The Myth of the Extra Mile:

Psychological Processes and Neural Mechanisms Underlying Overcompensation Effects

Word count: 17,650 (without the Appendices)

Abstract

Trust violations regularly occur under the form of distributive fairness violations. In response to such violations, the transgressor can signal his or her willingness to go the "extra mile" by compensating the victim beyond the inflicted damage, which is generally referred to as *overcompensation*. We conducted two behavioral studies (Studies 1 and 2) and one fMRI experiment (Study 3) to investigate the psychological processes and supporting neural systems that underlie the effectiveness of overcompensation as a strategy to enhance trust in interpersonal relationships. Towards this end, we investigated how people on the receiving end of the compensation experience being overcompensated. Our studies, first of all, revealed that after being overcompensated people did not report higher levels of trust in the transgressor than after being equally compensated, a finding that runs counter the "extra mile" logic. As expected, our behavioral findings additionally showed that, compared to equal compensation, overcompensation evoked more conflicting thoughts and more sense-making processes in the mind of the receiver. Converging evidence for these findings was provided by our neuroimaging results, which revealed higher activations in the conflict-monitoring and the mentalizing network of the brain after overcompensation compared to equal compensation. Finally, the results of our behavioral studies suggest that conflicting thoughts and sense-making serially mediate the effect that overcompensation has on trust perceptions. Together, these findings shed new light on *why* overcompensation can backfire and even lead to a further decline of trust. We discuss the theoretical implications of these findings and formulate suggestions for future research.

***Keywords*:** social neuroscience; overcompensation; conflicting thoughts; sense-making; trust perceptions; serial mediation

1. INTRODUCTION

Trust is a ubiquitous part of social life. In fact, almost any decision or exchange that a person engages in includes some sort of trust evaluation, which emphasizes the notion that trust represents a necessary ingredient to coordinate and facilitate social life (Bohnet & Croson, 2004; Mayer, Davis, & Schoorman, 2007). Although a wide range of trust definitions exists, a common understanding has grown that "trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rousseau, Sitkin, Burt, & Camerer, 1998, p. 395). The presence of trust has been shown to offer numerous benefits. At the level of the individual trust has, for instance, been linked to love and happiness in close relationships (Rempel, Holmes, & Zanna, 1985). Moreover, trust has also been identified as a key trademark of effective organizations, as it fosters cooperation and increases performance (Bromiley & Cummings, 1996; Dirks, 2000; Dirks & Ferrin, 2001, 2002).

Unfortunately, prior studies have also shown that trust is fragile and that people's everyday actions and decisions offer numerous opportunities to violate trust (see Bottom, Daniels, Gibson, & Murnighan, 2002; Kim, Dirks, Cooper, & Ferrin, 2006; Kim, Ferrin, Cooper, & Dirks, 2004; Schweitzer, Hershey, & Bradlow, 2006). Such violations emerge when the victim's positive expectations about the transgressor are disconfirmed, like in cases of romantic betrayal or when a friend does not repay a loan. The present research focusses on one particular type of violation; namely, unfair resource distributions. Prior research has shown that, when resources have to be divided between two or more parties, people generally prefer allocations to be divided in an equal way (Camerer & Thaler, 1995; Handgraaf, Van Dijk, & De Cremer, 2003; Van Dijk, De Cremer, & Handgraaf, 2004), and that a violation of this equality norm negatively impacts people's trust (Desmet, De Cremer, & van Dijk, 2010, 2011; Haesevoets, De Cremer, Van Hiel, & Van Overwalle, 2018). Given the many benefits

of trust, it is of great importance to better understand how damaged trust can be restored. The unequal division of financial means between two or more parties and the trust restoration process in such a situation is exactly what we investigate here.

1.1 Going the Extra Mile: Overcompensation as a Means to Enhance Trust

Which strategy can be expected to be most effective to enhance trust perceptions in the aftermath of an unequal resource distribution? Because an equality violation negatively affects tangible outcomes, it has been suggested that the disadvantaged party may look for something "substantive" before being willing to trust the transgressing party again (see Bottom et al., 2002; Sitkin & Roth, 1993). The idea behind this proposition is that a substantial response signals that the transgressor is willing to invest—even in tangible ways—in his or her relationship with the victim. That is, tangible responses are expected to lessen concerns about "cheap talk" and the ensuing limitations of mere words to enhance trust after a transgression (see Dirks, Kim, Ferrin, & Cooper, 2011; also see Bottom et al., 2002).

Recognizing the potential value of such tangible efforts, researchers have started to investigate the effectiveness of financial compensation as a means to enhance trust (e.g., Darley & Pittman, 2003; Desmet et al., 2010, 2011; Okimoto, 2008). But, how large should a financial compensation be in order to effectively increase trust? According to the retributive justice literature, compensation must be according to the reciprocal economic logic of identical exchange—that is, "an eye for an eye" or exact compensation of the inflicted harm (Fish, 2008; Miller, 2005). This logic assumes that the monetary reimbursement should thus be restricted to the value of the experienced loss (i.e., "*only* one eye for one eye"). As such, after a distributive fairness violation the provision of a compensation which exactly covers the inflicted harm (i.e., *equal compensation*) can be considered to be the norm.

It is, however, possible that the act of restoring trust asks more from the transgressor than simply restoring the outcome situation that the victim already expected to get in the first

place (i.e., an equal share of the resources). In other words, in the aftermath of an unequal resource distribution the process of restoring trust may call for a different strategy than building trust initially, requiring efforts that are larger in magnitude than those needed for initial trust development (Desmet et al., 2011; Kim et al., 2004). In this vein, Dirks and colleagues (2011, p. 88) aptly noted that, in contrast to building trust from scratch, "the repair of trust involves the additional challenge of overcoming the salient negative expectations that arose from the transgression." This argument is supported by the conventional wisdom that, if one aims to make up for wrongdoing, it is important to act generously. According to this perspective, it can thus be argued that, to effectively restore damaged trust, the transgressor should signal that the violation will not happen again and that he or she is willing to invest in this commitment. One way in which the transgressor can signal such assurance is by going the "extra mile" and thus provide the victim a reimbursement that exceeds the inflicted harm. This type of compensation that goes beyond the inflicted harm is generally referred to as *overcompensation* (Desmet et al., 2011; Haesevoets, Van Hiel, Reinders Folmer, & De Cremer, 2014; Haesevoets, Reinders Folmer, & Van Hiel, 2014).

Overcompensation regularly occurs in everyday life. For instance, after missing a connection flight due to delay the responsible airline company can provide customers a refund of expenses plus an additional free airline ticket (Boshoff, 1997). Similarly, in case of hotel overbooking, customers can be offered the finest suite of the hotel or, if no other room is available, a voucher- or cash-based compensation that is worth more than the original room price (Noone & Lee, 2011). And when a restaurant serves an incorrect dish, it can replace the dish and additionally offer the customer the meal for free (Hocutt, Bowers, & Donavan, 2006). Although the use of overcompensation has most prevalently been documented in customer-company relationships, such behaviors also regularly occur in interpersonal settings. For instance, when someone damages your newly purchased book (e.g., by spilling a can of

Coke or a cup of coffee on it), this person can undo the damage by buying you a new book plus giving you a little extra (Haesevoets, Van Hiel, et al., 2014).

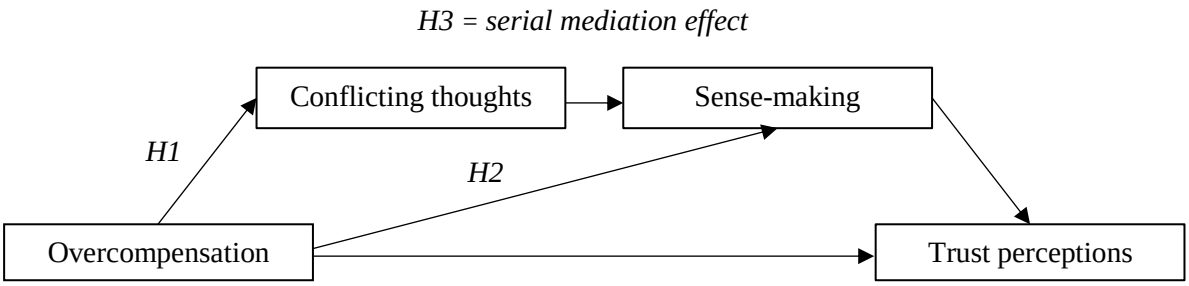
1.2 Theoretical Development and Hypotheses

Overcompensation can thus intuitively be expected to be a more effective strategy to enhance trust than equal compensation. Yet, although some prior studies have reported positive overcompensation effects (e.g., Boshoff, 1997; Gilly & Hansen, 1985; Hocutt et al., 2006), there seems to be a growing consensus, within various literatures, that overcompensation may not be more—and sometimes even less—effective than equal compensation to enhance trust and various trust-related outcomes (for a field study, see Goode, Hoehle, Venkatesh, & Brown, 2017; for a meta-analysis, see Gelbrich & Roschk, 2011; for a narrative review, see Davidow, 2003; also see the empirical work of Estelami & De Maeyer, 2002; Garrett, 1997; Haesevoets, Van Hiel, et al., 2014; Mack, Mueller, Crotts, & Broderick, 2000; Noone & Lee, 2011). Critically, however, is that these prior studies did not investigate the psychological mechanisms which are responsible for the reported lack of positive overcompensation effect. In other words, based on the current state of the literature it is unclear *why* financial overcompensation is not a more effective strategy than equal compensation to promote damaged trust.

The present research aims to provide deep insights into this "why" question by adopting a multi-method approach that includes both behavioral and neuroimaging methods, whereas most prior overcompensation studies exclusively relied on behavioral data. By integrating the psychological processes—and supporting neural systems—that people use when evaluating financial overcompensation, our research goes beyond what we might learn separately from either the social psychology or the neuroscience literatures. Instead, we capitalize on these two disciplines to elucidate the core processes by which people react to being overcompensated. More specifically, we posit that, compared to equal compensation,

overcompensation elicits more conflicting thoughts (*Hypothesis 1*) and more sense-making processes (*Hypothesis 2*) in the mind of the person who receives the compensation. We apply functional brain imaging in order to investigate if these assumptions are also supported by neurophysiological data. In this vein, we expect conflict and sense-making to be accompanied by increased activation in the conflict-monitoring network and the mentalizing network of the brain, respectively. Moreover, we also investigate how these processes relate to people's trust perceptions. In this regard, we predict that conflicting thoughts and sense-making serially mediate the effect that overcompensation has on trust perceptions (*Hypothesis 3*). The theoretical underpinnings of these predictions are explained in greater detail below. Figure 1 shows our conceptual model.

FIGURE 1. Conceptual Model of our Study.



1.2.1 Overcompensation Induces Conflicting Thoughts

Financial overcompensation is expected to elicit psychological conflicting associations in the mind of the person who receives the compensation. The basic tenet behind this prediction is that overcompensation is expected to simultaneously trigger both positive and negative thoughts. The positive thoughts relate to the egoism-based associations of receiving an outcome that clearly serves one's self-interest, whereas the negative thoughts concern associations relating to being unfairly advantaged (Adams, 1965; Jacques, 1961; Peters, Van den Bos, & Bobocel, 2004; Peters, Van den Bos, & Karremans, 2008). In other words, we expect that people will experience conflicting thoughts—that is, both positive and negative

thoughts—after being overcompensated. Equal compensation, on the other hand, results in a fair outcome for everyone, and also serves the recipient's self-interest (although to a lesser extent than overcompensation). After receiving equal compensation, there should thus be much more alignment between one's own interests and fairness concerns (Adams, 1965; Buunk & Van Yperen, 1989). Accordingly, we expect that after equal compensation mostly positive thoughts will be elicited, and no or only weak negative thoughts (cf. Van den Bos, Peters, Bobocel, & Ybema, 2006). Taken these prior arguments together, we thus expect that conflict levels will be higher after being overcompensated than after being equally compensated.

1.2.1.1 Neural predictions. How then does the brain react to being overcompensated? We expect that the higher levels of conflicting thoughts after receiving overcompensation will be supported by increased activation in the conflict-monitoring network of the brain, which detects and resolves conflicts between multiple inputs (Botvinick, Cohen, & Carter, 2004). Prior neuroimaging research has revealed that the conflict-monitoring network consists of the posterior medial frontal cortex (pmFC, including the dorsal part of the anterior cingulate cortex) and the lateral prefrontal cortex. The pmFC detects conflicts, whereas the lateral prefrontal cortex resolves these conflicts by modifying the attention to the different conflicting inputs (Botvinick et al., 2004). In addition to these conflict regions, brain areas that communicate with the conflict-monitoring system are also expected to show increased activation. Specifically, the bilateral insula—which is part of a larger system that integrates interoceptive information and important environmental inputs with conflict processing (Menon & Uddin, 2010; Taylor, Seminowicz, & Davis, 2009)—and the caudate nucleus—which is part of a network that regulates reward experiences (Shenhav, Botvinick & Cohen, 2013)—are also expected to show increased activation after being overcompensated. Based on the reasoning presented above, we can formulate the following hypothesis:

Hypothesis 1: People experience higher levels of conflicting thoughts (which is expected to be supported by increased activation in the conflict-monitoring network of the brain) after being overcompensated than after being equally compensated.

1.2.2 Overcompensation Evokes Sense-Making

As mentioned above, overcompensated people are expected to experience conflicting thoughts because the compensation automatically triggers both positive and negative thoughts, which basically represents an ambiguous situation. This ambiguity makes it difficult for the overcompensated person to know how to respond to the situation (Van den Bos et al., 2011; also see Van den Bos, Lind, Vermunt, & Wilke, 1997; Van den Bos, Wilke, Lind, & Vermunt, 1998). In line with this reasoning, empirical research has shown that experiences of ambiguity activate an information process in which people aim to make sense of the situation at hand (Van Hiel & Mervielde, 2002; Nohlen, Van Harreveld, Rotteveel, Barends, & Larsen, 2016). More specifically, this line of research suggests that ambiguity is an inherent motivator to search for additional information that can help to disambiguate the situation.

Critically, in the context of our research question, we expect that the conflicting thoughts that overcompensated people experience will stimulate them to infer the inner thoughts and intentions of the person who overcompensated them. More specifically, we expect that after being overcompensated people will try to make sense of this situation by figuring out what exactly motivated the provider of the compensation to reimburse them so generously. Several researchers have indeed argued that ambiguity is an important "occasion for sensemaking" (Maitlis, 2005, p. 21; Miller, Joseph, & Apker, 2000, p. 198; also see Brockner, 2002, Weick, 1995). In line with this reasoning, Steinel, Van Beest, and Van Dijk (2014) have found that, in negotiation situations, people indeed try to make sense of ambiguous situations (i.e., high offers that are beneficial to themselves, but not to their bargaining opponent). Equal compensation, on the other hand, represents low ambiguity as it

predominately activates positive thoughts. Prior research has shown that sense-making processes are indeed less likely to be activated under low ambiguity (Steinel et al., 2014; Van Hiel & Mervielde, 2002). As such, we expect that people will also engage in higher levels of sense-making after being overcompensated than after being equally compensated.

1.2.2.1 Neural predictions. Which neural systems can be expected to be associated with this sense-making process? Overcompensated people will try to make sense of the situation at hand, and therefore engage in a so-called mentalizing process (Frith & Frith, 2003). Mentalizing can be defined as the ability to understand the mental state of another person, and to use this information to predict and explain this other person's behavior (see Fletcher et al., 1995; Frith, Morton, & Leslie, 1991). Compared to equal compensation, overcompensation is expected to lead to increased activation in three well-defined areas of the mentalizing network, those being, the precuneus, the temporoparietal junction (TPJ), and the medial prefrontal cortex (mPFC; see Amodio & Frith, 2006; Mitchell, 2006; Saxe, 2006). Scholars have proposed that the TPJ is responsible for transient mental inferences about other people's intentions, while the mPFC subserves the attribution of more enduring traits and qualities about oneself and other people (Van Overwalle, 2009; Van Overwalle & Baetens, 2009). Based on the above presented reasoning, we formulate the following prediction:

> *Hypothesis 2:* People engage in higher levels of sense-making (which is expected to be supported by increased activation in the mentalizing network of the brain) after being overcompensated than after being equally compensated.

1.2.3 Conflicting Thoughts and Sense-Making as Serial Mediators

But how then do conflicting thoughts and sense-making relate to trust perceptions? Our conceptual model (see Figure 1) posits that after being overcompensated people first experience a conflict between what they want and what they believe is right. Furthermore, our model proposes that, only after this first conflictual experience, in a subsequent step, people

will engage in a sense-making process. The reason why people engage in such a sense-making process is that this inference can help them to figure out the intentions of the person who overcompensated them. The assumed sequence in our model is thus first: "Wow, I got a really good outcome, but that is not fair to the other" (conflicting thoughts), followed by: "Why is it that this person compensated me so generously?" (sense-making). This latter process may lead the recipient of the compensation to conclude that the overcompensation is just too generous (Steinel et al., 2014; Estelami & De Maeyer, 2002).

In a final step of our model, we therefore expect that, as an important downstream consequence, the sense-making process may negatively influence the recipient's assessment of the trustworthiness of the provider of the overcompensation ("I do not trust this person"). This is particularly the case because being unfairly advantaged has been related in previous research not only to conflicting thoughts and sense-making, but eventually, as the end-product of the process, to negative assessments of and judgements about this person (cf. Steinel et al., 2014; Peters et al., 2004, 2008); and such negative assessments and judgements are incompatible with high levels of trust. Taken the above provided arguments together, we formulate the following serial mediation hypothesis:

> *Hypothesis 3:* Conflicting thoughts and sense-making serially mediate the overcompensation effect on trust: Overcompensation induces conflicting thoughts, these conflicting thoughts subsequently result in a sense-making process, and this sense-making process, in turn, negatively affects perceptions of trustworthiness.

1.3. The present studies

To test our hypotheses, we conducted three studies. Study 1 set out to test if overcompensation indeed results in higher levels of conflict and sense-making than equal compensation. Additionally, this study also investigated if conflicting thoughts and sense-making constitute the predicted serial mediation chain. To induce a distributive fairness

violation, participants in this study received an unequal division, which clearly violates the allocation norm of equality (Deutsch, 1975). We specifically focused on violations of the equality norm as equality presents the clearest norm, and transgressions of this norm can be interpreted straightforwardly (see Haesevoets et al., 2018, for a similar argument). To make up for the inflicted harm, the transgressor subsequently provided the participants either an equal compensation or an overcompensation. Study 2 used a similar setup, but this time all participants were overcompensated. The main aim of our second study was to further investigate the role that conflictual thoughts and sense-making play in the relationship between overcompensation and perceptions of trustworthiness. Because within the literature trust is often considered to be a multidimensional concept that consists of an affective and a cognitive dimension (Lewis & Weigert, 1985; McAllister, 1995), this second study tested our conceptual model for both these trust dimensions.

To investigate the processes that underlie the impact of financial overcompensation on trust, in our first two studies (which are both behavioral studies) we explicitly asked participants to indicate the extent to which they experienced conflicting thoughts and engaged in a sense-making process after being overcompensated. However, by explicitly asking to what extent these two processes played a role in participants' assessment of the compensation, we actually made these two processes salient. To investigate if these two processes are also *spontaneously* activated in people's brain after receiving overcompensation, Study 3 consisted of a functional magnetic resonance imaging (fMRI) experiment. A specific advantage of neural measures in this light is that they can be assessed online during a psychological task, without having to interrupt the flow of a task to have the participant complete a self-report measure (Amodio, 2010). Another important advantage of this technique is that the use of fMRI allows the collection of biological data which are less susceptible to demand characteristics and other biases that plague self-report data, including response biases

associated with self-presentational and social desirability concerns. In light of our research question, it is, for instance, possible that, when being asked, participants feel that they "should" indicate that they dislike overcompensation, when in fact they like being overcompensated (for a similar reasoning, see Fliessbach et al., 2012). Moreover, neuroimaging methods have been proven to be particularly useful to probe into underlying processes that are difficult to measure with self-reports methods (cf. Dulebohn, Conlon, Sarinopoulos, Davison, & McNamara, 2009; Haesevoets et al., 2018), and as such they also constitute an important source of cross-validation.

All manipulations and exclusions in the studies are disclosed, as well as the method of determining the different sample sizes. The data and data analysis scripts of Studies 1 and 2 and the behavioral data of Study 3 are made publicly available, and can be accessed through Open Science Framework: weblink.

2. STUDY 1

The aim of our first study was twofold. The first aim of Study 1 was to test our assumption that overcompensation results in more conflicting thoughts (*Hypothesis 1*) and more sense-making (*Hypotheses 2*) than equal compensation. Moreover, we also aimed to illustrate the predicted serial-mediated relationship between overcompensation and trust perceptions via conflicting thoughts and sense-making processes (*Hypothesis 3*). Towards this end, we designed an experiment that contained one between-subjects variable with two conditions (equal compensation vs. overcompensation).

2.1. Method

2.1.1 Sample and Design. A priori power calculations with the R package Pwr (Champely, 2018) revealed that, under standard criteria ($\alpha = .05$, $1 - \beta = .80$), we needed at least 63 participants per condition in order to be able to detect such a medium effect size. Due to an administration error, only 122 (instead of 126) participants were solicited using Prolific

(www.prolific.co). We explicitly recruited working adults living in the United Kingdom. Participants received £1 for their participation plus a small bonus amount that depended upon their outcome during the experiment. Twenty-one participants had to be excluded from the analyses; the criteria for removal are outlined below. The final sample thus consisted of 101 participants (48 were male participants; M_{age} = 35.57, SD = 9.73), 52 of them were randomly assigned to the equal compensation condition (30 male participants; M_{age} = 36.19, SD = 9.22) and 49 to the overcompensation condition (18 male participants; M_{age}= 34.92, SD = 10.30).

A sensitivity analysis revealed that, given this final sample size, we had 80% power to detect an effect of size Cohen's d = 0.57. Since prior studies comparing the impact of equal compensation and overcompensation on trust perceptions found medium to large effect sizes (e.g., Cohen's d = 0.87 and 0.71, in respectively Study 3 and Study 4 of Haesevoets, Van Hiel, et al., 2014; and Cohen's d = 1.13 in Haesevoets, Reinders Folmer, et al., 2014), it can be concluded that the present study was properly powered to detect an effect of similar magnitude or an even somewhat smaller effect.

2.1.2 Procedure. At the start of the experiment, participants were informed that during the study they would be connected with two other Prolific participants who were simultaneously conducting the study. In reality, however, these two other participants and their behaviors were simulated. Participants were told that one participant would be assigned the role of allocator, whereas the other two participants would act as recipients. Subsequently, we informed participants that a network connection would be established between them and the other two participants. The participants then saw a screen that visually illustrated, by means of a line of running dots, that the server was connecting the participating individuals.

Once the network connection was established, participants were informed about the role that they would play during the study. Unbeknownst to them, all participants were told that one of the other two participants was assigned the role of allocator, and that they and the

remaining participant would engage in the role of recipient. Participants were then informed that they had to complete several tasks during the study. They were told that, in order to be able to carry out these tasks successfully, they needed a work budget. Participants were also informed that at the end of the experiment they would be paid in accordance with their earnings in these tasks.

It was then communicated to participants that the person who was assigned the role of allocator would soon divide the available work budget—that consisted of 100 units (1 unit being worth £0.01)—between them and the other participant. After participants had waited for some time, they received an email message in which the allocator informed them that he or she had decided to allocate only 20 units to them and the remaining 80 units to the other participant.

Participants were subsequently given the opportunity to let the allocator know how they perceived this unequal division. Towards this end, participants had to choose between two messages: The first message stated that they perceived the division as being fair, whereas the second message stated that they perceived the division as being unfair. Eleven participants selected the first message, indicating that they perceived the unequal division as fair, and were therefore excluded from the reported analyses.[1]

Participants subsequently received another email message from the allocator, which stated that he or she had decided to reallocate the work budget. This reallocation constituted our compensation manipulation. More specifically, in the *equal compensation* condition, the message from the allocator stated that he or she had decided to allocate an additional 30 units to the participant. As a result of this redistribution, in the equal compensation condition the

[1] We have reanalyzed the data of Study 1 including participants who indicated that they perceived the allocator's unequal division as fair. The inclusion of these participants did not change any of our key results. That is, in line with *Hypotheses 1* and *2*, the results of a *t*-tests showed that, compared to the equal compensation condition, participants in the overcompensation condition still scored significantly higher on the conflicting thoughts (M = 5.29, SD = 1.01 vs. M = 3.67, SD = 1.46, $t(109) = -6.70$, $p < .001$) and the sense-making scale (M = 5.64, SD = 1.15 vs. M = 4.05, SD = 1.79, $t(109) = -5.50$, $p < .001$). Moreover, in agreement with *Hypothesis 3*, the results of a serial mediation analysis showed that the indirect effect of the compensation manipulation on trust perceptions was still serially mediated by conflicting thoughts and sense-making (indirect effect: B = -0.41, boot SE = .15, 95% CI: [-.735, -.165]).

participant and the other person both received 50 units. In the *overcompensation* condition, the message from the allocator stated that he or she decided to allocate an additional 70 units to the participant. As a result of this, in the overcompensation condition the participant ended up with 90 units whereas the other person ended up with solely 10 units.

Participants were then asked some questions about how they experienced the allocator's redistribution of the work budget (see measures). After responding to these measures, the experiment was interrupted, ostensibly because of an error in the established connection with the other two participants. Participants were thanked for their efforts and informed that they would be paid in accordance with how the allocator had redistributed the work budget. Accordingly, participants received a bonus payment of £0.50 in the equal compensation condition and a bonus payment of £0.90 in the overcompensation condition.

2.1.3 Measures

2.1.3.1 Conflicting Thoughts. We first measured the extent to which participants experienced conflicting thoughts, using a self-developed six-item scale (based on items taken from various ambivalence scales; see Jamieson, 1993; Priester & Petty, 1996; also see Guarana & Hernandez, 2016; Rothman, 2011). A sample item is: "The way in which the allocator has redistributed the work budget makes me think about this whole situation in conflicting ways." The six items were all measured on a seven-point Likert scale (1 = *totally disagree*, 7 = *totally agree*), and aggregated into a general conflicting thoughts scale (Cronbach's alpha = .89). The full item list can be found in Appendix A.

To test more directly if equal compensation is simultaneously perceived as satisfying and fair and overcompensation as satisfying but unfair, we additionally also asked participants (1 = *not at all*, 7 = *very much so*): "To what extent do you consider the received compensation as being satisfying for your own self-interest?" (satisfaction item) and "To what extent do you

consider the received compensation as being fair for you and the other person?" (fairness item).

2.1.3.2 Sense-Making. Next, we measured the extent in which the provided compensation makes participants think about the intentions underlying the allocator's behavior, by means of a self-developed scale that also consists of six items (which are all based on the definition of the mentalizing concept that we discussed earlier; Fletcher et al., 1995; Frith et al., 1991). A sample item is: "The way in which the allocator has redistributed the work budget makes me question what has motivated this person to make this particular redistribution" (1 = *totally disagree*, 7 = *totally agree;* Cronbach's alpha = 96). The complete item list is included in Appendix A.

2.1.3.3 Trust Perceptions. After rating the conflict and sense-making items, participants were asked to rate their trust in the allocator. Towards this end, we used as seven-item trust scale (of which the items are based on the trust scales of Mayer & Davis, 1999; Shockley-Zalaba, Ellis, & Winograd, 2000; Spreitzer & Mishra, 1999). A sample item is: "To what extent do you think you can trust this person to be completely honest with you?" (1 = *not at all*, 7 = *very much so;* Cronbach's alpha = .96; see Appendix A for the full item list).

2.1.3.4 Manipulation Checks. We checked the effectiveness of our compensation manipulation with the following two check questions: "To what extent did the redistribution of the allocator result in an equal work budget for you and the other person?" (check 1) and "To what extent did the redistribution of the allocator result in a much larger work budget for you than for the other person?" (check 2). Both these check questions were administered at the end of the study, and rated on a seven-point Likert scale (1 = *not at all*, 7 = *very much so*).

2.1.3.5 Additional Selection Criteria. To exclude participants who answered with insufficient care, during the study we employed an attention check (i.e., "please select the first response box for this question;" based on Meade & Craig, 2012). Four participants failed this

check, and, as a result, their responses were excluded from the analyses. To ensure that participants indeed perceived the overcompensation as being larger than the inflicted damage, in addition to our two manipulation checks, at the end of the study we also asked participants in the overcompensation condition: "According to you, the number of additional units that you received from the allocator …. the harm that was inflicted by his or her initial unequal division of the work budget" (1 = *is smaller than*, 2 = *exactly covers*, 3 = *is larger than*). Six participants indicated that the overcompensation was smaller than or equal to the inflicted harm, and were therefore also excluded from the analyses.

2.2 Results

Table 1 reports the means, standard deviations, and intercorrelations among the scale measures.

TABLE 1. Means, Standard Deviations, and Intercorrelations (Pearson's *r*) among the Scale Measures Used in Study 1

Variable	*M*	*SD*	1.	2.
1. Conflicting thoughts	4.45	1.53		
2. Sense-making	4.79	1.74	.74***	
3. Trust perceptions	3.28	1.31	-.49***	-.58***

Note. *** $p < .001$.

2.2.1 Manipulation Checks. To test if our compensation manipulation was successful, we conducted a *t*-test on the two manipulation checks. The results of this test revealed that participants' scores on both check questions differed significantly between the equal compensation (check 1: $M = 6.87$, $SD = 0.53$ and check 2: $M = 1.54$, $SD = 1.54$) and the overcompensation condition (check 1: $M = 1.08$, $SD = 0.28$ and check 2: $M = 6.90$, $SD = 0.31$), $t(99) = 68.64$, $p < .001$, $d = 13.66$ and $t(99) = -23.92$, $p < .001$, $d = 4.83$, respectively.

2.2.2 Measurement Model. As shown in Table 1, the conflicting thoughts and the sense-making scale were highly correlated. To verify the independent nature of these constructs, we conducted a Confirmatory Factor Analysis (CFA). Toward this end, we constructed a three-factor model in which the scale items of each construct (conflicting thoughts, sense-making, and trust perceptions) loaded on a separate factor. The results of this analysis, which are summarized in Table B.1 of Appendix B, show that our three-factor model had a better fit than all alternative two-factor models and a one-factor model in which all items loaded onto the same factor, as such providing statistical evidence for the distinctiveness of our constructs.

2.2.3 Hypothesis Testing

2.2.3.1 Conflicting Thoughts. We first investigated if overcompensation induces more conflictual thoughts than equal compensation. In agreement with *Hypothesis 1*, the results of a *t*-test revealed that participants in the overcompensation condition scored significantly higher on the conflicting thoughts scale ($M = 5.31$, $SD = 1.02$) than those in the equal compensation condition ($M = 3.65$, $SD = 1.51$), $t(99) = -6.42$, $p < .001$, $d = 1.19$.

We subsequently conducted a one-sample *t*-test to investigate if participants in the equal compensation condition perceived the compensation as both satisfying and fair (two positive sources) and if participants in the overcompensation condition perceived the compensation as satisfying but unfair (one positive and one negative source). As expected, in the equal compensation condition, participants' scores on both the satisfaction and the fairness item were significantly higher than the scale's midpoint ($M = 6.06$, $SD = 1.31$ and $M = 6.50$, $SD = 1.06$, respectively), both ts > 11.35, ps < .001. In the overcompensation condition, participants' scores on the satisfaction item were also significantly higher than the scale's midpoint ($M = 5.49$, $SD = 1.95$), $t(48) = 5.35$, $p < .001$; conversely, on the fairness item, participants' scores were significantly lower than the scale's midpoint ($M = 1.31$, $SD =$

0.55), $t(48) = -34.42$, $p < .001$. As such, these findings illustrate that equal compensation is indeed perceived as both satisfying and fair, whereas overcompensation is perceived as satisfying but unfair.

2.2.3.2 Sense-Making. Next, we tested if overcompensated participants also engaged in more sense-making than those who were equally compensated. In support of *Hypothesis 2*, a *t*-test revealed that participants in the overcompensation condition scored significantly higher on the sense-making scale ($M = 5.65$, $SD = 1.14$) than participants in the equal compensation condition ($M = 3.98$, $SD = 1.83$), $t(99) = -5.45$, $p < .001$, $d = 1.10$.

2.2.3.3 Trust Perceptions. We conducted an additional *t*-test to investigate how our compensation manipulation affected participants' trust perceptions. The results of this test revealed that participants in the overcompensation condition scored significantly lower on the trust scale ($M = 2.51$, $SD = 1.00$) than those in the equal compensation condition ($M = 4.00$, $SD = 1.16$), $t(99) = 6.94$, $p < .001$, $d = 1.38$. Note that this finding is consistent with a large body of prior research reporting that overcompensation does not enhance trust beyond equal compensation, and even can have adverse effects.

2.2.3.4 Mediation Analysis. We subsequently tested our predicted serial mediation chain using Model 6 (serial mediation) of the PROCESS macro (Hayes, 2013). The advantage of this procedure is that it enables isolation of each mediator's indirect effect. Moreover, this approach also allows the investigation of "the indirect effect passing through both of these mediators in a series" (Van Jaarsveld, Walker, & Skarlicki, 2010, p. 1496). We used 5,000 bootstrap samples to estimate 95% bias-corrected confidence intervals. The compensation manipulation (X) was entered as the predictor, conflicting thoughts (M1) and sense-making (M2) were entered as the two mediators, and trust perceptions (Y) were entered as the outcome. Gender and age were controlled for as covariates. Model 1 tested if conflicting thoughts *alone* mediated the relationship between the compensation manipulation and trust

perceptions (X->M1->Y); Model 2 tested if sense-making *alone* mediated the relationship between the compensation manipulation and trust perceptions (X->M2->Y); and Model 3 tested if conflicting thoughts and sense-making *in a series* mediated the relationship between the compensation manipulation and trust perceptions (X->M1->M2->Y).

The results of this serial mediation analysis are presented in Figure 2; the indirect effects are summarized in Table 2. We found a significant direct effect of our compensation manipulation on trust perceptions (direct effect of X on Y: B = -1.03, SE = .24, 95% CI [-1.513, -.545]). Moreover, the total indirect effect of our compensation manipulation on trust perceptions was also significant (total indirect effect of X on Y: B = -0.47, boot SE = .18, 95% CI [-.878, -.158]). However, the indirect effect via conflicting thoughts alone (Model 1) was not significant (indirect effect of X on Y via M1: B = 0.05, boot SE = .18, 95% CI = [-.317, .403]). Similarly, the indirect effect via sense-making alone (Model 2) was also not significant (indirect effect of X on Y via M2: B = -0.12, boot SE = .13, 95% CI = [-.428, .075]). Most importantly, in line with our predictions made in *Hypothesis 3*, the indirect effect via both conflicting thoughts and sense-making in a series (Model 3) turned out to be significant (indirect effect of X on Y via M1 and M2: B = -0.40, boot SE = .15, 95% CI = [-.739, -.144]).

2.2.3.5 Supplementary Mediation Analysis. Although our predicted serial mediation hypothesis was supported by the data, the possibility remains that reverse causality exists. We therefore additionally tested an alternative model in which our compensation manipulation predicted trust perceptions via sense-making and conflicting thoughts (inverted causal order of the two serial mediators). The results of this analysis revealed that the indirect effect of our compensation manipulation on trust perceptions via sense-making and conflicting thoughts in a series was not significant (indirect effect of X on Y via M2 and M1: B = 0.03, boot SE = .10, 95% CI = [-.170, .224]). These findings hence suggest that it is more likely that

overcompensation induces conflicting thoughts followed by sense-making processes than that overcompensation evokes sense-making processes followed by conflicting thoughts.

FIGURE 2. Results of the Serial Mediational Analysis of Study 1. Values shown are unstandardized coefficients (with *SE* in parentheses).

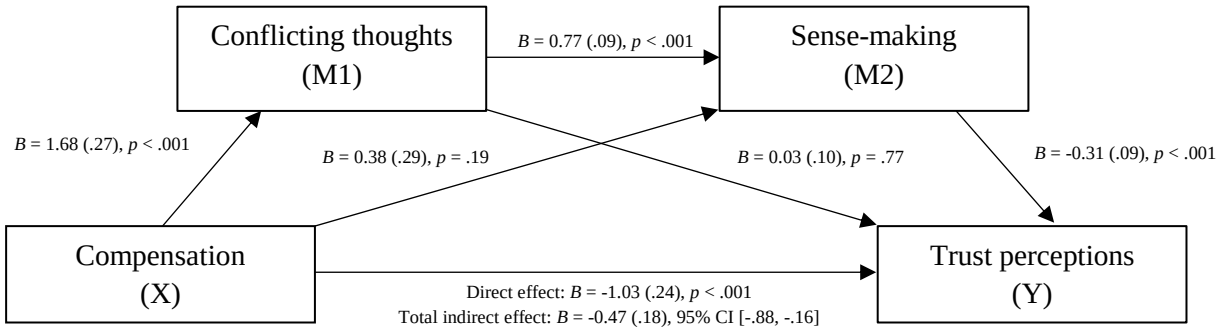


TABLE 2. Indirect Effects of the Compensation Manipulation on Trust Perceptions via Conflicting Thoughts and Sense-Making for Study 1.

Path	*B* (boot *SE*)	LL	UL
Model 1: Compensation -> Conflict -> Trust	0.05 (.18)	-.317	.403
Model 2: Compensation -> Sense-making -> Trust	-0.12 (.13)	-.428	.075
Model 3: Compensation -> Conflict -> Sense-making -> Trust	-0.40 (.15)	-.739	-.144
Total indirect effect	-0.47 (.18)	-.878	-.158

Note. LL = lower limit of 95% CI, UL = upper limit of 95% CI.

2.3 Discussion

In line with our predictions, the results of this first behavioral study indicate that overcompensation induces higher levels of conflicting thoughts (*Hypothesis 1*) and evokes higher levels of sense-making (*Hypothesis 2*) than equal compensation. Moreover, the data of the present study also provide cross-sectional evidence suggesting that conflictual thoughts and sense-making form the predicted serial mediation chain in the relationship between overcompensation and trust perceptions (*Hypothesis 3*).

3. STUDY 2

Study 2 aimed to deepen our understanding of the role that conflicting thoughts and sense-making play in the relationship between overcompensation and trust (cf. *Hypothesis 3*). A specific limitation of Study 1 is that the conflicting thoughts and the sense-making items were all formulated rather broadly. We did this in order to ensure that these items were applicable in both the equal compensation and the overcompensation condition. The present study used the exact same procedure as our prior study, except that we did not include an equal compensation condition. Because of this, in the present study we were able to formulate the conflicting thoughts and the sense-making items more specifically; that is, with a specific reference to the provided overcompensation.

For instance, a sample item of the conflicting thoughts scale that we used in Study 1 was: "The way in which the allocator has redistributed the work budget makes me think about this whole situation in conflicting ways." In the present study, we used the same item, but at the end we added: "Because I don't like it to be advantaged over others." Similarly, to measure the extent in which participants engaged in sense-making processes, in Study 1 we asked participants: "The way in which the allocator has redistributed the work budget makes me question what has motivated this person." In the present study, we added: "To allocate such a large part of the work budget to me."

Another potential limitation of Study 1 is that this study employed a one-dimensional measure of trust. However, within the literature trust is generally positioned as a multi-dimensional construct that consists of both an affective and a cognitive dimension (see the conceptual model of Lewis & Weigert, 1985; also see the empirical work of Cummings & Bromiley, 1996). The cognitive dimension of trust is rooted in rationality and competence, whereas the affective trust dimension is based on relationships and feelings (McAllister, 1995). Within this framework, cognition-based trust is generally seen as a less deep and more superficial type of trust than trust based on affect (see Dietz & Hartog, 2006; McAllister,

1995; Ranganathan, Madupu, Sen, & Brooks 2013). Because of the supposed multifaceted nature of trust, in the present study we decided to measure both the affective and the cognitive trust dimension.

3.1 Method

3.1.1 Sample. Sample size calculations were based on the relationships obtained in Study 1 for the overcompensation condition. Simulations showed that, under standard criteria, and given the Study 1 correlations, a sample of 200 participants would yield more than 99% power to reproduce the effect of conflicting thoughts (through sense-making) on trust perceptions of size beta = -.23 obtained in Study 1. Anticipating some dropout, we initially recruited 210 working adults from the United Kingdom through Prolific. We employed the exact same exclusion criteria as in Study 1, which led to the exclusion of 31 participants (see below). The samples of Study 1 and Study 2 were independent and did not overlap. Participants again received £1 for their participation plus a fixed bonus amount which equaled their earning during the study. The final sample consisted of 179 participants (91 male participants; M_{age} = 35.21, *SD* = 9.62). We used the exact same overcompensation amount as in the previous study, which covered the inflicted financial harm approximately two times.

3.1.2 Procedure. In the present study, we used almost the exact same paradigm as we used in Study 1; with the only difference that in the present study all participants were overcompensated. At the beginning of the study, participants were again said to be linked to two other participants who were supposedly also conducting the study. As in the prior study, one of the other participants was said to be assigned the role of allocator. This (preprogrammed) allocator again gave 20 units to the participant and 80 units to the other person. Like in the prior study, participants were again asked to indicate whether they perceived this unequal division as fair or unfair. Twelve participants selected the message that they perceived this unequal division as being fair, and were therefore excluded from the

reported analyses.[2] All participants subsequently received an *overcompensation* from the allocator, who allocated an additional 70 units to the participant. This redistribution again resulted in a final 90-10 distribution in favor of the participant. After receiving this overcompensation, participants were asked to answer our measures (see below), after which the study was interrupted. In accordance with the allocator's redistribution of the work budget, in this study all participants received a bonus payment of £0.90.

3.1.3 Measures. Similar to Study 1, in the present study we again first administered the conflicting thoughts items, followed by the sense-making items, and finally the affective and cognitive trust items.

3.1.3.1 Conflicting Thoughts and Sense-Making. As mentioned above, conflicting thoughts and sense-making were measured with similar items as we used as in the prior study, but this time these items included a specific reference to the provided overcompensation. A sample item of the conflicting thoughts scale is: "The way in which the allocator has redistributed the work budget makes me think about this whole situation in conflicting ways, because I don't like it to be advantaged over others." A sample items of the sense-making scale is: "The way in which the allocator has redistributed the work budget makes me question what has motivated this person to allocate such a large part of the work budget to me." The items were all rated on a seven-point Likert scale (1 = *totally disagree,* 7 = *totally agree;* Cronbach's alpha = .82 and .90 for the conflicting thoughts and the sense-making scale, respectively). The full item list is included in Appendix C.

3.1.3.2 Affective Trust and Cognitive Trust. Participants' affective and cognitive trust in the allocator were both measured with a five-item scale of which the items are based on the trust scales of Yang, Mossholder, and Peng (2009). Sample items are: "To what extent do you

[2] The data of Study 2 were also reanalyzed including those participants who perceived the unequal division of the allocator as being fair. In line with our predictions made in *Hypothesis 3*, the results of these analyses showed that, for both trust components, the indirect effect of conflicting thoughts on trust through sense-making was still significant (indirect effect for affective trust: *B* = -0.11, boot *SE* = .07, 95% CI: [-.281, -.015]; indirect effect for cognitive trust: *B* = -0.09, boot *SE* = .05, 95% CI: [-.213, -.002]).

think that you can openly communicate your feelings to this person?" (affective trust) and "To what extent do you think that you can depend on this person to meet his or her responsibilities?" (cognitive trust). The items were all rated on a seven-point Likert scale (1 = *not at all*, 7 = *very much so;* Cronbach's alpha = .91 for both the affective and the cognitive trust scale). The full item list can be found in Appendix C.

3.1.3.3 Additional Selection Criteria. We included the same attention check as in Study 1, which led to the exclusion of an additional six participants. Moreover, we again asked participants how the received compensation relates to the total harm that was inflicted by the allocator's unequal division of the work budget. An additional 13 participants were excluded from the analyses because they did not perceive the overcompensation as being larger than the inflicted harm.

3.2 Results

Table 3 includes the means, standard deviations, and intercorrelations among the scale measures.

TABLE 3. Means, Standard Deviations, and Intercorrelations (Pearson's *r*) among the Scale Measures Used in Study 2.

Variable	*M*	*SD*	1.	2.	3.
1. Conflicting thoughts	6.05	0.88			
2. Sense-making	6.04	0.97	.51***		
3. Affective trust	2.93	1.26	-.14†	-.24**	
4. Cognitive trust	2.67	1.17	-.23**	-.26***	.62***

Note. † $p = .061$, ** $p < .01$, *** $p < .001$.

3.2.1 Measurement Model. Because some of our constructs were highly correlated (see Table 3), we conducted a Confirmatory Factor Analysis (CFA) to verify the independent nature of these constructs. We constructed a four-factor model in which the scale items of each construct loaded on separate factors (conflicting thoughts, sense-making, affective trust,

and cognitive trust). As shown in Table D.1 of Appendix D, this four-factor model had a better fit than several alternative three-factor and two-factor models and a one-factor model in which all items loaded onto the same factor. As such, these findings again provide statistical evidence for the distinctiveness of our constructs.

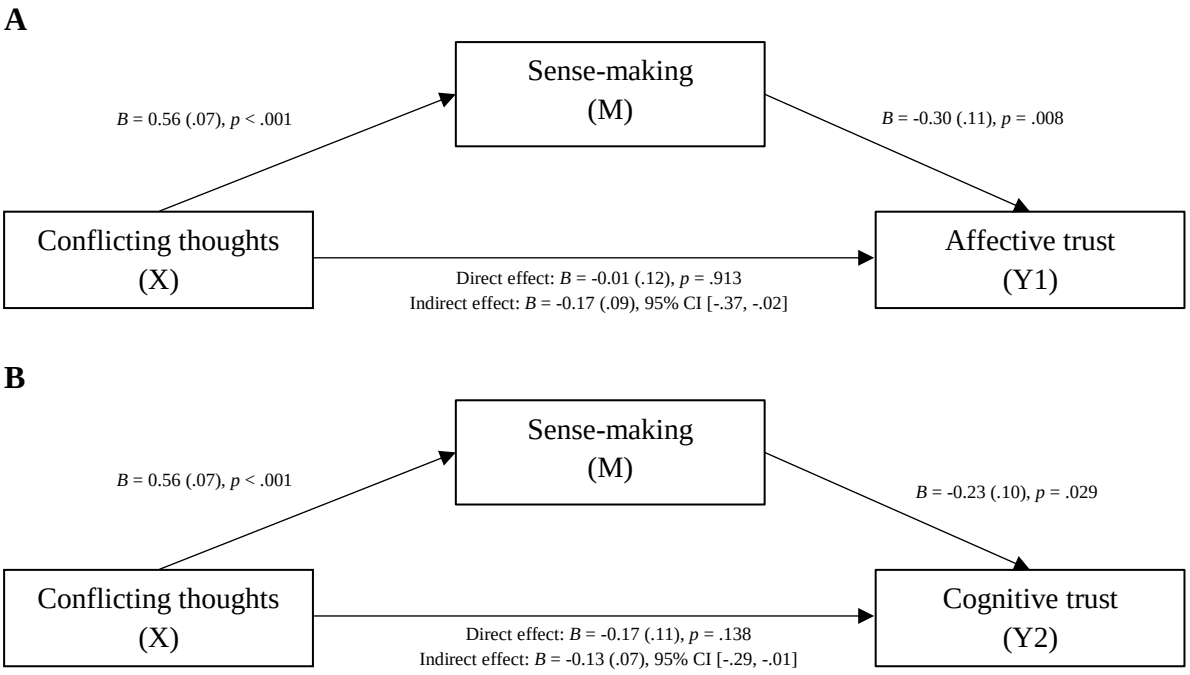
3.2.2 Hypothesis Testing

3.2.2.1 Mediation Analysis. We conducted two mediation analyses (using Model 4 of the PROCESS macro of Hayes, 2013; based on 5,000 bootstrap samples) to test if—in the aftermath of receiving overcompensation—conflicting thoughts negatively affect trust perceptions through sense-making processes. Gender and age were again controlled for as covariates. In a first analysis, conflicting thoughts (X) was included as the predictor, sense-making (M) as the mediator, and affective trust (Y1) as the outcome. The results of this analysis (see Figure 3A) revealed that sense-making mediated the relationship between conflicting thoughts and affective trust (indirect effect of X on Y1 via M: *B* = -0.17, boot *SE* = .09, 95% CI = [-.373, -.023]). The direct effect of conflicting thoughts on affective trust was non-significant (direct effect of X on Y1: *B* = -0.01, *SE* = .12, 95% CI = [-.254, .227]). We then conducted a similar analysis, but this time we used cognitive trust (Y2) as the outcome. The results of this analysis (see Figure 3B) revealed that the relationship between conflicting thoughts and cognitive trust was also mediated by sense-making (indirect effect of X on Y2 via M: *B* = -0.13, boot *SE* = .07, 95% CI = [-.289, -.005]). The direct effect of conflicting thoughts on cognitive trust was non-significant (direct effect of X on Y2: *B* = -0.17, *SE* = .11, 95% CI = [-.387, .054]). Overall, these findings indicate that conflicting thoughts resulting from overcompensation negatively affect trust perceptions through sense-making processes, which is consistent with our predictions made in *Hypothesis 3*.

3.2.2.2 Supplementary Mediation Analysis. To test if reverse causality exists between conflicting thoughts and sense-making, we conducted two additional mediation analyses in

which the order of these two variables was inverted. The results of these analyses revealed that, for both trust components, the indirect effect of sense-making on trust through conflicting thoughts was non-significant (indirect effect for affective trust: *B* = -0.01, boot *SE* = .07, 95% CI = [-.113, .148]; indirect effect for cognitive trust: *B* = -0.08, boot *SE* = .05, 95% CI = [-.182, .036]). These findings again suggest that it is more likely that conflicting thoughts following overcompensation induce sense-making processes than that sense-making processes following overcompensation induce conflicting thoughts.

FIGURE 3. Results of the Mediational Analyses of Study 2. Values shown are unstandardized coefficients (with *SE* in parentheses).



3.3 Discussion

Our second behavioral study aimed to replicate and extend our prior findings, by further investigating the role that conflicting thoughts and sense-making play in the relationship between overcompensation and trust perceptions. In line with the predictions made in *Hypothesis 3*, the results of the present study provide further cross-sectional evidence that the conflicting thoughts that result from being overcompensated negatively impact trust

perceptions through sense-making processes, and this seems to be true for both the affective and the cognitive trust dimension.

4. STUDY 3

The main aim of Study 3 was to provide converging evidence for the different psychological processes that occur when being exposed to overcompensation, using a neuroimaging approach. Studies 1 and 2 were both behavioral studies that explicitly asked participants to what extent they experienced conflictual thoughts and engaged in sense-making processes after being overcompensated. We acknowledge that by explicitly asking about conflict and sense-making we may have increased the base rate of these two processes. In order to examine if these two processes are also *spontaneously* activated after receiving overcompensation, in this study we investigated directly how people's brain react at the exact moment that they are overcompensated.

Another issue is that in our previous two studies, to provide the participant an overcompensation, the allocator took resources from another person and gave these resources to the participant—thereby creating yet another form of injustice. As such, it is possible that what is being captured in our prior two studies reflects a reaction to further unfairness, rather than a reaction to overcompensation *per se*. Therefore, in the present study, we used a somewhat different setup in which the allocator first divided the resources unequally between him or herself and the participant, and subsequently used his or her *own* resources to compensate the participant (for a similar approach, see Haesevoets et al., 2018).

We briefly recall the predictions we made in the introduction with respect to brain activation. After being overcompensated rather than being equally compensated, we predicted increased activations in two distinct brain networks. More specifically, we hypothesized that people will experience higher levels of conflicting thoughts and thus consequently show increased activation in the conflict-monitoring network (including the

key areas: pmFC and lateral prefrontal cortex; cf. *Hypothesis 1*) as well as higher levels of sense-making and thus also show increased activation in the mentalizing network (including the key areas: precuneus, TPJ, and mPFC; cf. *Hypothesis 2*).

4.1 Method

4.1.1. Sample. Given our available research budget, we were able to scan a total of 30 undergraduate students from a Belgian university (a sample size common in recent fMRI studies; cf. Haesevoets et al., 2018; Park & Young, 2020). The results of a sensitivity analysis revealed that, with the aforementioned sample size, and under standard criteria, our study had 80% power to detect a within-subject effect of size $f = .62$. One participant was excluded from the analyses due to movement artifacts. The remaining 29 participants (3 males) had a mean age of 22.79 years ($SD = 2.54$). All participants were right-handed and native Dutch speaking. Participants had no abnormal neurological history and had normal or corrected-to-normal vision. Informed consent was obtained in a manner approved by the Medical Ethics Committee at the University Hospital where the study was conducted (Ref: 2014/0865; Project title: Neural Correlates of Trust Violations and Trust Repair). Participants received €40 for their participation plus an additional amount. This additional amount was said to be determined by their outcome in the study. In reality, however, each participant received a surplus of €10.

4.1.2 Design. To ensure a necessary number of replications, we employed a within-subjects design that contained four conditions, which each consisted of 20 trials. Besides an equal compensation and an overcompensation condition, we also included a no compensation condition and an equal division condition (in which no violation occurred). We did this in order to make it more realistic that participants were allegedly playing against other students who could freely choose whether or not to divide the resources equally, and, in the aftermath of an unequal division, whether or not to compensate the participant. The overcompensations

were always exactly five time larger than the inflicted financial harm. We used such large overcompensation amounts to magnify the conflict between self-interest and fairness concerns—that is, the larger the overcompensation, the more the recipient's self-interest is served, but at the same time, the larger the overcompensation, the more it deviates from the norm of equality (for a similar reasoning, see Haesevoets, Van Hiel, Onraet, Joosten, & De Cremer, 2017; Haesevoets, Van Hiel, Pandelaere, & De Cremer, 2019).

4.1.3 Procedure and Measures.

4.1.3.1 Pre-Scanning. Before scanning, participants received written and oral instructions. Participants were informed that during the experiment they would engage in a series of allocation tasks in the role of recipient, with other students in the role of allocator. Participants were told that these tasks would be played for real money, and that the other students were simultaneously conducting a behavioral experiment at another building on campus. Participants learnt that they would be paired with a different partner in each round, and never with the same partner twice. Unknown to the participants, these other students were preprogrammed. To be as realistic as possible in making believe that participants were playing against other students, we used photographs that were selected from the student pool of our university to depict the allocators.

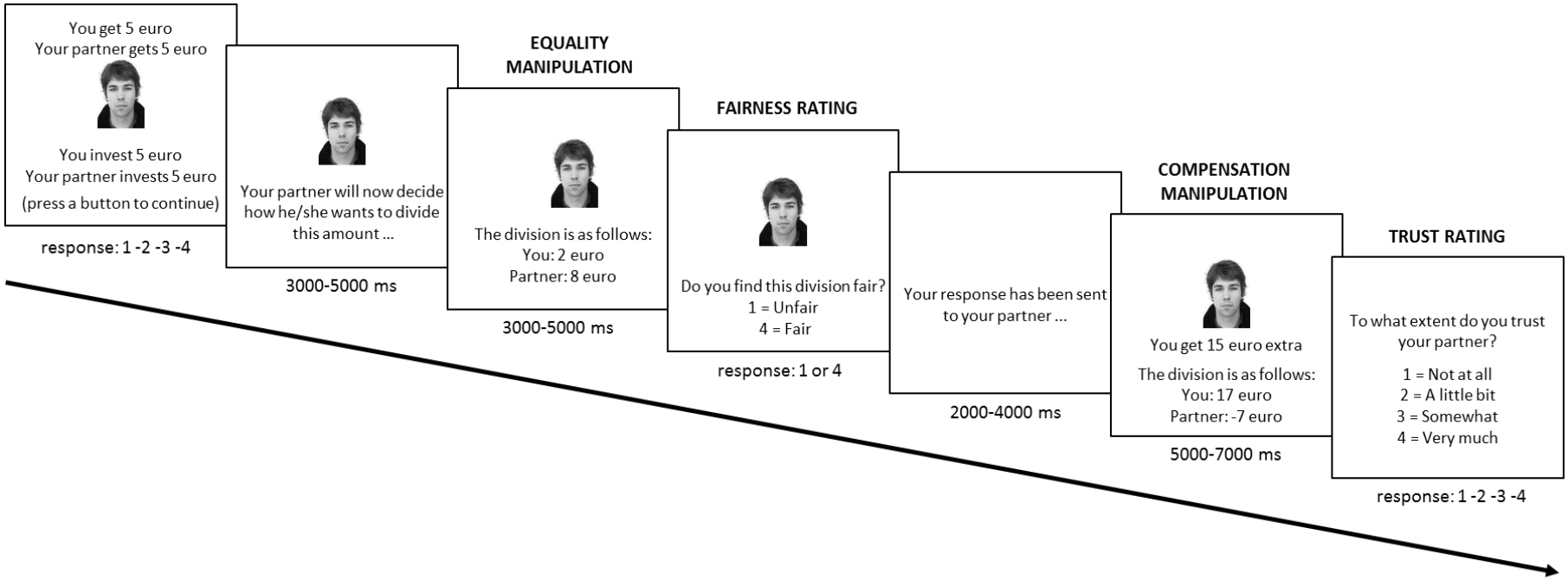
4.1.3.2 fMRI Experiment. During scanning, participants engaged in an experimental task, which was modelled after the dictator game (Kahneman, Knetsch, & Thaler, 1986). Figure 4 contains a graphical representation of what occurred in each trial. At the start of each trial, the participant and the other player were both provided with a starting budget, which they both had to invest in the task. The participants were then presented with a screen that stated that the other player was deciding how to divide the resources. On the next screen, participants received either an *equal division* or an *unequal division* from the other player (equality manipulation). Similar to our two prior behavioral studies, participants were

subsequently asked whether or not if they perceived the division as fair. After participants responded, a screen appeared which stated that the other player was being informed about the participant's decision. When participants received an unequal division and perceived this division as unfair, the other player provided them either *no compensation*, *equal compensation*, or *overcompensation* (compensation manipulation). After being confronted with this compensation manipulation, the participants proceeded to the final screen on which they were asked to indicate the extent in which they trust the allocator: "To what extent do you trust your partner?" (1 = *not at all*, 4 = *very much;* based on Haesevoets et al., 2018).[3] Participants who received an equal division and those who received an unequal division but perceived this division as fair immediately proceeded to rate their trust in the other player (and thus did not receive the compensation manipulation). To avoid order effects, the 80 trials were presented to the participants in a random order. After 20 trials, there was always a 30-second pause.

4.1.3.3 Post-Scanning. After scanning, participants were asked if they had recognized any of the allocators and if they were aware that they had played against preprogrammed players; which was not the case. Before leaving, participants were thoroughly debriefed, paid, and thanked for their participation.

[3] In the present study, participants had to rate their trust 80 times in total (i.e., at the end of each trial). Therefore, we used a single trust item to measure participants' trust in the allocators. This single item was also administered in Studies 1 and 2. Additional analyses suggest that this single-item measure of trust is an appropriate substitute for the multi-item trust measures that we reported in our behavioral studies. Specifically, we obtained high correlations between this single-item trust measure and composite scores of the multi-item trust measures in our behavioral studies (r = .89 and r = .71 for Studies 1 and 2, respectively).

FIGURE 4. Schematic Illustration of the Experimental Task of Study 3 (Example of an Overcompensation Trial).



Note. For better estimation of the hemodynamic blood function of each condition, it is necessary that screens do no start exactly at the beginning of each 2 seconds scan. Therefore, a jitter (or pause) was added lasting randomly between 550 and 6050 ms (for each trail and participant) before the onset of screens 1, 3, 4, 6, and 7.

4.1.4 Imaging Procedure. Images were collected using a 3 Tesla Magnetom Trio MRI scanner system (Siemens medical Systems, Erlangen, Germany), with a 32-channel radiofrequency head coil. Stimuli were projected onto a screen at the end of the magnet bore that participants viewed by way of a mirror mounted on the head coil. Stimulus presentation was controlled by E-Prime 2.0 (Psychology Software Tools) running under Windows XP. Foam cushions were placed within the head coil to minimize head movements. We first collected a high-resolution T1-weighted structural scan (MP-RAGE) followed by one functional run (30 axial slices; 4 mm thick; 1 mm skip). Functional scanning used a gradient-echo echo-planar pulse sequence (TR = 2 s; TE = 33 ms; 3.5 × 3.5 × 4.0 mm in-plane resolution).

4.1.5 Image Processing. The fMRI data were preprocessed and analyzed using the Statistical Parametric Mapping software package (SPM12; Wellcome Department of Cognitive Neurology, London, UK). For the functional run, data were preprocessed to remove sources of noise and artifact. Functional data were corrected for differences in acquisition time between slices for each whole-brain volume, realigned within and across runs to correct for head movement, and co-registered with each participant's anatomical data. The functional data were then transformed into a standard anatomical space (2 mm isotropic voxels) based on the ICBM152 brain template (Montreal Neurological Institute, MNI). Normalized data were spatially smoothed (6 mm full-width at half-maximum, FWHM) using a Gaussian Kernel. The preprocessed data were then examined using the Artifact Detection Tool software package (ART), for excessive motion artifacts and for correlations between motion and experimental design, and between global mean signal and experimental design. Outlier scans were identified in the temporal differences series by assessing between-scan differences using the following criteria in ART (Z-threshold: 3.0 mm, scan to scan movement threshold: 0.5 mm; rotation threshold: 0.02 radians). By default, these outliers were omitted

from the analyses by including a single regressor for each outlier. Six directions of motion parameters from the realignment step as well as outlier time points (defined by ART) were included as nuisance regressors. We used a default high-pass filter of 128 s and serial correlations were accounted for by the default auto-regressive AR(1) model.

4.1.6 Statistical Analysis

4.1.6.1 First-Level Analysis. The statistical analyses of the fMRI data involved first-level single participant analyses using the general linear model (GLM) of SPM12 for extracting a regressor for each condition time-locked at the presentation of the stimulus slide using the canonical response function with duration 0, six movement artifact regressors, and a variable amount of artifact regressors determined by ART. Specifically, we modelled the onset of the equality manipulation (slide 3 in Figure 4), fairness rating (slide 4), compensation manipulation (slide 6), and trust rating (slide 7). Given all possible conditions, this resulted in 12 regressors: Three regressors given an equal division (Equal Division, Equal-Fairness, Equal-Trust) and nine regressors given an unequal division (Unequal Division, Unequal-Fairness, Unequal-Unfairness) followed by compensation (Unequal-No Compensation, Unequal-Equal Compensation, Unequal-Overcompensation) and related trust rating (Unequal-No Compensation Trust, Unequal-Equal Compensation Trust, Unequal-Overcompensation Trust). We also modelled the evaluation of different amounts of monetary reward in a short task after the main experiment (that was not further analyzed), leading to an additional four regressors of no interest.

4.1.6.2 Second-Level Analysis. Analyses of interest were performed at the second, group level on the parameter estimates (regressors) associated with each condition using a random-effects model, which allows valid population level inferences. To investigate which brain regions are activated more immediately after receiving overcompensation (compared to after receiving equal compensation), whole-brain contrast analyses were computed for the

critical Overcompensation > Equal Compensation contrast. For completeness, we also contrasted these two conditions against the no compensation baseline condition. Moreover, we also computed the reverse contrasts (although we expect little activation here). A voxel-based statistical threshold of $p < .05$ (FWE-corrected) was used for all contrast analyses with a minimum cluster extent of 10 voxels.

4.2 Results

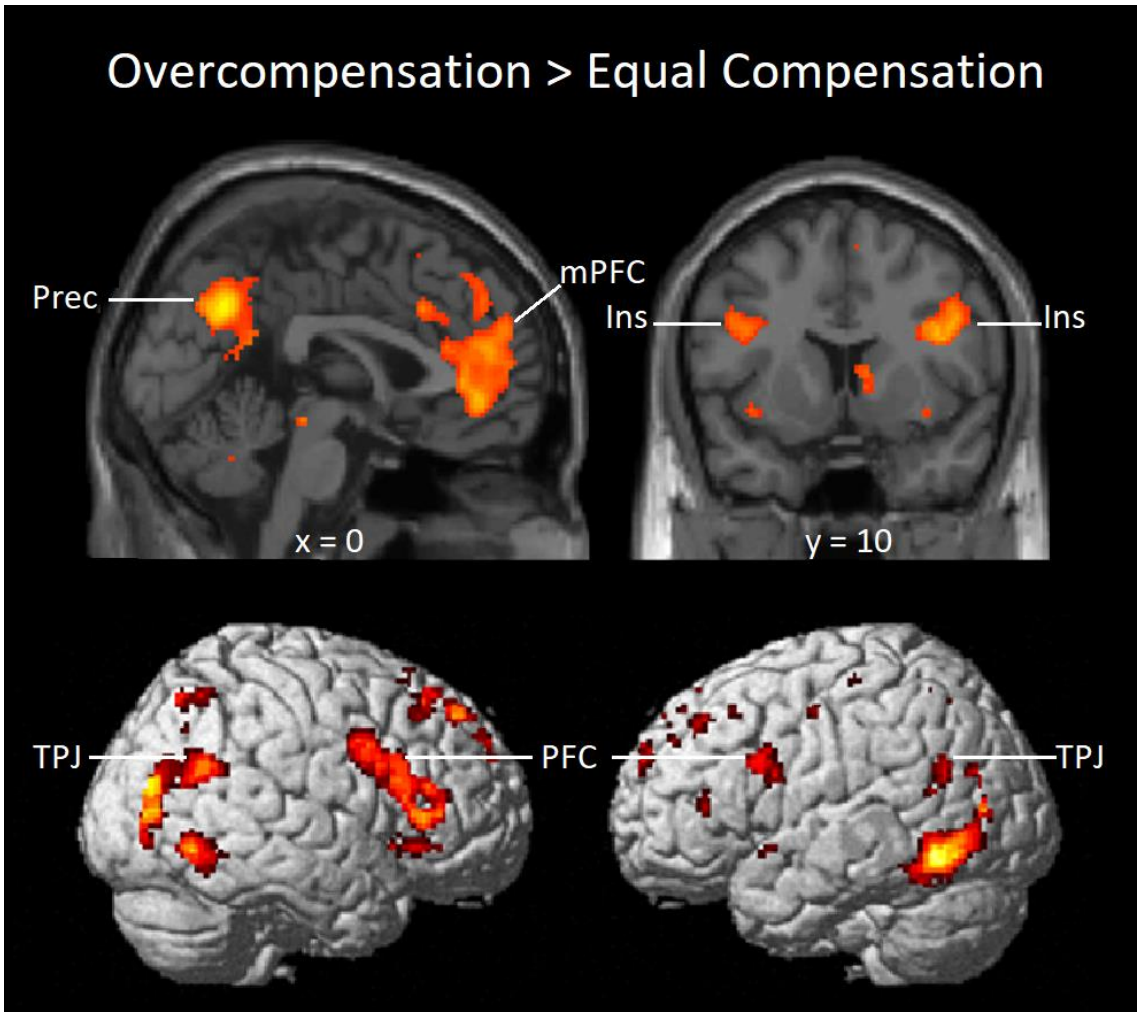
4.2.1 Behavioral Results. We first analyzed participants' trust ratings using a repeated-measures analysis of variance (ANOVA). We found a significant main effect of our four experimental conditions on trust perceptions, $F(3, 84) = 133.20$, $p < .001$, partial $\eta^2 = .83$. Subsequent pairwise comparison tests revealed that trust perceptions were significantly higher (all ps < .001, all ds > 0.77) in the equal division condition (in which no violation occurred; $M = 3.67$, $SD = 0.44$) than in the three compensation conditions. Moreover, after receiving equal compensation ($M = 2.89$, $SD = 0.51$) and overcompensation ($M = 3.10$, $SD = 0.68$) trust perceptions were significantly higher (both ps < .001, both ds > 2.23) than after receiving no compensation ($M = 1.29$, $SD = 0.37$). Critically, however, the perception of trust did not differ significantly between the equal compensation and the overcompensation condition ($p = .091$, $d = 0.33$). Mirroring the results of Study 1, this latter result provides further evidence that overcompensation does not result in higher trust levels than equal compensation.

4.2.2 Imaging Results

4.2.2.1 Whole-brain Contrast Analysis. To identify the brain areas that are involved in the processing of overcompensation, we next contrasted the brain responses from the overcompensation trials with those from the equal compensation trials. The resulting brain activations are visualized in Figure 5; Appendix E (Table E.1) provides a full list of the activated brain regions for all possible compensation contrasts. In support of *Hypothesis 1*,

the results for the Overcompensation > Equal Compensation contrast revealed increased activation in the bilateral inferior frontal gyrus (IFG), a part of the lateral prefrontal cortex of the *conflict-monitoring network* (Balodis et al., 2013; Gauvin, De Baene, Brass, & Hartsuiker, 2015; Ye & Zhou, 2009). As expected, increased activation was also found in areas that communicate with the conflict-monitoring system, including the bilateral insula (for receiving proprioceptive input; Klein, Ullsperger, & Danielmeier, 2013; Ullsperger, Harsay, Wessel, & Ridderinkhof, 2010) and the caudate nucleus (for regulating reward; Haruno et al., 2004; Haruno & Kawato, 2006). Moreover, the Overcompensation > Equal Compensation contrast additionally also showed increased activation in the *mentalizing network* of the brain, including the bilateral TPJ, mPFC, and precuneus (Fritt & Fritt, 2003; Hyatt, Calhoun, Pearlson, & Assaf, 2015; Mitchell, 2009; Van Overwalle, 2009; Van Overwalle & Baetens, 2009), a finding which corroborates *Hypothesis 2*. So, in line with our neural predictions, the results of the whole-brain contrast analysis indicate that, relative to equal compensation, overcompensation results in higher activations in areas of both the conflict-monitoring network (cf. *Hypothesis 1*) as well as the mentalizing network (cf. *Hypothesis 2*) of the brain.

FIGURE 5. Brain Activations for the Overcompensation > Equal Compensation Contrast in Study 3.



Note. Bain activations thresholded at $p < .05$ (FWE-voxel-corrected) with a minimum cluster extent of 10 voxels. PFC = bilateral prefrontal cortex, TPJ = bilateral temporoparietal junction, mPFC = medial prefrontal cortex, Prec = precuneus, Ins = insula.

4.2.2.2 NeuroSynth Analysis. But are the presently assumed psychological processes related to these specific neural areas the only possible interpretation? In order to verify to what extent other processes could underlie the activations, we subsequently inquired the NeuroSynth database (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). We did so by entering the peak MNI coordinates of the largest relevant clusters in Table E.1 obtained for the Overcompensation > Equal Compensation contrast, and searching in NeuroSynth for the strongest associated psychological function by means of the highest posterior probability,

functional connectivity, and meta-analytic coactivation parameters. For the *mentalizing* areas, the results were unambiguously related to sense-making process terms such as “theory of mind” (i.e., mentalizing) and “autobiographic” (i.e., self), given the peak activation of the mPFC (MNI coordinates -2 44 -4), left TPJ (-40 -60 24), right TPJ (46 -58 26), and precuneus (-4 -58 40). For the *conflict-monitoring* areas, the results were somewhat more equivocal, related to processes such as “conflicting” and more generally “tasks” for the right IFG (38 8 26), and also including “phonological” processes for the left IFG (-40 6 24). In sum, while the NeuroSynth data suggest that overcompensation indeed seems to result in higher levels of sense-making (which is in line with *Hypothesis 2*), the increased IFG activation may be related to conflicting thoughts (as was predicted by *Hypothesis 1*), but could be explained by other executive task processes as well.

4.2.2.3 Correlation Analysis. Finally, we also examined to what extent the conflict-related and the mentalizing-related brain responses are correlated with each other. Towards this end, we first extracted the neural response (i.e., percentage signal change) of each participant in the largest relevant clusters of the conflict resolution (i.e., left and right IFG) and sense-making processes (i.e., precuneus, left and right TPJ, and mPFC), by specifying regions of interest (ROI) as spheres of 15 mm around the peak coordinates of these clusters using the MarsBar toolbox (marsbar.sourceforge.net). We then computed correlations between the brain activations in these ROIs across participants. The correlation matrix (which is shown in Table 4) illustrates that the conflict-related and the mentalizing-related brain responses were mostly positively and significantly correlated.

TABLE 4. Correlations (Pearson *r*) among the Brain Activations for Study 3.

	Left IFG	Right IFG	Precuneus	Left TPJ	Right TPJ
Left IFG					
Right IFG	.67***				
Precuneus	.37*	.48**			
Left TPJ	.50**	.42*	.57**		
Right TPJ	-.01	.06	.41*	.49**	
mPFC	.29	.40*	.56**	.34†	.37*

Note. † $p = .071$, * $p < .05$, ** $p < .01$, *** $p < .001$.

4.3 Discussion

In line with the neural predictions made in *Hypotheses 1* and *2*, the results of our fMRI experiment indicate that, relative to equal compensation, overcompensation results in higher activation in both the conflict-monitoring network and the mentalizing network of the brain. As such, the present neuroimaging study provides converging evidence for Study 1, which behaviorally illustrated that overcompensation induces higher levels of conflicting thoughts and sense-making than equal compensation. What is critical, however, is that in the current study these activations occurred *spontaneously* in people's brain, that is, they occurred even though we did not ask participants whether they engaged in a conflict and/or a sense-making process.

5. GENERAL DISCUSSION

Trust violations are ubiquitous in human interactions, making efforts to repair trust highly relevant. In interpersonal settings, trust breaches often occur under the form of unfair resource distributions. One possible response to such violations involves the provision of a monetary reimbursement. In fact, it can be argued that, in order to completely regain broken trust, transgressors should show their willingness to go the "extra mile" by explicitly overcompensating the victim. The main objective of the present research was to uncover the psychological processes—and supporting neural systems—that are involved in the processing

of financial overcompensation. Additionally, we also examined how these processes relate to perceptions of trustworthiness. From our studies, it can be confidently inferred that despite the intuitive appeal of the "extra mile" logic, it in fact does not work. Instead, our findings show that overcompensation can even backfire and lead to a further decline of trust. Below, we will explain in detail *why* this is the case.

5.1 Main Findings

A first important finding of our research is thus that overcompensation does not result in higher trust levels than equal compensation (a result that mirrors prior studies; e.g., Garrett, 1997; Gelbrich & Roschk, 2011; Goode et al., 2017; Haesevoets, Reinders Folmer, De Cremer, & Van Hiel, 2013; Haesevoets, Van Hiel, et al., 2014; Mack et al., 2000; Noone & Lee, 2011). Specifically, our behavioral data revealed that, relative to equal compensation, overcompensation was less effective to enhance trust in Study 1 and equally effective in Study 3. This lack of positive overcompensation effects is actually in line with prior research on the *too-much-of-a-good-thing* effect (Grant & Schwartz, 2011; Pierce & Aguinis, 2013), which entails that positive phenomena can reach inflection points; exceeding these inflection points is undesirable because it leads either to no additional benefits (cf. Study 3) or, worse, even to undesirable outcomes (cf. Study 1). Although this phenomenon has been illustrated in various research domains (see Grant & Schwartz, 2011, for an overview), our study illustrates that this effect also applies to the relationship between compensation and trust, by showing that too much compensation can paradoxically have an adverse effect on trust.

But why is overcompensation not more effective than equal compensation to enhance trust? The results of Study 1 showed that people experience more conflicting thoughts (i.e., both positive and negative thoughts) and engage in more sense-making processes after being overcompensated than after being equally compensated. Moreover, our neuroimaging data (Study 3) revealed that overcompensation was associated with increased activation in a

central region of the conflict-monitoring network (i.e., the bilateral IFG) and in three important areas of the mentalizing network (i.e., the bilateral TPJ, the mPFC, and the precuneus). Taken these behavioral and neuroimaging findings together, the present research suggests that, compared to equal compensation, overcompensation results in higher levels of conflicting thoughts (which is supported by increased brain activations related to conflict; *Hypothesis 1*) and higher levels of sense-making (which is supported by increased mentalizing-related brain activations; *Hypothesis 2*).

However, it is important to stress that the brain regions that were activated in our study for the critical Overcompensation > Equal Compensation contrast are typically functionally heterogeneous, meaning that these brain regions are involved in multiple psychological processes. So, even though our NeuroSynth analysis revealed that the brain regions that were activated during our study have most often been associated with terms such as "conflicting" (in case of the conflict-related activations) and "theory of mind" (in case of the mentalizing-related activations) in other studies, these are not the only mental processes with which those regions have been associated. Our NeuroSynth analysis, for instance, revealed that the increased conflict-related activation that we found in our study may not only be related to conflict processing, but could possible also be explained by executive task processes. As such, based on the present findings we cannot straightforwardly assume that conflicting thoughts and sense-making processes were necessarily *evoked* based on the activity in those regions, but rather that these brain activations provide *converging evidence* for our behavioral findings.

Finally, to further extend this line of research into the domain of trust, we also examined how the conflict and the sense-making process relate to trustworthiness perceptions. The results of Study 1 suggest that, relative to equal compensation, overcompensation induces more conflict thoughts in the mind of the recipient, which

subsequently result in more sense-making processes, which in turn negatively affect their trust perceptions. Similar findings were also obtained in Study 2, which replicated these mediation effects for both the affective and the cognitive dimension of trust. It can thus be concluded that conflicting thoughts and sense-making both seem to play a critical role in explaining the lack of positive overcompensation effects on trust perceptions, and our findings suggest that they do so by forming the hypothesized serial mediation chain (*Hypothesis 3*).

5.2 Theoretical Implications

The present set of studies contributes to cumulative theoretical knowledge in psychology. As mentioned above, several prior studies have shown that overcompensation not always has the intended positive effect (for examples, see Estelami & De Maeyer, 2002; Garrett, 1997; Haesevoets et al., 2013; Haesevoets, Van Hiel, et al., 2014; Mack et al., 2000; Noone & Lee, 2011). However, from this prior body of research it is still unclear *why* overcompensation is not more effective than equal compensation to promote damaged trust. In this vein, a particular important implication of our work is that we are able to shed new light on this "why" question, and this by putting forward a *three-step sequential model* of how people respond to being overcompensated.

The *first step* of our model assumes that overcompensation simultaneously trigger both positive and negative thoughts in the mind of the recipient. In line with this assumption, the results of our studies suggest that people's first reaction when being confronted with an overcompensation is composed of both positive ("Wow, I got a really good outcome") and negative ("But that is not fair to the other") associations (see also Van den Bos et al., 2006; Van den Bos & Lind, 2013). Such conflicting thoughts in the face of ambivalence are activated simultaneously and unintentionally, and are thus triggered rather automatically (Berger, Hütter, & Corneille, 2019).

Moreover, our findings also suggest that after this first automatic conflict reaction—in a *second step* of our model—people will engage in a sense-making process, which can help them to figure out the inner thoughts and intentions of the person who overcompensated them ("Why is it that this person compensated me so generously?"). However, when people are influenced by conflicting sources, it is rather difficult for them to sort out how to respond toward these conflicting forces (e.g., Fiske & Taylor, 1991; Kunda, 1999). As such, this second sense-making reaction is expected to not be as spontaneous and automatic as the first reaction which elicited these thoughts, but instead requires cognitive effort to explicitly deal with this ambivalence when a decision has to be taken (Nohlen et al., 2016). Research has indeed shown that when people are influenced by two conflicting sources (like when being overcompensated)—and have to resolve this ambivalence—they are especially eager to systematically process information (Berger et al., 2019; Maio, Esses, & Bell, 2000; Nohlen et al., 2016), and this process takes some time.

In line with this reasoning about phase two, and directly applied to the context of overcompensation, using a reaction time paradigm, Peters and colleagues (2008) have found that people indeed need more time to make satisfaction judgements after being overpaid (which reflects an ambiguous situation) than after being equally paid (which reflects an unambiguous situation). This result is indicative of cognitive effort in the overcompensation condition. Another piece of evidence for the importance of systematic processing resides in the finding of Van den Bos and colleagues (2006) that cognitive busyness impacts people's satisfaction with advantageous inequity. According to these authors, if people are required to make a judgement about advantageous inequity, they need some time to discover, understand, and respond to the unfairness of a situation in which they are better off than others, and this correction process can only take place when people have enough cognitive resources available. In line with this expectation, Van den Bos et al. (2006) reported that people are less

satisfied with being advantaged when cognitive processing is only weakly (as opposed to strongly) limited. Applied to our own model, these two sets of studies indicate that, when overcompensated people have to decide on the trustworthiness of the provider, they will systematically process the positives and negatives.

Finally, our findings suggest that this deliberate, non-automatic sense-making process subsequently affects—in a *third step* of our model—people's judgement about the provider of the overcompensation. Indeed, our behavioral findings reveal that an important downstream consequence of step two (in which people tried to infer the inner intentions of the person who overcomposed them) is that people might become suspicious about the intentions of this person, which may eventually result in a negative assessment of this person's trustworthiness. This observation corroborates prior research of Steinel et al. (2014), who found that, in a bargaining context, receivers of an unusual high offer often consider such offers as "too good to be true," which in the end may lead them to reject it. In the context of service encounters, Estelami and De Maeyer (2002) similarly reported that "overgenerosity" (i.e., outperforming consumers' expectations through too generous actions) negatively impacts customers' evaluation of a service, and this because acts of overgenerosity triggers additional cognitive processes that put into question the truthfulness of the service provider.

5.3 Limitations and Future Research

Notwithstanding our study provides new and useful insights in the processes that underly overcompensation effects on trust, the present findings should be interpreted within the context of some limitations. A first and most important limitation of our work is that our data are all cross-sectional in nature. Although the present findings are in line with our proposed serial mediation model, any conclusions about causality cannot be drawn. Prior research has shown that inferring causality from mediation analysis can indeed be tedious

(Bullock, Green, & Ha, 2010), and some authors have even argued that it is impossible to establish causation by statistical analysis alone (e.g., Fiedler, Harris, & Schott, 2015).

In light of our research question, one could, for instance, argue that overcompensation reduces trust in the allocator, and that these diminished trust perceptions evoke conflicting thoughts and lead to sense-making processes. However, there is ample evidence that experiencing psychological conflict with respect to a supervisor or decision-maker is a precursor of low trust in that person (e.g., Conchie, Taylor, & Charlton, 2011), and, conversely, that lack of psychological conflict (e.g., psychological safety) powerfully predicts trust (e.g., Edmondson & Lei, 2014). Thus, although such an alternative model cannot be disregarded based on the current cross-sectional data, theoretically it makes less sense than the serial mediation chain that we postulated.

For Studies 1 and 2, we statistically examined the plausibility of a reverse-causal model, in which the order of conflicting thoughts and sense-making was inverted. Although the results of these analyses suggest that it is more likely that overcompensation induces conflicting thoughts followed by sense-making processes than that overcompensation evokes sense-making processes followed by conflicting thoughts, it is important to note that testing for reverse causality itself is not uncontested (e.g., see Lemmer & Gollwitzer, 2017). In this regard, both Spencer, Zanna, and Fong (2005) and Bullock, Green, and Ha (2008) have argued that only "experimental-causal-chain" designs, in which both mediators are sequentially manipulated in separate experiments, can resolve issues related to inferring causality from statistical analysis. To further disentangle the causal roles of conflicting thoughts, sense-making, and trust as a result of being overcompensated, we encourage future researchers in this domain to rely on such experimental-causal-chain designs.

Another important limitation of our work is that, although the brain activations that we observed in Study 3 for the critical Overcompensation > Equal Compensation contrast

were consistent with our hypotheses, this does not necessarily imply that the neural differences that we observed between these two conditions are specific to the processing of overcompensation. Indeed, the current task the Overcompensation > Equal Compensation contrast could involve other psychological processes in addition to overcompensation as well. Examples of such other psychological processes intertwined with overcompensation are, among others, the processes of self-payoff, other-payoff, and inequality aversion. Because the current experimental design does not allow to fully dissociate between these more specific processes, it is hard to know if the observed activations are driven by overcompensation *per se*. Note that this observation can possibly also explain why we observed the term "autobiographic" (i.e., processes related to self) in the NeuroSynth analysis, which may actually reflect self-payoff processing. Future research is therefore encouraged to further breakdown overcompensation into more specific processes in order to get a better view on what exactly makes overcompensation ineffective.

The overcompensation amounts that we investigated in our studies were always at least double the damage. We used such large amounts to ensure that the overcompensations that we investigated were situated within the range of what most people consider to be an explicit overcompensation. In this regard, it must also be mentioned that the presently employed overcompensations only took the inflicted financial damage into consideration, whereas fairness violations are often psychologically significant and salient, and thus not easy to quantify financially. Indeed, it is often very hard to quantify what is "fair" and from which specific point onwards an equal compensation becomes an overcompensation. Because the processes that we investigated might be more complex in everyday life, we encourage future overcompensation studies in more natural, mundane settings.

Our studies solely focused on violations of the equality norm, while other allocation norms exist as well. The equality norm was preferred in the present study because we

foremost were interested in distributive fairness violations, and we wanted to ascertain that our participants well understood that a norm was broken. One other prominent norm concerns the equity norm, which prescribes that rewards are distributed proportionally to people's contributions (Deutsch, 1975; Leventhal, 1976). However, violations of the equity norm require at least a comparative context in which performances of oneself vis-à-vis others can be assessed, and such assessments always have a subjective element in them. As a result, violations of the equity norm are less clear-cut and more open to interpretation than violations of the equality norm. Moreover, equity violations are much harder to achieve in an experimental context; and especially so in a scanning environment. An interesting avenue for future studies may, nonetheless, be to examine if the presently reported overcompensation effects also hold true for other violation types.

Although our findings indicate that going the "extra mile" after a trust violation may be deleterious to trust, it is important to realize that overcompensation may also have positive outcomes. Particularly interesting in this regard is a study by Haesevoets, Van Hiel, et al. (2014), which illustrated that trust and cooperation entail different effects with regard to overcompensation. More specially, these authors reported that even when there is a certain level of distrust on the part of the victim, he or she might still be willing to cooperate with the transgressor after being overcompensated. The reason for this cooperation may be due to the prospect of receiving specific benefits from this relationship. As such, these findings illustrate that, in order to better understand the psychological consequences of overcompensation for interpersonal relationships, it is important that future studies look beyond trust ratings, and this by also taking other outcomes into account.

A final interesting avenue for future research might be to investigate if and how positive overcompensation effects can be boosted. Prior research has shown that reactions to distributive unfairness are generally more positive when procedural fairness is high

(Brockner & Wiesenfeld, 1996). Future compensation studies are therefore encouraged to investigate if overcompensation can become a more effective strategy—and possibly even more effective than equal compensation—to enhance trust perceptions when it is accompanied by fair procedures. According to Thibaut and Walker (1975; also see Tyler, Rasinski, Spodick, 1985), one vital characteristic of fair procedures concerns process control. In light of this concept, we expect that it may be possible that people will also react more positively to overcompensation when they are given the opportunity to express their feelings about the compensation (cf. high process control).

6 CONCLUSION

Despite the intuitive appeal of overcompensation as a strategy to enhance trust, lately there seems to be a growing consensus in the literature that overcompensation does not promote trust levels beyond equal compensation. To provide more insights in the *in*effectiveness of financial overcompensation as a strategy to enhance trust in interpersonal relations, we investigated behavioral and neural responses to overcompensation. The results of our studies clarify that overcompensation does not result in higher trust levels than equal compensation, and this because overcompensation evokes more conflicting thoughts and more sense-making processes than equal compensation. The present research illustrates how behavioral and neuroimaging data can be integrated to deepen our understanding of the processes that underlie the important psychological phenomenon of trust.

References

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 267-299). New York: Academic Press.

Amodio, D. M. (2010). Can neuroscience advance social psychological theory? Social neuroscience for the behavioral social psychologist. *Social Cognition, 28,* 695-716.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268-277.

Balodis, I. M., Molina, N. D., Kober, H., Worhunsky, P. D., White, M. A., Sinha, R., ... & Potenza, M. N. (2013). Divergent neural substrates of inhibitory control in binge eating disorder relative to other manifestations of obesity. *Obesity, 21,* 367-377.

Berger, K. T., Hütter, M., & Corneille, O. (2019). Investigating attitudinal ambivalence via sequential priming: Evidence for the simultaneous and unintentional activation of opposite evaluations. *Journal of Experimental Psychology: General, 148,* 1269-1297.

Bohnet, I., & Croson, R. (2004). Trust and trustworthiness. *Journal of Economic Behavior and Organization, 55,* 443-445.

Boshoff, C. (1997). An experimental study of service recovery options. *International Journal of Service Industry Management, 8,* 110-130.

Bottom, W. P., Gibson, K., Daniels, S. E., & Murnighan, J. K. (2002). When talk is not cheap: Substantive penance and expressions of intent in rebuilding cooperation. *Organization Science, 13,* 497-513.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8,* 539-546.

Brockner, J. (2002). Making sense of procedural fairness: How high procedural fairness can reduce or heighten the influence of outcome favorability. *Academy of Management Review, 27,* 58-76.

Brockner, J., & Wiesenfeld, B. M. (1996). An integrative framework for explaining reactions to decisions: interactive effects of outcomes and procedures. *Psychological Bulletin, 120,* 189-208.

Bromiley, P., & Cummings, L. L. (1996). Transaction costs in organizations with trust. In R. J Bies, B. H. Sheppard & R. J Lewicki (Eds.). *Research on Negotiation in Organizations* (pp. 219–247). Greenwich, CT: JAI Press.

Bullock, J. G., Green, D. P., & Ha, S. E. (2008). Experimental Approaches to Mediation: A New Guide for Assessing Causal Pathways. *Unpublished manuscript, Yale University, New Haven, CT*.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, But What's the Mechanism? (Don't Expect an Easy Answer). *Journal of Personality and Social Psychology, 98,* 550.

Buunk, B. P., & Van Yperen, N. W. (1989). Social comparison, equality, and relationship satisfaction: Gender differences over a ten-year period. *Social Justice Research, 3,* 157-180.

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic perspectives, 9,* 209-219.

Champely, S (2018). *Pwr: Basic Functions for Power Analysis*. R package version 1.2-2.

Conchie, S. M., Taylor, P. J., & Charlton, A. (2011). Trust and distrust in safety leadership: Mirror reflections? *Safety Science, 49,* 1208-1214.

Cummings, L. L., & Bromiley, P. (1996). The organizational trust inventory (OTI). *Trust in Organizations: Frontiers of Theory and Research, 302,* 39-52.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review, 7,* 324-336.

Davidow, M. (2003). Organizational responses to customer complaints: What works and what doesn't. *Journal of Service Research, 5,* 225-250.

Desmet, P. T., De Cremer, D., & Van Dijk, E. (2010). On the psychology of financial compensations to restore fairness transgressions: When intentions determine value. *Journal of Business Ethics, 95,* 105-115.

Desmet, P. T., De Cremer, D., & Van Dijk, E. (2011). In money we trust? The use of financial compensations to repair trust in the aftermath of distributive harm. *Organizational Behavior and Human Decision Processes, 114,* 75-86.

Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues, 31,* 137-149.

Dirks, K. T. (2000). Trust in leadership and team performance: Evidence from NCAA basketball. *Journal of Applied Psychology, 85,* 1004-1012.

Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science, 12,* 450-467.Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology, 87,* 611-628.

Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology, 87,* 611-628.

Dirks, K. T., Kim, P. H., Ferrin, D. L., & Cooper, C. D. (2011). Understanding the effects of substantive responses on trust following a transgression. *Organizational Behavior and Human Decision Processes, 114,* 87-103.

Dulebohn, J. H., Conlon, D. E., Sarinopoulos, I., Davison, R. B., & McNamara, G. (2009). The biological bases of unfairness: Neuroimaging evidence for the distinctiveness of procedural and distributive justice. *Organizational Behavior and Human Decision Processes, 110,* 140-151.

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 23-43.

Estelami, H., & De Maeyer, P. (2002). Customer reactions to service provider overgenerosity. *Journal of Service Research, 4,* 205-216.

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests–An analysis of articles published in 2015. *Journal of Experimental Social Psychology, 75*, 95-102.

Fish, M. J. (2008). An eye for an eye: Proportionality as a moral principle of punishment. *Oxford Journal of Legal Studies, 28,* 57-71.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition.* New York: McGraw-Hill.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of theory of mind in story comprehension. *Cognition, 57,* 109-128.

Fliessbach, K., Phillipps, C. B., Trautner, P., Schnabel, M., Elger, C. E., Falk, A., & Weber, B. (2012). Neural responses to advantageous and disadvantageous inequity. *Frontiers in Human Neuroscience, 6*(165).

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358,* 459-473.

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neurosciences, 14,* 433-438.

Garrett, D. E. (1999). The effectiveness of compensation given to complaining customers: Is more better? *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 12,* 26-34.

Gauvin, H. S., De Baene, W., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *NeuroImage, 126,* 96-105.

Gelbrich, K., & Roschk, H. (2011). Do complainants appreciate overcompensation? A meta-analysis on the effect of simple compensation vs. overcompensation on post-complaint satisfaction. *Marketing Letters, 22,* 31-47.

Gilly, M. C., & Hansen, R. W. (1985). Consumer complaint handling as a strategic marketing tool. *Journal of Consumer Marketing*, *2*, 5-16.

Goode, S., Hoehle, H., Venkatesh, V., & Brown, S. A. (2017). User compensation as a data breach recovery action: An investigation of the Sony Playstation Network breach. *MIS Quarterly, 41,* 703-727.

Grant, A. M., & Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. *Perspectives on Psychological Science, 6,* 61-76.

Guarana, C. L., & Hernandez, M. (2016). Identified ambivalence: When cognitive conflicts can help individuals overcome cognitive traps. *Journal of Applied Psychology, 101,* 1013-1029.

Haesevoets, T., De Cremer, D., Van Hiel, A., & Van Overwalle, F. (2018). Understanding the positive effect of financial compensation on trust after norm violations: Evidence from fMRI in favor of forgiveness. *Journal of Applied Psychology, 103*, 578-590.

Haesevoets, T., Reinders Folmer, C., De Cremer, D., & Van Hiel, A. (2013). Money isn't all that matters: The use of financial compensation and apologies to preserve relationships in the aftermath of distributive harm. *Journal of Economic Psychology, 35,* 95-107.

Haesevoets, T., Reinders Folmer, C. & Van Hiel, A. (2014). More money, more trust? Target and observer differences in the effectiveness of financial overcompensation to restore trust. *Psychologica Belgica, 54,* 389-394.

Haesevoets, T., Van Hiel, A., Onraet, E., Joosten, A., & De Cremer, D. (2017). Low-price guarantees as advertisement strategy and compensation policy: The more, the better? *Journal of Consumer Behaviour, 16,* 389-402.

Haesevoets, T., Van Hiel, A., Pandelaere, M., & De Cremer, D. (2019). Towards identifying customer profiles in reactions to financial overcompensation: The role of self-interest and fairness sensitivity in explaining who prefers more money. *Journal of Consumer Behaviour, 18,* 261-271.

Haesevoets, T., Van Hiel, A., Reinders Folmer, C., & De Cremer, D. (2014). What money can't buy: The psychology of financial overcompensation. *Journal of Economic Psychology, 42,* 83-95.

Handgraaf, M. J., Van Dijk, E., & De Cremer, D. (2003). Social utility in ultimatum bargaining. *Social Justice Research, 16,* 263-283.

Haruno, M., & Kawato, M. (2006). Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *Journal of Neurophysiology, 95,* 948-959.

Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., ... & Kawato, M. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: A functional magnetic resonance imaging study of a stochastic decision task. *Journal of Neuroscience, 24,* 1660-1665.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York: The Guilford Press.

Hocutt, M. A., Bowers, M. R., & Donavan, D. (2006). The art of service recovery: fact or fiction? *Journal of Services Marketing, 20,* 199-207.

Hyatt, C. J., Calhoun, V. D., Pearlson, G. D., & Assaf, M. (2015). Specific default mode subnetworks support mentalizing as revealed through opposing network recruitment by social and semantic FMRI tasks. *Human brain mapping, 36,* 3047-3063.

Jacques, E. (1961). *Equitable payment.* London: Heinemann.

Jamieson, D. W. (1993). *The attitude ambivalence construct: Validity, utility, and measurement*. Paper presented at the annual meeting of the American Psychological Association, Toronto, Ontario, Canada.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business, 59,* 285-300.

Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence-vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes, 99,* 49-65.

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: the effects of apology versus denial for repairing competence-versus integrity-based trust violations. *Journal of Applied Psychology, 89,* 104.

Klein, T. A., Ullsperger, M., & Danielmeier, C. (2013). Error awareness and the insula: links to neurological and psychiatric diseases. *Frontiers in human neuroscience, 7,* 14.

Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: The MIT Press.

Lemmer, G., & Gollwitzer, M. (2017). The “true” indirect effect won't (always) stand up: When and why reverse mediation testing fails. *Journal of Experimental Social Psychology, 69,* 144-149.

Leventhal, G. S. (1976). The distribution of rewards and resources in groups and organizations. In *Advances in experimental social psychology* (pp. 91-131). Academic Press.

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces, 63,* 967-985.

Mack, R., Mueller, R., Crotts, J., & Broderick, A. (2000). Perceptions, corrections and defections: Implications for service recovery in the restaurant industry. *Managing Service Quality: An International Journal*, *10,* 339-346.

Maio, G. R., Esses, V. M., & Bell, D. W. (2000). Examining conflict between components of attitudes: Ambivalence and inconsistency are distinct constructs. *Canadian Journal of Behavioural Science, 32,* 71-83.

Maitlis, S. (2005). The social processes of organizational sensemaking. *Academy of Management Journal, 48,* 21-49.

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, *84,* 123-136.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, 32, 344-354.

McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38,* 24-59.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17,* 437-455.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function, 214,* 655-667.

Miller, W. I. (2005). *Eye for an Eye.* Cambridge University Press.

Miller, K.I., Joseph, L., & Apker, J. (2000). Strategic ambiguity in the role development process. *Journal of Applied Communication Research, 28,* 193-214.

Mitchell, J. P. (2006). Mentalizing and Marr: an information processing approach to the study of social cognition. *Brain Research, 1079,* 66-75.

Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364,* 1309-1316.

Nohlen, H. U., Van Harreveld, F., Rotteveel, M., Barends, A. J., & Larsen, J. T. (2016). Affective responses to ambivalence are context-dependent: A facial EMG study on the role of inconsistency and evaluative context in shaping affective responses to ambivalence. *Journal of Experimental Social Psychology, 65,* 42-51.

Noone, B. M., & Lee, C. H. (2011). Hotel overbooking: The effect of overcompensation on customers' reactions to denied service. *Journal of Hospitality & Tourism Research, 35,* 334-357.

Okimoto, T. G. (2008). Outcomes as affirmation of membership value: Material compensation as an administrative response to procedural injustice. *Journal of Experimental Social Psychology, 44,* 1270-1282.

Park, B., & Young, L. (2020). An association between biased impression updating and relationship facilitation: a behavioral and fMRI investigation. *Journal of Experimental Social Psychology, 87,* 103916.

Peters, S. L., Van den Bos, K., & Bobocel, D. R. (2004). The moral superiority effect: Self versus other differences in satisfaction with being overpaid. *Social Justice Research, 17,* 257-273.

Peters, S. L., Van den Bos, K., & Karremans, J. C. (2008). On the psychology of the advantaged: How people react to being overpaid. *Social Justice Research, 21,* 179-191.

Pierce, J. R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management*, *39,* 313-338.

Priester, J. R., & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality and Social Psychology, 71,* 431-449.

Ranganathan, S. K., Madupu, V., Sen, S., & Brooks, J. R. (2013). Affective and cognitive antecedents of customer loyalty towards e-mail service providers. *Journal of Services Marketing, 27,* 195-206.

Rempel, J. K., Holmes, J. G. & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49***,** 95-112.

Rothman, N. B. (2011). Steering sheep: How expressed emotional ambivalence elicits dominance in interdependent decision making contexts. *Organizational Behavior and Human Decision Processes*, *116,* 66-82.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*, 393-404.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16,* 235-239.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101*, 1-19.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron, 79,* 217-240.

Shockley-Zalabak, P., Ellis, K., & Winograd, G. (2000). Organizational trust: What it means, why it matters. *Organization Development Journal*, *18*, 35-48.

Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science, 4,* 367-392.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*, 845-851.

Spreitzer, G. M., & Mishra, A. K. (1999). Giving up control without losing control: Trust and its substitutes' effects on managers' involving employees in decision making. *Group & Organization Management, 24*, 155-187.

Steinel, W., Van Beest, I., & Van Dijk, E. (2014). Too good to be true: Suspicion-based rejections of high offers. *Group Processes & Intergroup Relations, 17*, 682-698.

Taylor, K. S., Seminowicz, D. A., & Davis, K. D. (2009). Two systems of resting state connectivity between the insula and cingulate cortex. *Human Brain Mapping, 30*, 2731-2745.

Thibaut, J., & Walker, L. (1975). *Procedural justice: A psychological analysis*. New York: Erlbaum Wiley.

Tyler, T. R., Rasinski, K. A., & Spodick, N. (1985). Influence of voice on satisfaction with leaders: Exploring the meaning of process control. *Journal of Personality and Social Psychology*, *48*, 72-81.

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function, 214*, 629-643.

Van den Bos, K., & Lind, E. A. (2013). On sense-making reactions and public inhibition of benign social motives: An appraisal model of prosocial behavior. In *Advances in experimental social psychology* (Vol. 48, pp. 1-58). Academic Press.

Van den Bos, K., Lind, E. A., Vermunt, R., & Wilke, H. A. (1997). How do I judge my outcome when I do not know the outcome of others? The psychology of the fair process effect. *Journal of Personality and Social Psychology, 72*, 1034-1046.

Van den Bos, K., Peters, S. L., Bobocel, D. R., & Ybema, J. F. (2006). On preferences and doing the right thing: Satisfaction with advantageous inequity when cognitive processing is limited. *Journal of Experimental Social Psychology, 42,* 273-289.

Van den Bos, K., Van Lange, P. A., Lind, E. A., Venhoeven, L. A., Beudeker, D. A., Cramwinckel, F. M., ... & van der Laan, J. (2011). On the benign qualities of behavioral disinhibition: Because of the prosocial nature of people, behavioral disinhibition can weaken pleasure with getting more than you deserve. *Journal of Personality and Social Psychology, 101,* 791-811.

Van den Bos, K., Wilke, H. A., Lind, E. A., & Vermunt, R. (1998). Evaluating outcomes by means of the fair process effect: Evidence for different processes in fairness and satisfaction judgments. *Journal of Personality and Social Psychology, 74,* 1493-1503.

Van Dijk, E., De Cremer, D., & Handgraaf, M. J. (2004). Social value orientations and the strategic use of fairness in ultimatum bargaining. *Journal of Experimental Social Psychology, 40,* 697-707.

Van Jaarsveld, D. D., Walker, D. D., & Skarlicki, D. P. (2010). The role of job demands and emotional exhaustion in the relationship between customer and employee incivility. *Journal of Management, 36,* 1486-1504.

Van Hiel, A., & Mervielde, I. (2002). Effects of ambiguity and need for closure on the acquisition of information. *Social Cognition, 20,* 380-408.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping, 30,* 829-858.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage, 48,* 564-584.

Weick, K. (1995). *Sensemaking in Organizations*. Newbury Park, CA: Sage.

Yang, J., Mossholder, K. W., & Peng, T. K. (2009). Supervisory procedural justice effects: The mediating roles of cognitive and affective trust. *The Leadership Quarterly, 20,* 143-154.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods, 8,* 665-670.

Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *Neuroimage, 48,* 280-290.

APPENDIX A

Overview of the Scale Measures Used in Study 1

Conflicting thoughts

The way in which the allocator has redistributed the work budget …

1. Makes me experience both positive and negative thoughts about the redistribution of this person.
2. Makes me think about this situation in conflicting ways.
3. Makes me experience a lot of mixed-thoughts towards this person.
4. Does not feel right to me.
5. Makes me both satisfied and dissatisfied at the same time.
6. Makes me feel conflicted about our relationship.

Sense-making

The way in which the allocator has redistributed the work budget …

1. Makes me question what has motivated this person to make this particular redistribution.
2. Makes me question what exact reasons might underlie this person's redistribution.
3. Makes me think about what was going on in this person's mind when making this specific redistribution.
4. Makes me wonder why this person has redistributed the units in this particular way.
5. Makes me reflect on the true intentions of this person.
6. Puts me in a state of mind that I want to figure out what this person's true intentions are for redistributing the units in this way.

Trust perceptions

The following questions are about the allocator. To what extent do you think that ...

1. You can trust the values of this person.
2. You can trust that this person will keep the promises that he or she makes.
3. You can trust this person to be relied on.
4. You can trust this person to be sincere to you.
5. You can trust this person to be completely honest with you.
6. You can trust this person to care about your well-being.
7. You would let this person have influence over issues that are important to you.

APPENDIX B

Comparison of Measurement Models for Study 1

Table B.1

Results of the Confirmatory Factor Analysis (CFA)

Model	Factors	χ^2	*df*	χ^2/df	$\Delta\chi^2$	CFI	RMSEA	SRMR
Null model	All indicators are independent.	2285.39	117	19.53				
Baseline Model	Three factors: Conflicting thoughts, sense-making, and trust perceptions.	310.69	149	2.09		0.92	0.10	0.08
Model 1	Two factors: Conflicting thoughts and sense-making were combined into one factor.	452.62	151	3.00	141.93***	0.86	0.14	0.08
Model 2	Two factors: Conflicting thoughts and trust perceptions were combined into one factor.	604.13	151	4.00	293.44***	0.79	0.17	0.17
Model 3	Two factors: Sense-making and trust perceptions were combined into one factor.	852.58	151	5.65	541.89***	0.67	0.21	0.17
Model 4	One factor: All three factors were combined into one factor.	978.19	152	6.44	667.50***	0.61	0.23	0.15

Note. *** $p < .001$. CFI = comparative fit index; RMSEA = root mean squared error of approximation; SRMR = standardized root mean square residual.

APPENDIX C

Overview of the Scale Measures Used in Study 2

Conflicting thoughts

The way in which the allocator has redistributed the work budget ...

1. Clearly serves my own self-interest, but not in ways I like because this redistribution disadvantages the other person.
2. Makes me experience a lot of mixed-thoughts, because this redistribution is unfair for the other.
3. Makes me think about this whole situation in conflicting ways, because I don't like it to be advantaged over others.
4. Makes me satisfied about my own good outcome but also dissatisfied about the bad outcome of the other.
5. Makes me feel conflicted because this redistribution is unfair.
6. Makes me simultaneously experience positive and negative thoughts.

Sense-making

The way in which the allocator has redistributed the work budget ...

1. Makes me question what has motivated this person to allocate such a large part of the work budget to me.
2. Makes me question what reasons might underlie this person's redistribution.
3. Makes me wonder why this person has redistributed the units in this particular way.
4. Makes me reflect on the intentions of this person for compensating me so generously.
5. Puts me in a state of mind wanting to figure out what this person's intentions were for giving me so many units.
6. Makes me think about what was going on in this person's mind when providing me this generous offer.

Affective trust

The following questions are about the allocator. To what extent do you think that …

1. This person cares about your personal needs.
2. If you share a problem with this person, he or she will respond with care.
3. You can share your difficulties with this person.
4. You can openly communicate your feelings to this person.
5. You can feel secure with this person because of his or her sincerity.

Cognitive trust

The following questions are about the allocator. To what extent do you think that …

1. You can depend on this person to meet his or her responsibilities.
2. You can rely on this person to do what is best.
3. This person will follow through with commitments he or she makes.
4. There is no reason to doubt this person's competence.
5. You can be confident that this person will act professional.

APPENDIX D

Comparison of Measurement Models for Study 2

Table D.1

Results of the Confirmatory Factor Analysis (CFA).

Model	Factors	χ^2	*df*	χ^2/df	$\Delta\chi^2$	CFI	RMSEA	SRMR
Null model	All indicators are independent.	2838.97	231	12.290				
Baseline Model	Four factors: Conflicting thoughts, sense-making, affective trust, and cognitive trust.	436.10	203	2.15		0.91	0.08	0.06
Model 1	Three factors: Conflicting thoughts and sense-making were combined into one factor.	665.99	206	3.23	229.89***	0.82	0.11	0.09
Model 2	Three factors: Affective trust and cognitive trust were combined into one factor.	681.85	206	3.31	245.75***	0.82	0.11	0.08
Model 3	Two factors: Conflicting thoughts and sense-making were combined into one factor and affective trust and cognitive trust were combined into one factor.	909.16	208	4.37	473.06***	0.73	0.14	0.10
Model 4	One factor: All four factors were combined into one factor.	1657.69	209	7.93	1221.60***	0.45	0.20	0.19

Note. *** $p < .001$. CFI = comparative fit index; RMSEA = root mean squared error of approximation; SRMR = standardized root mean square residual.

APPENDIX E

Overview of the Brain Activations for Study 3

Table E.1

Whole-Brain Comparisons for all Possible Compensation Contrasts

Comparison and Anatomical Area	Voxels	x	y	z	*t*	
Overcompensation > Equal Compensation						
R Middle Occipital Gyrus	939	36	-80	16	8.810	***
R Middle Occipital Gyrus		30	-68	26	7.947	***
R Superior Parietal Lobule		24	-60	50	7.008	***
L Middle Occipital Gyrus	15	-42	-74	26	5.248	*
L Middle Occipital Gyrus	235	-24	-66	36	6.370	***
L Middle Occipital Gyrus		-28	-76	22	6.330	***
L Superior Occipital Gyrus		-24	-68	28	6.211	***
R Inferior Temporal Gyrus	510	46	-62	-6	9.034	***
R Inferior Temporal Gyrus		44	-52	-10	8.718	***
R Fusiform Gyrus		32	-42	-10	7.756	***
L Inferior Temporal Gyrus	989	-44	-60	-8	11.583	***
L Inferior Temporal Gyrus		-52	-58	-10	8.868	***
L Fusiform Gyrus		-36	-40	-20	7.955	***
L Precuneus		-4	-58	40	8.203	***
L Precuneus		-8	-54	28	7.613	***
L Precuneus	68	-6	-46	58	5.813	**
L Angular Gyrus (ext. to TPJ)	69	-40	-60	24	5.951	***
R Angular Gyrus (ext. to TPJ)	258	46	-58	26	7.001	***
R Superior Temporal Gyrus		58	-60	22	5.083	*
L Superior Parietal Lobule	123	-26	-58	50	7.281	***
		-26	-48	40	5.003	*
Cerebellar Vermis 8	24	2	-58	-28	5.673	**
L Middle Temporal Gyrus	11	-52	-56	16	5.210	*
L Lingual Gyrus	17	-24	-52	-8	5.548	**
		-18	-44	52	5.698	**
L ParaHippocampal Gyrus	108	-22	-32	-12	6.622	***
		-2	-30	-12	6.126	***

Table E.1 continued

L Lingual Gyrus		-14	-34	-4	6.026	***
L Heschls Gyrus	26	-40	-26	10	5.767	**
L Precentral Gyrus	59	-30	-24	54	5.950	***
L Middle Cingulate	73	-6	-18	48	6.656	***
L Postcentral Gyrus	10	-52	-10	48	5.257	*
R Precentral Gyrus	1142	46	6	32	8.128	***
R Inferior Frontal Gyrus		38	8	26	7.582	***
R Inferior Frontal Gyrus		44	22	14	7.297	***
L Inferior Frontal Gyrus	262	-40	6	24	7.796	***
L Inferior Frontal Gyrus		-46	14	28	6.338	***
R Caudate Nucleus	46	10	12	6	5.939	***
L Insula Lobe	27	-38	10	-8	5.598	**
R Insula Lobe	70	28	22	-2	6.242	***
R Insula Lobe		34	18	-8	5.670	**
L Inferior Frontal Gyrus	23	-50	34	10	6.089	***
L Middle Frontal Gyrus	163	-22	34	42	7.323	***
L Middle Frontal Gyrus		-22	20	44	6.290	***
R Superior Frontal Gyrus (mPFC)	82	20	34	56	5.791	**
R Middle Frontal Gyrus		26	24	44	5.593	**
L Anterior Cingulate (mPFC)	2257	-2	44	-4	7.720	***
R Anterior Cingulate (mPFC)		2	50	14	7.340	***
R Anterior Cingulate (mPFC)		4	42	12	7.040	***
L Superior Frontal Gyrus (mPFC)	19	-18	46	36	5.537	**
Overcompensation > No Compensation						
L Precuneus	72	-4	-66	44	5.940	***
L Cerebellum VI	33	-24	-66	-28	5.452	**
L Cerebellum VI		-28	-60	-34	5.143	*
L Inferior Temporal Gyrus	502	-52	-58	-10	8.429	***
L Inferior Temporal Gyrus		-52	-50	-16	8.147	***
L Inferior Occipital Gyrus		-52	-72	-4	4.991	*
L Superior Parietal Lobule (ext. to TPJ)	1010	-26	-58	50	6.921	***
L Middle Occipital Gyrus		-26	-66	40	6.613	***
L Inferior Parietal Lobule		-38	-42	40	6.605	***

Table E.1 continued

R Middle Cingulate	26	2	-36	32	5.428	**
R Inferior Parietal Lobule (ext. to TPJ)	787	48	-32	48	7.829	***
R Inferior Parietal Lobule		36	-42	44	6.531	***
R Angular Gyrus		30	-58	44	6.199	***
R Precentral Gyrus	407	44	8	30	8.095	***
L Precentral Gyrus	504	-48	10	34	7.065	***
L Precentral Gyrus		-40	4	30	7.004	***
R Posterior Medial Frontal (pmFC/mPFC)	503	6	18	48	7.535	***
R Posterior Medial Frontal (pmFC/mPFC)		6	30	46	6.847	***
R Superior Frontal Gyrus (mPFC)		14	38	48	5.544	**
L Insula Lobe	10	-34	18	-4	5.147	*
R Insula Lobe	101	30	24	-2	6.572	***
R Middle Frontal Gyrus	380	48	40	16	6.744	***
R Middle Frontal Gyrus		44	32	18	6.285	***
R Middle Frontal Gyrus		40	20	32	6.151	***
L Inferior Frontal Gyrus	146	-42	42	10	6.007	***
L Inferior Frontal Gyrus		-48	36	16	5.932	***
L Inferior Frontal Gyrus	18	46	44	-4	5.279	*
Equal Compensation > No Compensation						
R Cerebellum Crus 1	35	34	-68	-32	5.723	**
L Inferior Parietal Lobule (ext. to TPJ)	391	-40	-42	46	6.678	***
L Inferior Parietal Lobule		-38	-40	38	6.268	***
L Inferior Parietal Lobule		-42	-54	50	5.481	**
R Inferior Parietal Lobule (ext. to TPJ)	547	48	-32	48	7.996	***
R Inferior Parietal Lobule		38	-46	44	6.499	***
R SupraMarginal Gyrus		40	-32	40	6.389	***
L Superior Medial Gyrus (pmFC)	45	-2	28	46	5.505	**
R Posterior Medial Frontal (pmFC)		4	20	46	5.154	*
Reverse Contrasts						
Equal Compensation > Overcompensation			No clusters			
No Compensation > Overcompensation			No clusters			
No Compensation > Equal Compensation			No clusters			

Note. x, y, and z = Montreal Neurological Institute (MNI) coordinates of the peak values; *t* = t-score of the peak values; L = left, R = Right, ext. = extending. Whole brain analysis thresholded at $p < .05$ FWE voxel-corrected with cluster extent > 10 voxels. * $p < .05$, ** $p < .01$, *** $p < .001$.