (2018). Cooperation and decision time. *Current Opinion in Psychology, 26*, 67-71. https://doi.org/10.1016/j.copsyc.2018.05.007

Evans, J. S. B. (1982). *The psychology of deductive reasoning (Psychology revivals)*. New York, NY: Psychology Press.

Evans, J. S. B. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*(4), 451-468. https://doi.org/10.1111/j.2044-8295.1984.tb01915.x

Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13*(4), 321-339. https://doi.org/10.1080/13546780601008825

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Everett, J., Ingbretsen, Z., Cushman, F., & Cikara, M. (2017). Deliberation erodes cooperative behaviour – Even towards competitive outgroups, even when using a control condition, and even when controlling for sample bias. *Journal of Experimental Social Psychology, 73*, 76-81. https://doi.org/10.1016/j.jesp.2017.06.014

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*, 785-791. https://doi.org/10.1038/nature02043

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*, 980-994. https://doi.org/10.1257/aer.90.4.980

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137-140. https://doi.org/10.1038/415137a

Fehr, E., & Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. *Current Opinion in Neurobiology, 14*(6), 784-790. https://doi.org/10.1016/j.conb.2004.10.007

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817-868. https://doi.org/10.1162/003355399556151

Fiedler, K. (2004). Tools, toys, truisms, and theories: Some thoughts on the creative cycle of theory formation. *Personality and Social Psychology Review, 8*(2), 123-131. https://doi.org/https://doi.org/10.1207/s15327957pspr0802_5

Fiedler, S., Glöckner, A., & Nicklisch, A. (2012). The influence of social value orientation on information processing in repeated voluntary contribution mechanism games: An eyetracking analysis. In A. Innocenti & A. Sirigu (Eds.), *Neuroscience and the economics of decision making* (Vol. 5, pp. 21-53). New York: Routledge.

Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social Value Orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes, 120*(2), 272-284. https://doi.org/10.1016/j.obhdp.2012.07.002

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*(3), 397-404. https://doi.org/10.1016/S0165-1765(01)00394-9

Fischbacher, U., Hertwig, R., & Bruhin, A. (2013). How to model heterogeneity in costly punishment: Insights from responders' response times. *Journal of Behavioral Decision Making, 26*(5), 462-476. https://doi.org/10.1002/bdm.1779

Frankish, K., & Evans, J. S. B. T. (1983). The duality of mind: An historical perspective. In K. Frankish & J. S. B. T. Evans (Eds.), *In two minds: Dual processes and beyond* (pp. 1-30). Oxford, UK: Oxford University Press.

Glöckner, A. (2019, January). *Cumulative science. From transparency to efficient theory development.* Paper presented at the Workshop "Cumulative Science: From Transparency to Efficient Theory Development", Saarland University.

Glöckner, A., & Betsch, T. (2011). The empirical content of theories in judgment and decision making: Shortcomings and remedies. *Judgment and Decision Making, 6*(8), 711-721. Retrieved from http://journal.sjdm.org/11/m38/m38.pdf

Glöckner, A., & Witteman, C. (2010). Beyond dual-process models: A categorisation of processes underlying intuitive judgement and decision making. *Thinking & Reasoning, 16*(1), 1-25. https://doi.org/10.1080/13546780903395748

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3*(4), 367-388. https://doi.org/10.1016/0167-2681(82)90011-7

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*(3859), 1243-1248. https://doi.org/10.1126/science.162.3859.1243

Haruno, M., Kimura, M., & Frith, C. D. (2014). Activity in the nucleus accumbens and amygdala underlies individual differences in prosocial and individualistic economic choices. *Journal of Cognitive Neuroscience, 26*(8), 1861-1870. https://doi.org/10.1162/jocn_a_00589

Hauge, K. E., Brekke, K. A., Johansson, L.-O., Johansson-Stenman, O., & Svedsäter, H. (2016). Keeping others in our mind or in our heart? Distribution games under cognitive load. *Experimental Economics, 19*(3), 562-576. https://doi.org/10.1007/s10683-015-9454-z

Head, J. G. (1974). *Head, public goods and public welfare*. Durham, North Carolina: Duke University Press.

Hechter, M. (1987). *Theories of group solidarity*. Berkeley, CA: University of California Press.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Henrich, N. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767-1770. https://doi.org/10.1126/science.1127333

Hilbig, B. E., Glöckner, A., & Zettler, I. (2014). Personality and pro-social behavior: Linking basic traits and social value orientations. *Journal of Personality and Social Psychology, 107*(3), 529–539. https://doi.org/10.1037/a0036074

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*, 598–603. https://doi.org/10.1016/j.paid.2012.11.008

Horstmann, N., Hausmann, D., & Ryf, S. (2009). 12 methods for inducing intuitive and deliberate processing modes. In A. Glöckner & C. Witteman (Eds.), *Foundations for tracing intuition: Challenges and methods* (pp. 219-236). London, UK: Psychology Press & Routledge.

Isaac, R. M., Walker, J. M., & Thomas, S. H. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, *43*(2), 113-149. https://doi.org/10.1007/BF00140829

Isler, O., Maule, J., & Starmer, C. (2018). Is intuition really cooperative? Improved tests support the social heuristics hypothesis. *PloS One, 13*(1), e0190560. https://doi.org/10.1371/journal.pone.0190560

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review, 93*(5), 1449-1475. https://doi.org/10.1257/000282803322655392

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Karagonlar, G., & Kuhlman, D. M. (2013). The role of social value orientation in response to an unfair offer in the ultimatum game. *Organizational Behavior and Human Decision Processes, 120*(2), 228-239. https://doi.org/10.1016/j.obhdp.2012.07.006

Kelley, H. H., & Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology, 16*(1), 66-91. https://doi.org/10.1037/h0029849

Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making, 9*(6), 510-522. Retrieved from http://journal.sjdm.org/14/14808/jdm14808.pdf

Klauer, K. C. (2014). Random-walk and diffusion models. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 139-152). New York: Guilford Publications.

Köbis, N., Verschuere, B., Bereby-Meyer, Y., Rand, D. G., & Shalvi, S. (2019). Intuitive honesty versus dishonesty: Meta-analytic evidence. *Perspectives on Psychological Science*, 1-19. https://doi.org/10.1177/1745691619851778

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology, 24*(1), 183-214. https://doi.org/10.1146/annurev.soc.24.1.183

Konovalov, A., & Krajbich, I. (2019). Revealed strength of preference: Inference from response times. *Judgment and Decision Making, 14*(4), 381–394. Retrieved from http://journal.sjdm.org/18/181101/jdm181101.pdf

Krajbich, I., Bartling, B., Hare, T., & Fehr, F., E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications, 6*, 7455. https://doi.org/10.1038/ncomms8455

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences, 108*(33), 13852-13857. https://doi.org/10.1073/pnas.1101328108

Kreps, D. M. (1990). Corporate culture and economic theory. In J. E. Alt & K. A. Shepsle (Eds.), *Perspectives on positive political economy* (pp. 90–143). Oxford: Oxford University Press.

Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory, 27*(2), 245-252. https://doi.org/10.1016/0022-0531(82)90029-1

Kruglanski, A. W. (2004). The quest for the gist: On challenges of going abstract in social and personality psychology. *Personality and Social Psychology Review, 8*(2), 156-163. https://doi.org/10.1207/s15327957pspr0802_9

Kvarven, A., Strømland, E., Wollbrant, C., Andersson, D., Johannesson, M., Tinghög, G., . . . Myrseth, K. O. R. (2019). *The intuitive cooperation hypothesis revisited: A meta-analytic examination of effect-size and between-study heterogeneity*. Retrieved from https://osf.io/preprints/metaarxiv/kvzg3/

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 111-194). Princeton, NJ: Princeton University Press.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*(2), 329-358. https://doi.org/10.1207/s15327906mbr3902_8

Lemmers-Jansen, I. L. J., Krabbendam, L., Amodio, D. M., van Doesum, N. J., Veltman, D. J., & van Lange, P. A. M. (2018). Giving others the option of choice: An fMRI study on low-cost cooperation. *Neuropsychologia*, 1-9. https://doi.org/10.1016/j.neuropsychologia.2017.12.009

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science, 18*(5), 421-428. https://doi.org/10.1111/j.1467-9280.2007.01916.x

Liebrand, W. B. G., Jansen, R. W. T. L., Rijken, V. M., & Suhre, C. J. M. (1986). Might over morality: Social values and the perception of other players in experimental games. *Journal of Experimental Social Psychology, 22*(3), 203-215. https://doi.org/10.1016/0022-1031(86)90024-7

Liebrand, W. B. G., & McClintock, C. G. (1988). The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *European Journal of Personality, 2*(3), 217-230. https://doi.org/10.1002/per.2410020304

Lind, J. T., & Mehlum, H. (2007). Utest: Stata module to test for a U-shaped relationship (Version revised 09 Feb 2019). Boston: College Department of Economics. Retrieved from https://ideas.repec.org/c/boc/bocode/s456874.html

Lind, J. T., & Mehlum, H. (2010). With or without U? The appropriate test for a U-shaped relationship. *Oxford Bulletin of Economics and Statistics, 72*(1), 109-118. https://doi.org/10.1111/j.1468-0084.2009.00569.x

Lohse, J., Goeschl, T., & Diederich, J. H. (2016). Giving is a question of time: Response times and contributions to an environmental public good. *Environmental and Resource Economics, 3*(67), 455-477. https://doi.org/10.1007/s10640-016-0029-z

Lotito, G., Migheli, M., & Ortona, G. (2013). Is cooperation instinctive? Evidence from the response times in a public goods game. *Journal of Bioeconomics, 15*(2), 123-133. https://doi.org/10.1007/s10818-012-9141-5

Lotz, S. (2015). Spontaneous giving under structural inequality: Intuition promotes cooperation in asymmetric social dilemmas. *PloS One, 10*(7), e0131562. https://doi.org/10.1371/journal.pone.0131562]

Manktelow, K. I. (2012). *Thinking and reasoning. An introduction to the psychology of reason, judgment and decision making*. London: Psychology Press.

Marwell, G., & Ames, R. E. (1979). Experiments on the provision of public goods. I. Resources, interest, group size, and the free-rider problem. *American Journal of Sociology, 84*(6), 1335-1360. https://doi.org/10.1086/226937

Mayerl, J. (2009). *Kognitive Grundlagen sozialen Verhaltens: Framing, Einstellungen und Rationalität*. Wiesbaden: VS Verlag für Sozialwissenschaften.

McAuliffe, W. H., Forster, D. E., Pedersen, E. J., & McCullough, M. E. (2018). Experience with anonymous interactions reduces intuitive cooperation. *Nature Human Behaviour, 2*(12), 909-914. https://doi.org/https://doi.org/10.1038/s41562-018-0454-9

Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A., & Marotzke, J. (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences, 105*(7), 2291-2294. https://doi.org/10.1073/pnas.0709546105

Mischkowski, D., & Glöckner, A. (2016). Spontaneous cooperation for prosocials, but not for proselfs: Social value orientation moderates spontaneous cooperation behavior. *Scientific Reports, 6*, 21555. https://doi.org/10.1038/srep21555

Mischkowski, D., Glöckner, A., & Lewisch, P. (2018). From spontaneous cooperation to spontaneous punishment – Distinguishing the underlying motives driving spontaneous behavior in first and second order public good games. *Organizational Behavior and Human Decision Processes, 149*, 59-72. https://doi.org/10.1016/j.obhdp.2018.07.001

Mischkowski, D., Thielmann, I., & Glöckner, A. (2018). Think it through before making a choice? Processing mode does not influence social mindfulness. *Journal of Experimental Social Psychology, 74*, 85-97. https://doi.org/http://dx.doi.org/10.1016/j.jesp.2017.09.001

Mischkowski, D., Thielmann, I., & Glöckner, A. (2019). Minimizing inequality versus maximizing joint gains: On the relation between personality traits and different prosocial motivations. *Journal of Research in Personality, 81*, 153-157. https://doi.org/10.1016/j.jrp.2019.06.001

Montealegre, A., & Jimenez-Leal, W. (2019). The role of trust in the social heuristics hypothesis. *PloS One, 14*(5), e0216329. https://doi.org/10.1371/journal.pone.0216329

Mosleh, M., & Rand, D. G. (2018). Population structure promotes the evolution of intuitive cooperation and inhibits deliberation. *Nature*, *8*(1), 6293. https://doi.org/10.1038/s41598-018-24473-1

Mrkva, K. (2017). Giving, fast and slow: Reflection increases costly (but not uncostly) charitable giving. *Journal of Behavioral Decision Making, 30*(5), 1052-1065. https://doi.org/10.1002/bdm.2023

Murphy, R. O., & Ackermann, K. A. (2014). Social Value Orientation: Theoretical and measurement issues in the study of social preferences. *Personallity and Social Psychology Review, 18(1)* 13–41. https://doi.org/10.1177/1088868313501745

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making, 6*(8), 771–781. Retrieved from http://journal.sjdm.org/11/m25/m25.pdf

Myrseth, K. O. R., & Wollbrant, C. E. (2017). Cognitive foundations of cooperation revisited: Commentary on Rand et al. (2012, 2014). *Journal of Behavioral and Experimental Economics, 69,* 133-138. https://doi.org/http://dx.doi.org/10.1016/j.socec.2017.01.005

Nielsen, U. H., Tyran, J.-R., & Wengström, E. (2014). Second thoughts on free riding. *Economics Letters, 122*(2), 136-139. https://doi.org/10.1016/j.econlet.2013.11.021

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics, 92*(1), 91-112. https://doi.org/10.1016/j.jpubeco.2007.04.008

Nikiforakis, N., & Engelmann, D. (2011). Altruistic punishment and the threat of feuds. *Journal of Economic Behavior & Organization, 78*(3), 319-332. https://doi.org/10.1016/j.jebo.2011.01.017

Nishi, A., Christakis, N. A., & Rand, D. G. (2017). Cooperation, decision time, and culture: Online experiments with American and Indian participants. *PloS One, 12*(2), e0171252. https://doi.org/10.1371/journal.pone.0171252

Oliver, P. (1980). Rewards and punishments as selective incentives for collective action: theoretical investigations. *American Journal of Sociology, 85*(6), 1356-1375. https://doi.org/10.1086/227168

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). https://doi.org/10.1126/science.aac4716

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59*(1), 185-198. https://doi.org/10.1016/S1071-5819(03)00017-X

Pashler, H., & Wagenmakers, E. J. (2012). Editors’ introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528-530. https://doi.org/10.1177/1745691612465253

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology, 80*, 34-72. https://doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G., Fugelsang, J. A., Koehler, D. J., & Thompson, V. A. (2016). Commentary: Rethinking fast and slow based on a critique of reaction-time reverse inference. *Frontiers in Psychology, 7*, 1174. https://doi.org/10.3389/fpsyg.2016.01174

Peters, E., & Slovic, P. (2007). Affective asynchrony and the measurement of the affective attitude component. *Cognition and Emotion, 21*(2), 300-329. https://doi.org/10.1080/02699930600911440

Pettit, P., & Sugden, R. (1989). The backward induction paradox. *The Journal of Philosophy, 86*(4), 169-182. https://doi.org/10.2307/2026960

Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications, 5*, 4939. https://doi.org/10.1038/ncomms5939

Peysakhovich, A., & Rand, D. G. (2015). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science, 62*(3), 631-647. https://doi.org/10.1287/mnsc.2015.2168

Piovesan, M., & Wengström, E. (2009). Fast or fair? A study of response times. *Economics Letters*, *105*(2), 193-196. https://doi.org/10.1016/j.econlet.2009.07.017

Platt, J. R. (1964). Strong inference. *Science, 146*(3642), 347-353. Retrieved from http://science.sciencemag.org/content/sci/146/3642/347.full.pdf

Popper, K. R. (1934/2005). *Logik der Forschung* (11th ed.). Tübingen: Mohr Siebeck.

Protzko, J., Ouimette, B., & Schooler, J. (2016). Believing there is no free will corrupts intuitive cooperation. *Cognition, 151*, 6-9. https://doi.org/10.1016/j.cognition.2016.02.014

Rand, D. G. (2016). Cooperation, fast and slow. *Psychological Science, 27*(9), 1192-1206. https://doi.org/10.1177/0956797616654455

Rand, D. G. (2017a). Reflections on the time-pressure cooperation registered replication report. *Perspectives on Psychological Science, 12*(3), 543-547. https://doi.org/10.1177/1745691617693625

Rand, D. G. (2017b). Social dilemma cooperation (unlike dictator game giving) is intuitive for men as well as women. *Journal of Experimental Social Psychology, 73*, 164-168. https://doi.org/10.1016/j.jesp.2017.06.013

Rand, D. G. (2018). Non-naïvety may reduce the effect of intuition manipulations. *Nature Human Behaviour, 2*(9), 602. https://doi.org/10.1038/s41562-018-0404-6

Rand, D. G., Brescoll, V. L., Everett, J. A., 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., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*, 427-430. https://doi.org/10.1038/nature11467

Rand, D. G., Greene, J. D., & Nowak, M. A. (2013). Rand et al. reply. *Nature, 498*(7452), E2-E3. https://doi.org/10.1038/nature12195

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences, 17*(8), 413-425. https://doi.org/10.1016/j.tics.2013.06.003

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

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation*. Ann Arbor, MI: University of Michigan Press.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*(2), 333. https://doi.org/10.1037/0033-295X.111.2.333

Renkewitz, F., Fuchs, H. M., & Fiedler, S. (2011). Is there evidence of publication biases in JDM research? *Judgment and Decision Making, 6*(8), 870-888. Retrieved from http://journal.sjdm.org/11/m27/m27.pdf

Rubinstein, A. (2007). Instinctive and cognitive reasoning: A study of response times. *The Economic Journal, 117*(523), 1243-1259. https://doi.org/10.1111/j.1468-0297.2007.02081.x

Santa, J. C., Exadaktylos, F., & Soto-Faraco, S. (2018). Beliefs about others' intentions determine whether cooperation is the faster choice. *Scientific Reports, 8*, 7509. https://doi.org/10.1038/s41598-018-25926-3

Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology, 41*, 77-87. https://doi.org/10.1016/j.joep.2012.08.007

Selten, R. (1978). The chain store paradox. *Theory and Decision, 9*(2), 127-159. https://doi.org/10.1007/BF00131770

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science, 23*(10), 1264-1270. https://doi.org/10.1177/0956797612443835

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127-190. https://doi.org/10.1037/0033-295X.84.2.127

Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science, 1*(4), 538-555. https://doi.org/10.1177/2515245918805755

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22. https://doi.org/10.1037/0033-2909.119.1.3

Sloman, S. A. (2014). Two systems of reasoning: An update. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 69-79). New York: Guilford Publications.

Smith, A., Bernheim, B. D., Camerer, C. F., & Rangel, A. (2014). Neural activity reveals preferences without choices. *American Economic Journal: Microeconomics, 6*(2), 1-36. https://doi.org/10.1257/mic.6.2.1

Spiliopoulos, L., & Ortmann, A. (2018). The BCD of response time analysis in experimental economics. *Experimental Economics, 21*(2), 383–433. https://doi.org/10.1007/s10683-017-9528-1

Stagnaro, M. N., Arechar, A. A., & Rand, D. G. (2017). From good institutions to generous citizens: Top-down incentives to cooperate promote subsequent prosociality but not norm enforcement. *Cognition, 167*, 212-254. https://doi.org/10.1016/j.cognition.2017.01.017

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220-247. https://doi.org/10.1207/s15327957pspr0803_1

Stromland, E., Tjotta, S., & Torsvik, G. (2018). *Understanding and intuitive cooperation*. Retrieved from https://ssrn.com/abstract=3120986

Tetlock, P. E., & Mellers, B. A. (2002). The great rationality debate. *Psychological Science, 13*(1), 94-99. https://doi.org/10.1111/1467-9280.00418

Thielmann, I., & Hilbig, B. E. (2015). Trust: An integrative review from a person-situation perspective. *Review of General Psychology, 19*(3), 249-277. https://doi.org/10.1037/gpr0000046

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., . . . Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature, 498*, 427-430. https://doi.org/10.1038/nature12194

Tinghög, G., Andersson, D., Bonn, C., Johannesson, M., Kirchler, M., Koppel, L., & Västfjäll, D. (2016). Intuition and moral decision-making–the effect of time pressure and cognitive load on moral judgment and altruistic behavior. *PloS One, 11*(10), e0164012. https://doi.org/10.1371/journal.pone.0164012

Tucker, A. W. (1983). The mathematics of tucker: A sampler. *The Two-Year College Mathematics Journal, 14*(3), 228-232. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/00494925.1983.11972687

Van't Veer, A., Stel, M., & van Beest, I. (2014). Limited capacity to lie: Cognitive load interferes with being dishonest. *Judgment and Decision Making, 9*(3), 199-206. Retrieved from http://journal.sjdm.org/13/131120/jdm131120.pdf

Van de Calseyde, P. P., Keren, G., & Zeelenberg, M. (2014). Decision time as information in judgment and choice. *Organizational Behavior and Human Decision Processes, 125*(2), 113-122. https://doi.org/10.1016/j.obhdp.2014.07.001

Van Doesum, N. J., Van Lange, D. A. W., & Van Lange, P. A. M. (2013). Social Mindfulness: Skill and Will to Navigate the Social World. *Journal of Personality and Social Psychology, 105*(1), 86–103. https://doi.org/10.1037/a0032540

Van Lange, P. A. M. (1992). Confidence in expectations: A test of the triangle hypothesis. *European Journal of Personality, 6*(5), 371-379. https://doi.org/10.1002/per.2410060505

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77*(2), 337-349. https://doi.org/10.1037/0022-3514.77.2.337

Van Lange, P. A. M., Joireman, J. A., Parks, C. D., & van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes, 120*, 125-141. https://doi.org/10.1016/j.obhdp.2012.11.003

Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., & Joireman, J. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73*(4), 733-746. https://doi.org/10.1037/0022-3514.73.4.733

Van Lange, P. A. M., & Van Doesum, N. J. (2015). Social mindfulness and social hostility. *Current Opinion in Behavioral Sciences, 3*, 18–24. https://doi.org/10.1016/j.cobeha.2014.12.009

Verkoeijen, P., & Bouwmeester, S. (2014). Does intuition cause cooperation. *PloS One, 9*(5), e96654. https://doi.org/10.1371/journal.pone.0096654

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*(5765), 1301-1303. https://doi.org/10.1126/science.1121448

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*(3), 129-140. https://doi.org/10.1080/17470216008416717

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology, 60*, 53-85. https://doi.org/10.1146/annurev.psych.60.110707.163633

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*(1), 110-116. https://doi.org/10.1037/0022-3514.51.1.110

Yamagishi, T. (1988). Seriousness of social dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly, 51*(1), 32-42. https://doi.org/10.2307/2786982

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., . . . Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences, 109*(50), 20364-20368. https://doi.org/10.1073/pnas.1212126109

Yamagishi, T., Matsumoto, Y., Kiyonari, T., Takagishi, H., Li, Y., Kanai, R., & Sakagami, M. (2017). Response time in economic games reflects different types of decision conflict for prosocial and proself individuals. *Proceedings of the National Academy of Sciences, 114*(24), 6394-6399. https://doi.org/10.1073/pnas.1608877114

Yamagishi, T., Terai, S., Kiyonari, T., Mifune, N., & Kanazawa, S. (2007). The social exchange heuristic: Managing errors in social exchange. *Rationality and Society, 19*(3), 259-291. https://doi.org/10.1177/1043463107080449

Yong, E. (2012). Replication studies: Bad copy. *Nature, 485*(7398), 298-300. https://doi.org/10.1038/485298a

Zaki, J., & Mitchell, J. P. (2013). Intuitive prosociality. *Current Directions in Psychological Science, 22*(6), 466 –470. https://doi.org/10.1177/0963721413492764

Zettler, I., Hilbig, B. E., & Heydasch, T. (2013). Two sides of one coin: Honesty–Humility and situational factors mutually shape social dilemma decision making. *Journal of Resarch in Personality, 47*(4), 286–295. https://doi.org/10.1016/j.jrp.2013.01.012

Statement of Originality

I hereby declare that I am the sole author of this thesis and have made use of no other sources than those cited in this work.

Göttingen, September 13th 2019

Anna Dorothee Mischkowski