



Absolute vs. relative success: Why overconfidence is an inefficient equilibrium

Alice Soldà

Changxia Ke

William von Hippel

Lionel Page

AWI DISCUSSION PAPER SERIES NO. 700

March 2021

Absolute vs. relative success:

Why overconfidence is an inefficient equilibrium

Accepted for publication at Psychological Science

Alice Soldà¹, Changxia Ke², William von Hippel³ and Lionel Page⁴

¹ Heidelberg University, Department of Economics

² Queensland University of Technology, Department of Economics

³ University of Queensland, Department of Psychology

⁴ University of Technology Sydney, Department of Economics

12 March 2021

Abstract

Overconfidence is one of the most ubiquitous biases in the social sciences, but the evidence regarding its overall costs and benefits is mixed. To test the possibility that overconfidence might yield important relative benefits that offset its absolute costs, we conducted an experiment (N=298 university students) in which pairs of participants bargain over the unequal allocation of a prize that was earned via a joint effort. We manipulated confidence using a binary noisy signal to investigate the causal effect of negotiators' beliefs about their relative contribution on the outcome of the negotiation. Our results provide evidence that high levels of confidence lead to relative benefits (how much one earns compared to one's partner) but absolute costs (how much money one receives overall). These results suggest that overconfidence creates an inefficient equilibrium whereby overconfident negotiators benefit over their partners even as they bring about joint losses.

Keywords: Overconfidence, motivated beliefs, negotiation

Statement of Relevance

Overconfidence is one of the most well-documented biases in the social sciences, but the evidence regarding its costs and benefits is mixed. We present an experiment that attempts to reconcile these inconsistent findings by investigating the tension between relative and absolute costs and benefits. Our findings suggest that high levels of confidence can lead to an inefficient equilibrium. They also provide an explanation for why overconfidence can be both costly and ubiquitous; high levels of confidence can enhance relative success at an absolute cost.

Overconfidence is one of the most ubiquitous biases in the social sciences, but the evidence regarding its overall costs and benefits is mixed. For example, on the cost side of the equation, overconfident people often fail for lack of sufficient planning (Shipman & Mumford, 2011; Xiao et al., 1997) and fail to reach negotiated settlements because they think they can get a better deal (Babcock *et al.*, 1995; Bazerman & Neale, 1982; Neale & Bazerman, 1985). On the benefit side, people often defer to their overconfident peers (Anderson *et al.*, 2012; Murphy *et al.*, 2015), in part because overconfidence is persuasive (Schwardmann & van der Weele, 2019; Soldà *et al.*, 2019; von Hippel & Trivers, 2011). For example, overconfident job applicants are perceived as more qualified than well-calibrated ones (Ronay *et al.*, 2019), and overconfident negotiators generate higher gains than well-calibrated ones *when they reach an agreement* (Benos, 1998; Kyle & Wang, 1997). Indeed, the more people demand the more they get (Galinsky & Mussweiler, 2001; Moore, 2004; White & Neale, 1994), and greater confidence leads to greater demands (Kramer *et al.*, 1993; McGillicuddy *et al.*, 1984; Thompson & Loewenstein, 1992).

One way to reconcile the costs and benefits of overconfidence is to consider whether a contested outcome is perceived in *relative* or *absolute terms*. Absolute benefits matter, of course, but relative outcomes loom large whenever people jockey for status, which is determined entirely by how one compares to others (Brosnan & De Waal, 2003; Buss, 1989). As a consequence, people will sometimes reduce their absolute outcomes if that sacrifice enables them to improve their ranking relative to others (Charness *et al.*, 2014). In the case of overconfidence, the costs tend to manifest in absolute terms (e.g., failure or injury), but the benefits emerge primarily in relative or interpersonal terms (Anderson *et al.*, 2012; Murphy *et al.*, 2015; Murphy *et al.*, 2018). Although relative and absolute benefits typically go hand-in-hand, relative benefits can accrue at an absolute cost (e.g., when a competition injures or depletes both parties but yields a clear winner).

In the current investigation we explore the hypothesis that overconfidence yields relative gains even in the presence of absolute costs. To test this possibility, we examine the costs and benefits of high levels of confidence in a *distributive negotiation* – a form of dispute resolution in which win/win or pie-expanding outcomes are not possible. Distributive negotiations emerge whenever people compete over assets that have equal value to both parties, and hence negotiators are prevented from arriving at a mutually satisfying resolution in which each party chooses their preferred assets. Although many negotiations have the potential to yield integrative, or win/win

solutions, distributive solutions are nonetheless common, as both sides often fail to recognize win/win opportunities (Thompson & Hrebec, 1996). Such outcomes are particularly likely when competing parties perceive their interests as diametrically opposed (Bazerman & Neale, 1983). For example, despite the fact that both workers and management are highly dependent on the success of their company, they often see their interests as being in direct opposition to each other. When their disagreements are sufficient to lead workers to strike, both parties enter distributive negotiations with the added quality that the time spent to reach a settlement becomes an additional and substantial cost to both parties. This form of “shrinking pie” negotiation is the canonical model of bargaining in economics (Rubinstein, 1982) – as it captures the costs of both delays and failure to reach a settlement – and is the situation we simulated in the current research.

Such negotiations are an example of “mixed motives” games, as they are characterized by a blend of cooperative incentives (reaching a deal) and conflictual incentives (reaching a better deal) (Schelling, 1961). This blend of incentives creates a trade-off between being conciliatory and intransigent. In such a context, overconfidence may simultaneously generate relative benefits and absolute costs. By enhancing intransigence, overconfidence might improve ones’ outcome relative to others while simultaneously introducing delays and reducing the chances of a successful negotiation, to the detriment of both parties.

To test this possibility, we adapted an experimental design from the game “Divided” (van Dolder *et al.*, 2015), in which pairs of participants must agree on how to allocate a prize resulting from a joint effort. The prize can only be allocated unequally, to avoid the common default to share any joint outcome 50/50, and participants negotiate over who gets the larger share. To isolate the causal role of confidence in this context, we chose to manipulate confidence rather than measure it as an individual difference variable. In service of this goal, prior to the negotiation we provided “noisy” feedback to participants, which indicated they were likely to have performed either better or worse than their partner.

There are two important points to make with regard to our manipulation of confidence: First, our experiment focuses on the form of overconfidence in which people over-place their performance relative to others (the second type of overconfidence defined by Moore and Healy, 2008). Second, by manipulating confidence we have removed the aspect of *overconfidence* that is clearly an error in judgment; our ‘overconfident’ participants have been told they probably outperformed their

partner, and hence are arguably engaged in a Bayesian integration of the feedback they have received. For this reason, from this point forward we use the more neutral terms *high* and *low levels of confidence* to refer to levels of over- or under-confidence that emerged as a result of our manipulated feedback.

Method

The experimental design and hypotheses were pre-registered on AsPredicted.¹ Note that we have occasionally clarified the hypotheses for expositional purposes and we relegated the hypotheses and analyses on agreement failures to the SOM-R (section 2). No data point was excluded from the analyses.

Participants: We recruited a total of 298 participants via Hroot (Bock *et al.*, 2014), mainly among students from local engineering, business, and medical schools. The experiment was conducted over a series of 21 sessions that involved an average of 14 participants per session, and participation took place in GATE-lab (Ecully, France). Overall, 54% of the participants were female and the average age was 23 years (SD = 5.48).

Participants were paid the sum of their earnings for each phase in addition to a five-euro show-up fee. The experiment took an average of one hour and the average payoff was 15.71 euros (SD = 6.72). Participants received their payment in private at the end of the experiment. Our intent was to run 300 participants. With that sample size, the minimum detectable effect size with statistical power at the recommended .80 level was Cohen's $d=0.32$ for mean comparisons between participants who received a good signal vs. a bad signal and Cohen's $d=0.46$ for mean comparisons between the four possible combinations of signals (Cohen, 2013).

Procedure and Measures: The experiment was programmed using o-Tree (Chen *et al.*, 2016) and was composed of four parts, which we refer to as the 'individual phase', the 'partner phase', the 'manipulation phase', and the 'negotiation phase'. We used the individual phase to match participants in pairs. In the partner phase, both participants in a pair answer general knowledge questions to build a joint prize. In the manipulation phase, we elicit participants' beliefs about their performance in the partner phase relative to their partner's performance. In the negotiation phase, pairs of participants negotiate the allocation of their shared prize. The unfolding of the experiment

¹ Link: <https://aspredicted.org/bj9er.pdf>

is displayed in Figure 1. Translated experimental instructions are available in the SOM-U (section 1.2).

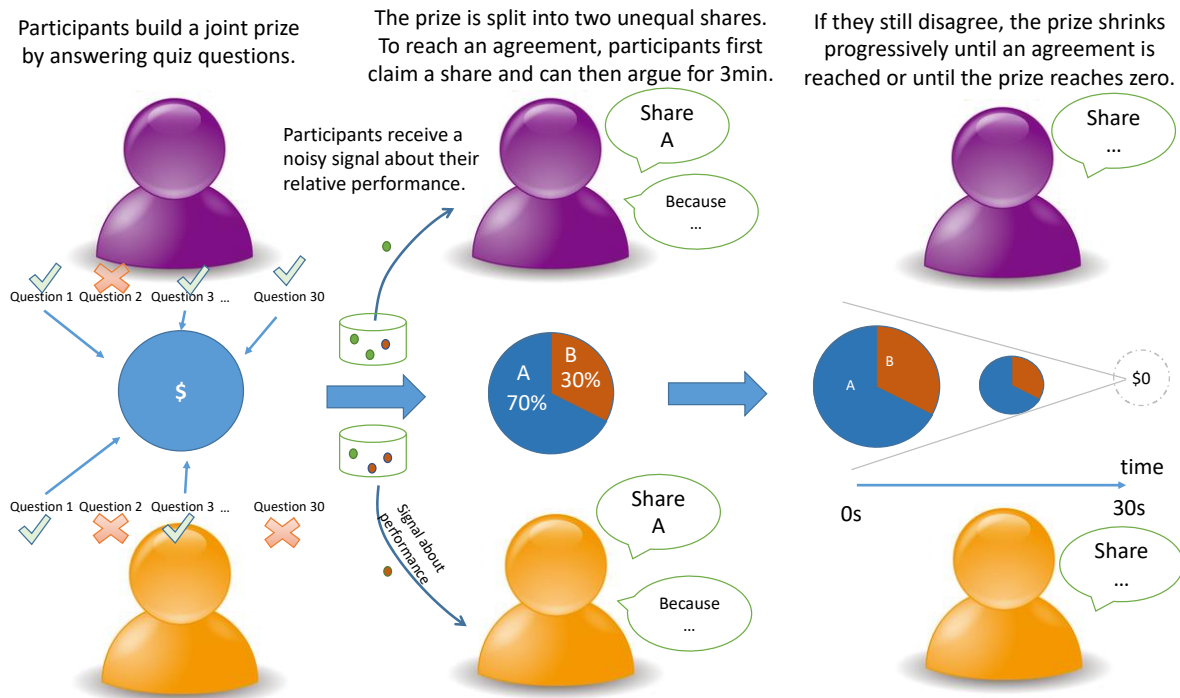


Fig. 1. Schematic representation of the experimental design. Schematic representation of the experimental design. Participants take a quiz and their correct answers contribute to a joint prize. After the quiz, they receive noisy feedback about who is likely the best performer in the pair. The prize money is then divided into two unequal shares and participants claim the share they want. If they claim the same share, they can argue via chat for up to three minutes, at which point the prize starts shrinking. The value of the prize diminishes steadily to zero after 30 seconds if no agreement is reached.

Individual phase. Participants answered ten general knowledge questions individually. For each question, they choose the correct answer from among four options. Participants received 0.2 euro for each correct answer. At the end of the individual phase, participants were ranked according to their performance on the quiz. The participant with the highest score was ranked 1 and the participant with the lowest score was ranked n (with n , the total number of participants in the

session). Participants were not informed of their rank, and they only receive information about their score and their payoff for this stage at the end of the experiment.

Partner phase. Participants were matched in pairs according to their rank: The participant ranked n was matched with the participant ranked $n - 1$, the participant ranked $n-2$ was matched with the participant ranked $n-3$, and so on until all participants were paired. This pairing procedure was common knowledge among participants and was intended to clarify that both members are able to contribute to the joint prize more or less equally. Participants then answered 30 general knowledge questions individually. The questions used in both stages of the experiment are available in the SOM-U (section 1.3). As in Part 1, the questions were the same for all participants and they were to choose the correct answer from four options. Participants received 0.67 euro for each correct answer and the money earned by both participants in each pair was allocated to a joint account. To prevent participants from inferring each other's performance from the value of their joint account, the productivity p of the pair $\{i, j\}$ was randomly drawn from a uniform distribution $U(0.75, 1.25)$. If we denote c_i the number of correct answers of participant i and c_j the number of correct answers of participant j from the pair $\{i, j\}$, the value v of the joint prize of the pair $\{i, j\}$ is computed as follows: $v_{\{i,j\}} = 0.67 * p_{\{i,j\}} * (c_i + c_j)$.

Manipulation phase. After participants completed the 30 general knowledge questions, we elicited their beliefs about their absolute and relative performance in the partner phase. First, participants were asked to report their beliefs about the number of questions they answered correctly in the partner phase. Participants received 1 euro if their estimate was exact or deviated from their true performance by only one question, and 0.50 euro if their estimate deviated from their true performance by two questions. If the estimate deviated by more than two questions, they did not earn or lose anything. Participants were then asked how likely it is that they outperformed their partner in the partner stage. Participants indicated their belief on a scale from 0 to 100% on a slider without incentives.

We then manipulated participants' belief about their relative performance by giving them a private 'noisy' binary signal, using a procedure adapted from Schwardmann and van der Weele (2019). The 'noisy' feedback is designed to create random variation in participants' confidence about their relative contribution, without deception. Each participant was shown two cyber-urns containing twenty balls of two different colours (red and green) and was told that the computer program

selects a ball from one of these two urns. If the participant performed better than his partner in the partner phase, the ball is drawn from the urn with fifteen green balls and five red balls. If the participant performed worse than his partner in the partner phase, the ball is drawn from the urn with five green balls and fifteen red balls. Therefore, a participant who outperformed his partner is more likely to see a green ball and a participant who was outperformed by his partner is more likely to see a red ball.

After the ball was shown to participants, we then elicited their beliefs about their relative performance in the partner phase again. After this final belief elicitation, the value v_{ij} of the joint prize was displayed on the screen and participants were asked to decide how to share their prize via a three-stage negotiation process. Participants had been told that they would split the prize in the partner phase, but were only given instructions on the details of the split and the negotiation procedure after the last belief elicitation.

Negotiation phase. Participants were informed that their joint prize has been divided into two unequal shares. Their task is to reach an agreement on the allocation of these shares. The large share is equal to 70% of the joint prize ($0.7v_{ij}$) and the small share is equal to 30% of the joint prize ($0.3v_{ij}$). The negotiation process is divided into three stages. Participants have the opportunity to reach agreement in each of the three stages. However, v_{ij} decreases in Stage 3. The unfolding of the stages was described to participants before they entered the negotiation process.

In Stage 1, participants were asked to claim either the high share or the low share and to write a message to their partner to justify their choice. There was no time constraint in this stage. If the negotiators from the same pair claimed different shares, an agreement was reached: the participant who claimed the large share received $0.7v_{ij}$ and the participant who claimed the small share received $0.3v_{ij}$. In this case, the negotiation process ended in Stage 1 and participants did not enter Stage 2 or Stage 3. If both negotiators claimed the large share in Stage 1, they proceeded to Stage 2. Both participants would also enter Stage 2 if they both choose the small share in Stage 1, however, this situation never occurred in our experiment.

In Stage 2, participants who did not agree in Stage 1 were given three additional minutes to try to reach an agreement. During these three minutes, participants could communicate via a chat box with their partner. The communication within pairs was only restricted in two ways: participants were not allowed to reveal the colour of the ball that was shown to them nor any private

information that would remove their anonymity. They were reminded of the amount allocated to each share, their own decision in Stage 1, and their partner's decision in Stage 1. They could decide to switch from the large share to the small share at any time by hitting the corresponding button on their screen. An agreement was reached when one of the negotiators in the pair switched from the large share to the small share. In this case, the participant who claimed the large share received $0.7v_{ij}$, the participant who claimed the small share received $0.3v_{ij}$, and the negotiation ended. If no agreement was reached within the allocated time, participants proceeded to Stage 3.

In Stage 3, participants were given thirty additional seconds to try to reach an agreement. For each second spent in this stage, the value of joint prize and hence amount in each share decreased linearly and proportionally, such that both shares were equal to zero at the end of the thirty seconds. Participants could observe on their screen the value of the shares decreasing in real time (i.e., shrinking every second). The shares stopped shrinking when one participant chose the small share. In this case, the participant who chose the large share received the remaining amount allocated to the large share, and the participant who switched to the small share received the remaining amount allocated to the small share. If no one switched before the end of the thirty seconds, both negotiators received nothing, and the total value of the joint prize was lost.

At the end of the negotiation phase, participants completed a demographics questionnaire in which they were asked to report their gender, age, and their risk preferences. We elicit risk preferences by asking participants to indicate how willing there were to take risks in general on a scale from 0 to 10 (Dohmen *et al.*, 2005).

Results

For each pair of participants $\{i, j\}$ we have four possible combinations of signals: i received a bad signal/ j received a good signal, both i and j received a bad signal, both i and j received a good signal, i received a good signal/ j received a bad signal. The payoffs associated with each of these combinations are displayed Table 1, with absolute outcomes depicted in the left panel and relative outcomes depicted in the right panel. Absolute payoffs are measured as the percentage of the initial prize received by each participant after the negotiation. This measure incorporates the possible loss of prize money in the negotiation process: When the prize money shrinks, the sum of absolute payoffs progressively approaches zero. In contrast, relative payoffs are measured as the share of

the final prize (i.e., the remaining amount of the joint prize at the end of the negotiation) received by each participant. Relative payoffs only reflect how well players do compared to each other, not relative to the initial amount they could have received.

The average absolute payoffs were lower when both participants received good signals than when both received bad signals. For pairs of participants who received opposite signals, the participant with the good signal received more than the participant with the bad signal in both relative and absolute terms.

Table 1. Players' payoffs as a function of their own signal and the signal of their partner.

| Absolute Payoffs | | | | Relative Payoffs | | | |
|------------------|------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|--|--|
| Player j | | | | Player j | | | |
| Signal | | Good | Bad | Good | Bad | | |
| Player i | Good | 26.96% 26.96% (n=46) | 35.89% 44.46% (n=102) | 50.00% 50.00% (n=46) | 44.67% 55.33% (n=102) | | |
| | Bad | 44.46% 35.89% (n=102) | 39.18% 39.18% (n=48) | 55.33% 44.67% (n=102) | 50.00% 50.00% (n=48) | | |

Identification Strategy

Because a participant's confidence in our experiment is likely to be correlated with unobserved characteristics (for instance dominance, extraversion, dispositional overconfidence, etc.), we cannot rely on raw confidence levels to establish a causal relationship between participants' confidence and their outcome in the negotiation stage. Instead, to avoid the pitfalls of endogeneity, we used the exogenous variation in confidence that emerged from the noisy component of the signal given to participants. The signal is informative about the true state of the world (i.e., whether the participant performed better than her partner) because it is accurate in 75% of cases. Hence, it is no surprise that it shifted participants' beliefs (see Figure 2).

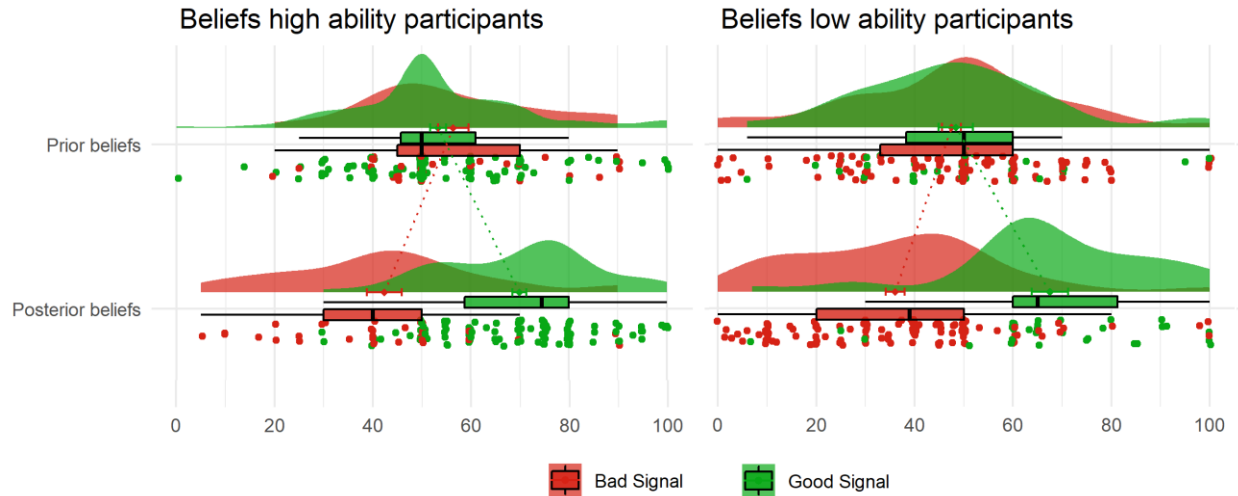


Fig. 2. Distribution of prior and posterior beliefs for participants who are the best performer in the pair (top panel) and participants who are not the best performer in the pair (bottom panel).

Nevertheless, the signal observed by a participant is completely random and exogenous *conditional on the true state of the world* (i.e., a participant gets a good signal randomly with 75% probability if he/she was the best in the pair and with 25% probability otherwise). Because we know whether the participant was the best performer of the pair, we can perfectly control for whether they outperformed their partner by using the noisy signal as an instrumental variable (Woodbridge, 2010) for participants' posterior beliefs.²

Figure 2 shows the distribution of posterior beliefs separately for the better and worse participants and whether they received a good or a bad signal. For both types of participants, we found no difference in beliefs prior to the signal (Mann-Whitney tests (best): $Z = 0.597$; $p = 0.551$;³ Somer's $d = -0.0264$; with 95% CI = [-0.117 ; 0.0643]; (not best) $Z = 0.114$; $p = 0.909$; Somer's $d = -0.00484$; with 95% CI = [-0.867 ; 0.0770]) and a strong difference in beliefs after observing the

² Indeed, when we enter performance and the signal into a simultaneous regression predicting confidence, being the best only has an impact on confidence prior to seeing the signal. Once people receive the noisy signal, being the best has no remaining impact on their confidence (see Table S2.9 in SOM-U).

³ All p-values are two tailed throughout.

signal (Mann-Whitney tests (best): $Z = -6.416$; $p < 0.001$; Somer's $d = 0.271$ with 95% CI = [0.187 ; 0.356]; (not best) $Z = -6.442$; $p < 0.001$ Somer's $d = 0.264$ with 95% CI = [0.182; 0.346]).⁴

These results show that participants are very reactive to the signal, making it a strong instrument for our subsequent analyses. To ensure that the feedback could only influence outcomes through its impact on private beliefs, participants were not allowed to discuss their feedback directly with their partner. These features of the design ensured the validity of our instrumental analysis.

Negotiation Outcomes

Next, we implemented the instrumental variable strategy described above to assess whether the patterns observed in Table 1 are due to the causal effect of participants' confidence on their negotiation outcomes. In service of this goal, we estimated the effect of both participants i and j from the pair $\{i, j\}$ instrumented by the signals received on participant i 's absolute and relative outcome using 3-stage least square (3SLS) regressions. The marginal effects from the 3SLS regressions of a participant's increase in confidence conditional on her partner's confidence on both her absolute and relative payoff from the negotiations are displayed in Figure 3.

The upper panel of Figure 3 shows that an increase in participant i 's confidence leads to an increase in her absolute payoff when her partner j has low confidence (belief $j = 20\%$, see the blue line). However, when the partner j has high confidence (belief $j = 80\%$), participant i is penalized for having higher confidence (see the red line). This interaction between the participant's confidence and her partner's confidence is significant ($\beta = -0.0148$, $Z = -2.67$, $p = 0.008$, 95% CI = [-0.0257, -0.0039]), regardless of whether the participant is the best performer of the pair. These results confirm that the pattern observed in the left panel of Table 1 is driven by variation in confidence generated by the noisy feedback participants received rather than by some combination of pre-existing beliefs of the participants.

⁴ If we assume that the mapping of beliefs to responses is symmetric for good and bad news, Figure 2 also suggest an asymmetry in the assimilation of good vs. bad news relative to the Bayesian benchmark (illustrated by a flatter dotted line from the mean prior belief to the mean posterior belief for participants who received a bad signal). We confirm this pattern in SOM-U Table 2.7, using parametric analyses of belief updating (Eil & Rao, 2011; Möbius et al., 2014; Sharot et al., 2012).

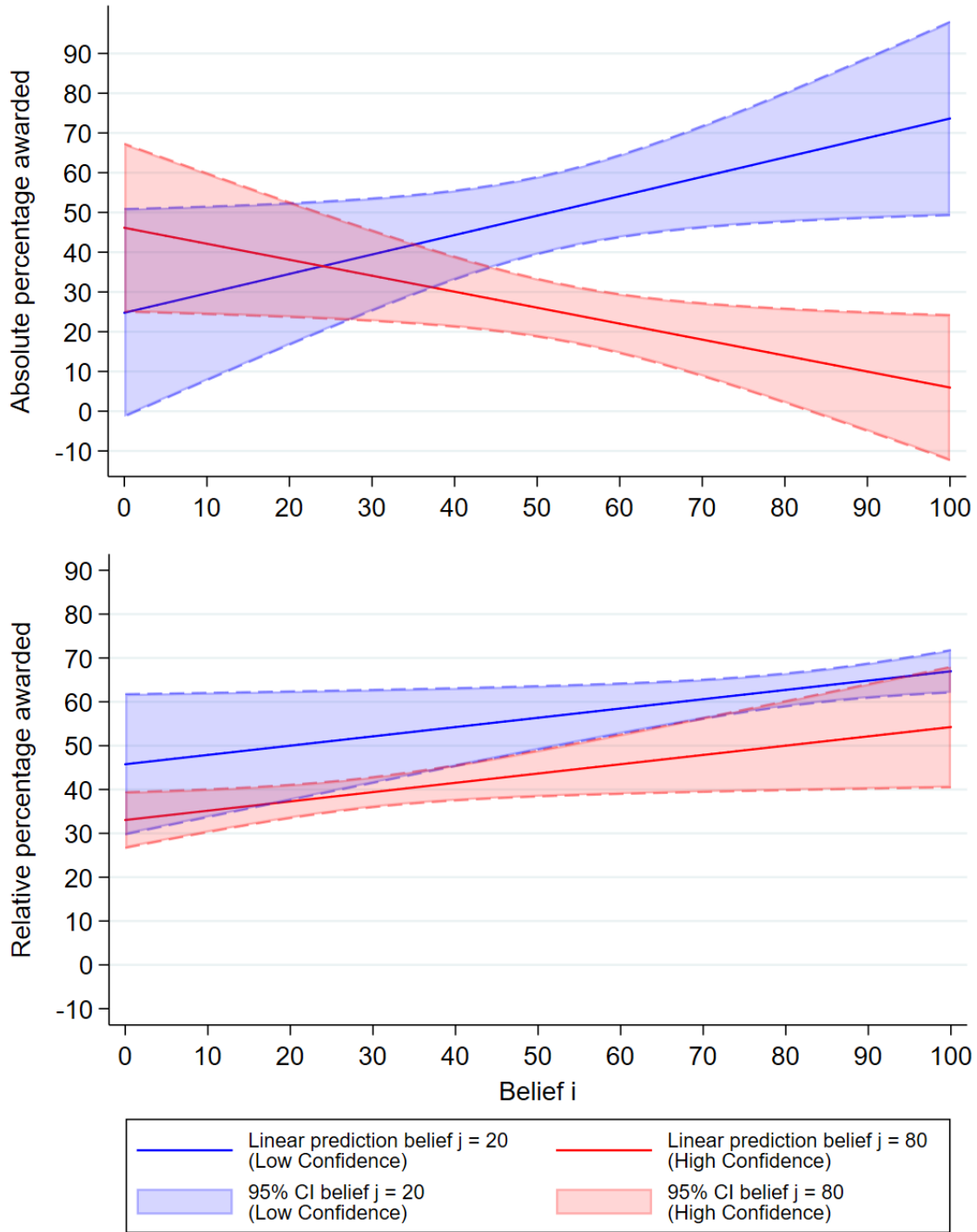


Fig. 3. Linear prediction (with 95% confidence interval) of the effect of participant i 's belief on participant i 's absolute (upper panel) and relative (lower panel) payoffs when participant j 's belief is low (red line) or high (blue line). The estimates are from models (1) and (9) in Table S1 of the SOM-R, respectively.

In contrast to these findings with the absolute payoff, the relative payoffs are symmetric when both participants have the same degree of confidence (either high or low), with no such interaction in the lower panel of Figure 3. The effect of a participant's own confidence is significant ($\beta = 0.212$, $Z = 2.22$, $p = 0.026$, 95% CI = [0.0250, 0.399]), and having greater confidence than one's partner has a positive effect on relative payoffs secured ($\beta = 0.216$, $Z = 14.69$, $p < 0.001$, 95% CI = [0.188, 0.245]), regardless of whether the participant is the best performer of the pair. These estimates confirm that the pattern observed in the right panel of Table 1 reflects the causal effects of beliefs on payoffs. Overall, these results suggest that higher levels of confidence can be beneficial in relative terms, even when they come at a cost in absolute terms. Table S2.3 in the SOM-U shows that these results hold when excluding pairs of participants who agreed in Stage 1, suggesting that the results are not just driven by participants who did not interact with their partner.

Agreements

Other aspects of the negotiation process also suggest that high levels of confidence within pairs of negotiators lead to conflictual negotiations and hence to a smaller final prize. Figure 4 displays the percentage of agreements reached in each stage of the negotiation. Overall, 6.04% of the pairs reached an agreement in Stage 1, 36.24% in Stage 2 (seconds 1 to 180) and 42.95% in Stage 3 (seconds 181 to 210). 14.77% of pairs did not reach an agreement at all and came up empty-handed. The spikes in agreement around 180 seconds suggests that, perhaps unsurprisingly, most pairs agreed either at the end of Stage 2 (14.77%) or immediately when the shares start to shrink in Stage 3 (32.89%). Figure 4 also shows that delays or failures to reach an agreement occurred frequently, suggesting a substantial amount of the initial prize was lost in the negotiation process. Such delays and failures are driven in part by pairs of negotiators who are both high in confidence. The left panel of Table 1 shows that the average percentage of the initial prize awarded at the end of the negotiation among pairs of participants who received two good signals is 26.43% lower than the share awarded to pairs of participants who received signals of opposite valence (Mann-Whitney test: $Z = 2.453$, $p = 0.014$, Somers' $d = -0.12$ with 95% CI = [-0.217, -0.0136]).

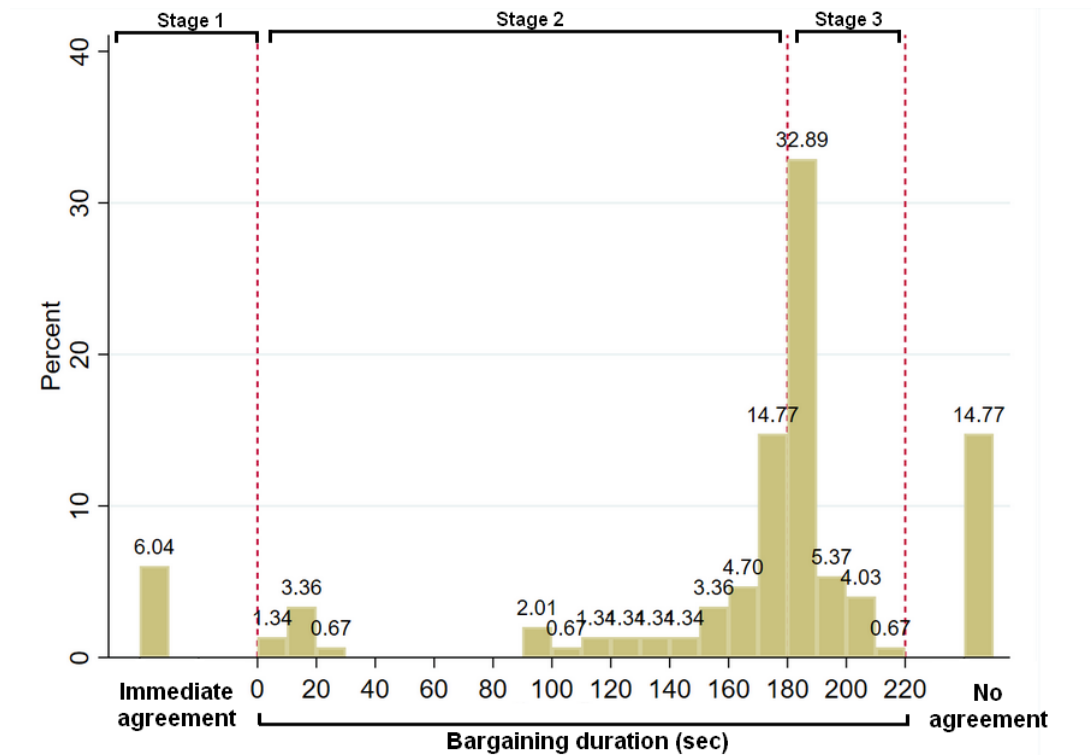


Fig. 4. Distribution of the time needed to reach an agreement (in seconds) across all pairs (n=149). Stage 1: participants only state their preferred share. Stage 2: participants can argue about their preferred share. Stage 3: the shares shrinks to zero over 30 seconds if participants do not reach an agreement.

To further investigate the causal effect of confidence on the outcome of the negotiation, we estimated the effect of the sum of participants’ beliefs on the percentage of the initial prize that remained to be shared after the negotiation. We instrumented the sum of beliefs with the noisy signals using two-stage least squares (2SLS) regression, as the noisy signals create exogenous variation in the overall sum of confidence in the pair. Results are displayed in Table 2. Models (1) and (2) in Table 2 show that a 10-percentage point increase in confidence at the pair level leads to a 4 percentage point decrease in the percentage of the initial prize that is awarded after the negotiation ($\beta = -0.422$, $Z = -2.33$, $p = 0.020$, 95% CI = [-0.778, -0.0671]). Models (3) and (4) show a similar effect when considering only pairs of negotiators who reached an agreement before the end of the negotiation process ($\beta = -0.363$, $Z = -3.01$, $p = 0.003$, 95% CI = [-0.599 to -0.127]), suggesting that losses are not only driven by agreement failures but by delays as well. These results

provide evidence that high levels of confidence lead to conflict that is detrimental to the outcome of the negotiation process.

Table 2. Effect of confidence on the outcome of the negotiation.

| Dependent var.: | Remaining percentage of the initial prize | | | |
|--|---|--------------------|------------------|------------------|
| | (1) | (2) | (3) | (4) |
| Belief_i + Belief_j | -0.401 | -0.412 | -0.340 | -0.362 |
| p-values | 0.022 | 0.034 | 0.003 | 0.005 |
| 95% CI | -0.744 to -0.0587 | -0.0793 to -0.0311 | -0.562 to -0.118 | -0.616 to -0.108 |
| Performance | Yes | Yes | Yes | Yes |
| Demographics | No | Yes | No | Yes |
| Constant | 111.20 | 133.09 | 100.99 | 107.25 |
| p-values | <0.001 | <0.001 | <0.001 | <0.001 |
| 95% CI | 69.31 to 153.08 | 62.98 to 203.20 | 74.62 to 127.35 | 63.65 to 150.86 |
| Observations | 149 | 149 | 127 | 127 |
| F-Stat first stage | 17.66 | 10.73 | 14.67 | 9.13 |
| Chi ² | 5.62 | 14.60 | 13.42 | 17.44 |

Note: Columns (1) to (4) report the 2SLS estimates of the sum of beliefs of participant i and j from the pair $\{i, j\}$, instrumented by both i and j signals on the percentage of the initial prize that is awarded at the end of the negotiation process. Columns (1) and (2) show the results for all pairs. Columns (3) and (4) show the results only for pairs of participants who reached an agreement. We control for both participants i and j 's performance in the partner phase in all models. We control for both i and j 's demographics (gender, age, and risk preferences) in columns (2) and (4).

Communication

Finally, it is worth asking how the effects of confidence manifested themselves via the actual negotiation process.⁵ To address this question, the initial messages participants sent each other

⁵ We thank an anonymous reviewer for questions and suggestions underlying these analyses.

were coded for the presence or absence of two types of claims, which we labelled *Entitlement* (e.g., “I deserve the larger share because I performed better”) and *Assertiveness* (e.g., “I’m not going to back down and take the smaller share”). Next, three raters evaluated the *Perceived Confidence* in the messages on a 6-point scale (0=*not at all confident* to 5=*completely confident*; $Kappa = 0.24$). These initial messages that participants sent each other can be considered independent, as they had not yet read what their partner wrote. Recall that after that initial message, participants were also given a chance to chat further, and this subsequent chat was coded for confidence on the same 6-point scale. Four findings of note emerged from these analyses.

First, participants’ post-signal self-reported confidence had a positive impact on the likelihood that their messages were assertive and made claims of entitlement, and assertiveness and entitlement in turn mediated the positive effect of participants’ self-reported confidence on rater-perceived confidence in their initial message (see Figure S2.6 in SOM-U). Second, rater-perceived confidence in participants’ initial message was positively associated with rater-perceived confidence in their subsequent chat, and a positive correlation emerged between the perceived confidence in the chat of the two partners (see Figure S2.5 in SOM-U). Third, when we enter assertiveness, entitlement, and self-reported confidence into a regression predicting absolute and relative outcomes, the effects of self-reported confidence remain the same as those reported above (see Table S2.8 in SOM-U). Fourth, assertiveness mirrored the effect of confidence, in that people who made claims that they would not back down performed better in relative but not absolute terms.

Discussion

The results with absolute payoffs suggest that when both players have a high level of confidence, the situation is not in equilibrium. When one player is highly confident the other player is better off being less confident, as under those circumstances they both receive a higher payoff. In contrast, when one player is not very confident the other player is better off being highly confident, as under that circumstance the more confident player earns more. These findings suggest that in terms of absolute payoffs, high levels of confidence are beneficial in negotiation *only* when your partner is low in confidence. These findings are broadly consistent with the prior literature, which shows that overconfidence is beneficial when you reach an agreement (which is more likely when your partner is low in confidence) but costly when you do not (which is more likely when your

partner is also high in confidence). But neither this pattern of findings nor the prior literature explains why overconfidence remains so common despite these clear costs.

To answer that question, we must turn to relative payoffs. When we examine the relative outcomes, the situation where both players have high levels of confidence is in equilibrium because neither negotiator has an incentive to be less confident. If Player i is not confident, Player j is better off being confident and if Player i is confident, Player j is still better off being confident. Importantly, this positive effect of confidence is independent of whether an agreement is reached, as our experiment captured the type of costs that emerge when overconfidence creates an impasse as well as when overconfidence simply introduces delays. These results demonstrate that the goal to achieve higher *relative outcomes* will favor the development of overconfidence despite the fact that overconfidence creates *absolute costs*. Thus, our research clarifies why the obvious costs of overconfidence do not lead people to become better calibrated. By focusing on relative rather than absolute outcomes – as people often do (Buss, 1989; Charness *et al.*, 2014; von Hippel, 2018) – overconfident negotiators are prevented from feeling they have suffered a loss even when they walk away with little or nothing.

Beyond what these results reveal about overconfidence, they also elucidate the inter-relationships between competence, confidence, assertiveness, and entitlement. People who received a positive signal reported greater levels of confidence, which in turn led them to be more assertive and make greater claims of entitlement in their message to their negotiation partner. Assertiveness and entitlement caused their messages to be perceived as more confident, but controlling for assertiveness and entitlement did not impact the effect of confidence on either relative or absolute outcomes in the negotiation. As a side point, it is worth noting that the effects of assertiveness were similar to the effects of confidence; people who claimed they would not back down also performed better in relative but not absolute terms (consistent with the results of van Dolder *et al.*, 2015). The results also indicate that the effects of confidence can emerge independently of those of competence-based status, as the instrumental analyses removed the effect of being the best performer in the pair, and being the best was not predictive of confidence once participants received the noisy signal.

The results of the current experiment also provide a bridge between research on confidence, coordination problems in bargaining situations, and dominance complementarity. Prior research

has shown that a clear status hierarchy can benefit negotiators by providing cues on which they can coordinate when they have conflicting goals, thereby avoiding impasses (De Kwaadsteniet, & Van Dijk, 2010; Halevy *et al.*, 2011). In contrast, when negotiation partners both see themselves as high or low in status, this coordination advantage is lost. Our results with high and low levels of confidence yielded similar effects with regard to absolute outcomes – if one member of the negotiating pair is confident, absolute outcomes are better when the other is not, and vice-versa. Thus, different confidence levels between competing parties can facilitate solutions to coordination problems.

Similar results emerge when people experience dominance complementarity, as dyads and groups tend to perform better when their members vary in dominance than when their members are all high or low in dominance (Ronay *et al.*, 2012; Wiltermuth *et al.*, 2015). Despite these similarities across the literatures on hierarchy, dominance, and confidence, it is notable that levels of confidence only showed these complementary effects with regard to absolute outcomes. In relative terms, people were always better off being high rather than low in confidence. Given the similarities between these literatures, our results raise the possibility that hierarchy and dominance would reveal similar relative benefits and absolute costs when all members of a group are high in status or dominance.

Caveats

The current results suggest why overconfidence is common even when it is costly for everyone in absolute terms, but there are several caveats to keep in mind. First, our negotiation task was explicitly designed to incorporate a shrinking pie, with delay costly to both sides. Given the luxury of more time, it is possible that universally high levels of confidence would have been less costly. Second, our experiment was also designed to increase the probability of disagreement by mandating that the distribution be unfair, as people often reject unfair offers to protect their reputation (Nowak *et al.*, 2000). Of course, many real-life negotiations also result in outcomes that one or both sides regard as unfair, and hence in that sense the situation we created in the lab is quite common. Lastly, our experiment was also a distributive negotiation with no opportunity for a compromise that would benefit both parties. As noted above, many negotiations are inherently of this type and many fall into this pattern unnecessarily (Thompson & Hrebec, 1996), but many do not, as opportunities to expand the pie are common when negotiations involve a variety of

issues and people have diverse preferences. Thus, future research might consider whether the current findings would emerge in less contentious circumstances.

Conclusions

The results of the current research suggest that high levels of confidence can emerge as an inefficient equilibrium in negotiation: people are *relatively* better off being more confident independent of whether their partner is also confident. As a consequence, situations in which people focus on relative outcomes are likely to select for strategically inflated self-beliefs. Thus, these findings provide clear predictions that overconfidence should be most evident whenever people prioritize relative over absolute gains (as in status competitions or among people who are particularly attuned to relative standing). Nevertheless, strategically inflated self-beliefs come at a cost, as both partners earn more in absolute terms when they are less confident. By considering both relative and absolute outcomes simultaneously, the current findings provide an explanation for why overconfidence is both costly and ubiquitous, as evolution can select for traits that enhance relative success even at an absolute cost.

References

- Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology, 103*(4), 718.
- Babcock, L., Loewenstein, G., Issacharoff, S., & Camerer, C. (1995). Biased judgments of fairness in bargaining. *The American Economic Review, 85*(5), 1337-1343.
- Bazerman, M. H., & Neale, M. A. (1982). Improving negotiation effectiveness under final offer arbitration: The role of selection and training. *Journal of Applied Psychology, 67*(5), 543.
- Bazerman, M. H., & Neale, M. A. (1983). Heuristics in negotiation: Limitations to effective dispute resolution. In M. Bazerman & R. Lewicki (Eds.), *Negotiating in Organizations*. Beverly Hills: Sage (pp. 51-67).
- Benos, A. V. (1998). Aggressiveness and survival of overconfident traders. *Journal of Financial Markets, 1*(3-4), 353-383.
- Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review, 71*, 117-120.
- Brosnan, S. F., & De Waal, F. B. (2003). Monkeys reject unequal pay. *Nature, 425*(6955), 297-299.
- Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12*(1), 1-14.
- Charness, G., Masclet, D., & Villeval, M. C. (2014). The dark side of competition for status. *Management Science, 60*(1), 38-55.
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88-97.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- De Kwaadsteniet, E. W., & Van Dijk, E. (2010). Social status as a cue for tacit coordination. *Journal of Experimental Social Psychology, 46*(3), 515-524.

Dohmen, T. J., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2005). Individual risk attitudes: New evidence from a large, representative, experimentally validated survey, DIW Discussion Papers, No. 511, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2), 114-38.

Galinsky, A. D., & Mussweiler, T. (2001). First offers as anchors: the role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology*, 81(4), 657.

Halevy, N., Y. Chou, E., & D. Galinsky, A. (2011). A functional model of hierarchy: Why, how, and when vertical differentiation enhances group performance. *Organizational Psychology Review*, 1(1), 32-52.

Kramer, R. M., Newton, E., & Pommerenke, P. L. (1993). Self-enhancement biases and negotiator judgment: Effects of self-esteem and mood. *Organizational Behavior and Human Decision Processes*, 56(1), 110-133.

Kyle, A. S., & Wang, F. A. (1997). Speculation duopoly with agreement to disagree: Can overconfidence survive the market test? *The Journal of Finance*, 52(5), 2073-2090.

McGillicuddy, N. B., Pruitt, D. G., & Syna, H. (1984). Perceptions of firmness and strength in negotiation. *Personality and Social Psychology Bulletin*, 10(3), 402-409.

Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). Managing self-confidence: Theory and experimental evidence (No. w17014). National Bureau of Economic Research.

Moore, D. A. (2004). Myopic prediction, self-destructive secrecy, and the unexpected benefits of revealing final deadlines in negotiation. *Organizational Behavior and Human Decision Processes*, 94(2), 125-139.

Murphy, S. C., Barlow, F. K., & von Hippel, W. (2018). A longitudinal test of three theories of overconfidence. *Social Psychological and Personality Science*, 9(3), 353-363.

Murphy, S. C., von Hippel, W., Dubbs, S. L., Angilletta Jr, M. J., Wilson, R. S., Trivers, R., & Barlow, F. K. (2015). The role of overconfidence in romantic desirability and competition. *Personality and Social Psychology Bulletin*, *41*(8), 1036-1052.

Neale, M. A., & Bazerman, M. H. (1985). The effects of framing and negotiator overconfidence on bargaining behaviors and outcomes. *Academy of Management Journal*, *28*(1), 34-49.

Nowak, M. A., Page, K. M., & Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science*, *289*(5485), 1773–1775.

Ronay, R., Greenaway, K., Anicich, E. M., & Galinsky, A. D. (2012). The path to glory is paved with hierarchy: When hierarchical differentiation increases group effectiveness. *Psychological science*, *23*(6), 669-677.

Ronay, R., Oostrom, J. K., Lehmann-Willenbrock, N., Mayoral, S., & Rusch, H. (2019). Playing the trump card: Why we select overconfident leaders and why it matters. *The Leadership Quarterly*, *30*(6), 101316.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, 97-109.

Schelling, T. C. (1961). Experimental games and bargaining theory. *World Politics*, *14*, 47.

Schwardmann, P., & Van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, *3*(10), 1055-1061.

Sharot, T., Kanai, R., Marston, D., Korn, C. W., Rees, G., & Dolan, R. J. (2012). Selectively altering belief formation in the human brain. *Proceedings of the National Academy of Sciences*, *109*(42), 17058-17062.

Shipman, A. S., & Mumford, M. D. (2011). When confidence is detrimental: Influence of overconfidence on leadership effectiveness. *The Leadership Quarterly*, *22*(4), 649-665.

Soldà, A., Ke, C., Page, L., & von Hippel, W. (2020). Strategically delusional. *Experimental Economics*, *23*, 604-631.

Thompson, L., & Hrebec, D. (1996). Lose–lose agreements in interdependent decision making. *Psychological bulletin*, *120*(3), 396.

Thompson, L., & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes*, 51(2), 176-197.

van Dolder, D., van den Assem, M. J., Camerer, C. F., & Thaler, R. H. (2015). Standing united or falling divided? High stakes bargaining in a TV game show. *American Economic Review*, 105(5), 402-07.

von Hippel, W. (2018). *The Social Leap: The new evolutionary science of who we are, where we come from, and what makes us happy*. HarperCollins: NY.

von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34(1), 1-16.

White, S. B., & Neale, M. A. (1994). The role of negotiator aspirations and settlement expectancies in bargaining outcomes. *Organizational Behavior and Human Decision Processes*, 57(2), 303-317.

Wiltermuth, S., Tiedens, L. Z., & Neale, M. (2015). The benefits of dominance complementarity in negotiations. *Negotiation and Conflict Management Research*, 8(3), 194-209.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Xiao, Y., Milgram, P., & Doyle, D. J. (1997). Capturing and modeling planning expertise in anesthesiology: Results of a field study. In C. E. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (pp. 197–205). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Supplementary Online Material (Reviewed)

1. Main text Analyses

To investigate the causal effect of the participants' confidence on their payoffs received from the negotiation, we regressed both the absolute payoff (as a percentage of the initial joint prize) and the relative payoff (as a percentage of the final prize)¹ on the posterior beliefs of participants i and j about their relative contribution to the joint prize, as well as the interaction between the two players' posterior beliefs.

Table S1. Effect of beliefs on participants' payoff from the negotiation.

| Dependent var: | Absolute percentage awarded to participant i | | | | | | Relative percentage awarded to participant i | | | | | |
|--|--|--------------------|------------------|------------------|-----------------|-----------------|--|-------------------|-------------------|-------------------|------------------|------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Belief_i | 0.786 | 0.802 | 0.0071 | 0.0024 | - | - | 0.212 | 0.249 | 0.212 | 0.184 | - | - |
| p-values | 0.020 | 0.021 | 0.957 | 0.986 | | | 0.522 | 0.460 | 0.026 | 0.072 | | |
| 95% CI | 0.126 to 1.445 | 0.122 to 1.481 | -0.249 to 0.263 | -0.264 to 0.269 | | | -0.437 to 0.861 | -0.412 to 0.910 | 0.0250 to 0.399 | -0.0161 to 0.384 | | |
| Belief_j | 0.356 | 0.382 | -0.422 | -0.423 | - | - | -0.212 | -0.142 | -0.212 | -0.208 | - | - |
| p-values | 0.273 | 0.257 | 0.001 | 0.002 | | | 0.519 | 0.671 | 0.022 | 0.029 | | |
| 95% CI | -0.281 to 0.994 | -0.280 to 1.045 | -0.682 to -0.163 | -0.672 to -0.382 | | | -0.856 to 0.432 | -0.798 to 0.514 | -0.393 to -0.0312 | -0.394 to -0.0218 | | |
| Belief_i*Belief_j | -0.0148 | -0.0154 | - | - | - | - | 0.0000 | -0.0013 | - | - | - | - |
| p-values | 0.008 | 0.009 | | | | | 1.000 | 0.839 | | | | |
| 95% CI | -0.0257 to -0.0039 | -0.0269 to -0.0039 | | | | | -0.0118 to 0.0118 | -0.0133 to 0.0108 | | | | |
| Belief_i-belief_j | - | - | - | - | 0.219 | 0.229 | - | - | - | - | 0.216 | 0.204 |
| p-values | | | | | 0.040 | 0.034 | | | | | <0.001 | <0.001 |
| 95% CI | | | | | 0.0100 to 0.428 | 0.0177 to 0.440 | | | | | 0.188 to 0.245 | 0.166 to 0.241 |
| Best | -4.629 | -4.015 | -4.629 | -4.189 | -4.971 | -5.133 | -5.304 | -4.428 | -5.304 | -4.444 | -5.639 | -5.006 |
| p-values | 0.508 | 0.571 | 0.434 | 0.478 | 0.410 | 0.399 | <0.001 | 0.015 | <0.001 | <0.001 | <0.001 | <0.001 |
| 95% CI | -18.35 to 9.090 | -17.91 to 9.884 | -16.22 to 6.966 | -15.76 to 7.380 | -16.79 to 6.849 | -17.06 to 6.791 | -8.464 to -2.144 | -7.978 to -0.877 | -7.325 to -3.283 | -6.802 to -2.086 | -7.038 to -4.239 | -6.917 to -3.095 |
| Performance | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Demographics | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes |
| Constant | 20.24 | 24.74 | 58.80 | 64.94 | 43.29 | 50.83 | 52.65 | 50.75 | 52.65 | 54.02 | 52.82 | 53.14 |
| p-values | 0.290 | 0.260 | <0.001 | <0.001 | <0.001 | <0.001 | 0.007 | 0.015 | <0.001 | <0.001 | <0.001 | <0.001 |
| 95% CI | -17.25 to 57.74 | -18.33 to 67.82 | 38.23 to 79.37 | 37.34 to 92.55 | 25.78 to 60.80 | 25.36 to 76.29 | 14.49 to 90.82 | 9.882 to 91.61 | 31.41 to 73.89 | 28.16 to 79.88 | 35.75 to 69.89 | 30.22 to 76.06 |
| Obs. | 298 | 298 | 298 | 298 | 298 | 298 | 298 | 298 | 298 | 298 | 298 | 298 |
| R-Squared | -0.0830 | -0.0815 | 0.0023 | 0.0064 | 0.0022 | 0.0007 | 0.0129 | 0.0345 | 0.0129 | 0.0353 | 0.010 | 0.0311 |
| Chi2 | 22.33 | 22.71 | 14.66 | 15.16 | 8.29 | 9.80 | 17.17 | 26.96 | 17.17 | 27.00 | 397.50 | 245.99 |

Note: Table S1 shows the effect of participant i 's and j 's beliefs on participant i 's absolute payoffs (columns (1) to (6)) and relative payoffs (columns (7) to (12)). Columns (1) and (2) display the 3SLS estimates of participant i 's and j 's beliefs, as well as their interaction on participant i 's absolute payoff. Columns (3) and (4) display the same estimates after removing the interaction term. Columns (5) and (6) display the 2SLS estimates of the difference in beliefs between participant i and j on participant i 's absolute payoff. Columns (7) to (12) display the same set of estimates using participant i 's relative payoffs as the dependent variable. We control for participant i relative performance (i.e. whether participant i was the best performer in the pair in the partner phase) in all models. Control variables include participant i and j 's absolute performance in the partner phase, as well as participant i 's demographics (gender, age and risk preference). We use 1000 bootstrap replications with samples of one draw per pair.

¹ Both participant i and j 's relative payoff are set to 50% for pairs who failed to reach an agreement.

We instrumented each participants' posterior belief in a pair (*belief_i* or *belief_j*) by a dummy variable equal to 1 if a good signal was observed, and 0 otherwise. In addition, we instrumented the interaction between the two players' beliefs by the combination of signals received by the pair $\{i,j\}$. Because payoffs within a pair are dependent, as they are determined by only one participant (who switched from the high share to the low share), we use 1000 bootstrap replications with samples of one draw per pair (i.e., for each replication, one participant from each pair is randomly selected as an observation). To control for the potential correlation between the error terms of the three equations, we estimate the models using 3SLS. The estimated coefficients are displayed in columns (1) and (7) in Table S1. Columns (2) and (8) show the same estimation with additional control variables (*i*'s gender, age, and risk preference). To ensure the validity of our instrumental procedure, we control in all regressions for a dummy variable equal to 1 if participant *i* was the best performer in the pair in Part 2, and 0 if participant *j* was the best performer. In addition, we control for both participant *i* and *j*'s performance in Part 2.

Because the interaction term is not significant in the context of relative payoff, we can remove it from the equation.² We re-estimate models (7) and (8) without the interaction term and the new estimates are displayed in columns (9) and (10), respectively. Because the coefficients associated with participant *i* and *j*'s posterior beliefs are perfectly symmetric, we can also estimate a joint coefficient using the difference in beliefs between participant *i* and participant *j* as the main independent variable. Column (11) shows the estimates of the 2SLS regression of participant *i*'s relative payoff on the difference in beliefs between participant *i* and participant *j*, instrumented by both participant *i*'s and participant *j*'s signals. Column (12) displays the same estimates, when adding the same control variables as in models (8) and (10).

For completeness, we also performed the same estimations as in columns (9) to (12) using the absolute payoff received by participant *i* as the dependent variable. The estimates are displayed in columns (3) to (6). However, these models do not consider the existence of an interaction between participant *i*'s and participant *j*'s beliefs and should be interpreted in this light.

2. Pre-registered Analyses

This section reports the pre-registered analyses that are missing from the main text, as they are secondary to our primary research question.

2.1 Confidence and likelihood to concede

We pre-registered that we would investigate the effect of confidence on the likelihood of accepting the small share. Table S2 displays the percentage of participants who accept the small share for pairs that agreed in (1) Stage 2, (2) in Stage 1 or 2, in (3) Stage 3, and (4) overall, conditional on the signal they received.³

Participants who received a good signal are less likely to accept the small share in Stage 2 than those who received a bad signal and the difference remains significant when adding participants who agreed in Stage 1 (two-tailed tests of proportion:⁴ Stage 2: $z = 4.25$, $p < 0.001$, Somers' $d = -0.41$ with 95% CI = [-0.582, -0.233], Stage 1 or 2: $z = 4.83$, $p < 0.001$, Somers' $d = -0.43$ with 95% CI = [-0.587, -0.270]). This difference disappears in Stage 3, when the share starts to shrink (PR test: $z = -0.01$, $p = 0.993$, Somers' $d = -0.01$ with 95% CI = [-0.174, 0.176]), which might be driven by the toughest negotiators being self-selected to enter the

² The fact that relative payoffs are perfectly symmetric (50% each) when both participants have the same degree of confidence (either high or low) is a strong indication that there should be no interaction between participants' beliefs within the same pair.

³ We pool observation from Stage 1 and 2 due to the very small number of pairs who reached an agreement in Stage 1 (9 pairs only).

⁴ PR test hereafter.

last stage. Over all stages, among all the pairs that researched agreement, those who received a good signal are significantly less likely to accept the smaller share than participants who received a bad signal (PR test: $z = 3.39$, $p < 0.001$, Somers' $d = -0.21$ with 95% CI = [-0.334,-0.092]). Probit regressions of the likelihood to switch from the large share to the small share on the difference in beliefs between participants i and j from a pair $\{i,j\}$, instrumented by the signals received by both i and j produce similar results (see SOM-U section 2.2).

Table S2 – Proportion of participants who accept the small share, by stage and signal received.

| Signal | % of participants who accept the small share in | | | |
|--------|---|---------------------|------------------|------------------|
| | Stage 2 | Stage 1 or 2 | Stage 3 | All stages |
| B | 68.97%*** (6.13) | 69.57%*** (5.58) | 49.18% (6.45) | 60%*** (4.31) |
| G | 28% (6.41) | 26.32% (5.88) | 49.25% (6.15) | 38.71% (4.39) |
| Obs. | 108 | 126 | 128 | 254 |

*Note: Table S2 shows the proportion of participants who concede to the small share in different stages of the negotiation process, conditional on the signal they received. Standard errors in parentheses. Asterisks indicate the results of two-sample tests of proportions. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.*

2.2 Confidence and agreements

To investigate the causal effect of confidence on the likelihood to reach an agreement, we pre-registered that we would estimate a probit regression of a dummy variable equal to 1 if the pair reached an agreement before the end of the negotiation process (and 0 otherwise), on the difference in posterior beliefs between participants i and j instrumented by both i 's and j 's signals. We also pre-registered that we would estimate a 2SLS regression and a Poisson regression of the stage at which an agreement is reached on the difference in posterior beliefs between participants i and j , instrumented by both i 's and j 's signals. We later realized that the estimations of these models had some limitations. These limitations are discussed in section 2.4 of the SOM-U where we propose new estimation strategies.