

SCIENTIFIC REPORTS



OPEN

Motives relate to cooperation in social dilemmas but have an inconsistent association with leadership evaluation

Christian Wolff ^{1,2} & Nina Keith²

A common assumption is that good leaders are driven by a power motive that motivates them to influence others. However, leaders need to restrain themselves in social dilemmas where cooperation maximizes collective outcomes. We theorize that in social dilemmas, a desire for positive relationships (affiliation motive) is more beneficial than a power motive because it draws attention away from short-term self-interest towards understanding others. In a game of *Settlers of Catan* in the laboratory, we find that a functional variant of the affiliation motive relates to verbal encouragement of cooperation, to fewer occurrences of oil spills, to higher ratings of transformational leadership and, in a field survey, to fewer selfish business decisions. Furthermore, a dysfunctional variant of the power motive relates to two of three indicators of selfishness. Group members perceive selfish individuals as assuming leadership roles which indirectly relates to slightly higher ratings of transformational leadership. This pattern of evaluation may privilege men who, on average, show more selfish behaviour which can be partially attributed to their motives. Mere awareness of gender-based discrimination does not enable raters to circumvent this pattern of evaluation. This work suggests a need for interventions that increase appreciation of cooperative leaders.

Mastering social dilemmas is essential. Humans achieve many kinds of progress through cooperation in social dilemmas^{1,2}, such as when business departments share their knowledge to create a useful product³. In social dilemmas, each party is best off in the short term by minimizing their personal costs⁴ whereas in the long term, wisely chosen cooperation can pay off way beyond its initial costs⁵ and increases prosperity.

In social dilemmas, leader behaviour matters. Leaders often have discretion to choose a course of action^{6,7}. Different leaders make different choices⁸. Here we argue that leaders' choices are affected by their motives. Motives are stable preferences for particular classes of states or activities. Leaders' behaviour in social dilemmas is of high importance for their organizations even beyond the scope of a single situation⁹. It sparks lasting reciprocity from stakeholders and observers¹⁰, signals trustworthiness¹¹, builds reputation⁵, and maintains existing relationships. Moreover, leaders serve as role models for followers who imitate them and who are inspired by leaders with integrity¹².

Motives shape how people understand social dilemmas. Resolving social dilemmas requires an unbiased understanding of the situation^{13,14}. Actors need to (i) recognize interdependencies in the distribution of everyone's outcomes and (ii) anticipate what others will do¹⁵. Empathy (being able to perceive others' mental states¹⁶) allows both—an appraisal of dilemma outcomes from the perspective of others^{17–20} and, based on that, the drawing of inferences about others' intentions¹⁷. Empathy itself depends heavily on motivation¹⁷. Some motives have excitatory or inhibitory effects on it. In this way, motives can determine behaviour in social dilemmas.

The affiliation motive and cooperation. Here we first propose that an affiliation motive relates positively to cooperation. The affiliation motive refers to a desire to build and maintain positive relationships. Affiliation attracts people to situations in which they can connect with others¹⁷ and motivates them to attend to others'

¹Work and Organizational Psychology, University of Bamberg, Bamberg, Germany. ²Organizational and Business Psychology, Technische Universität Darmstadt, Darmstadt, Germany. Correspondence and requests for materials should be addressed to C.W. (email: christian.wolff@uni-bamberg.de)

mental states¹⁷. In social dilemmas, attending to others enables an accurate understanding of the situation, which in turn fosters cooperation^{18–22}.

However, we limit this proposition to a *functional* variant of the affiliation motive. We theorize that those high in this variant particularly enjoy being considerate and cooperative rather than being popular or being validated by others. This conceptualization deviates from prior research, which often cast affiliation in a negative light^{23,24}. Affiliation motivated individuals have often been seen as desperately wanting to be liked, fearing rejection, avoiding conflicts, and favoring their in-group at the expense of everyone else²³. In this study, we refer to this as the *dysfunctional* affiliation motive and include it only for comparison. We do not expect the dysfunctional affiliation motive to relate to caring and trusting nor, in turn, to cooperation. We use the terms *functional* and *dysfunctional* to allude to the assumed implications that each motive variant was theorized to have for social interactions, especially in the context of leadership^{23,25–28}.

The power motive and selfishness. Second, we propose that a power motive relates positively to selfishness. The power motive refers to the desire to influence or control people or processes. Experiencing power heightens sensitivity for rewards²⁹, narrows focus of attention by suppressing constraining information³⁰, makes people play down risks³¹, and increases overconfident decisions³². In social dilemmas, this should direct individuals toward selfish choices, which typically provide the most salient rewards. Experiencing power particularly affects the processing of social information. That is, power can decrease taking others' perspectives³³ or their advice³⁴ and sometimes undermines coordination with others³⁵. People who are motivated to pursue self-interest often reduce empathy¹⁷. In this way, a strong power motive may deter individuals from recognizing how cooperation benefits everyone in the long run and what is wrong with a selfish choice. This may cause biased understandings of social dilemmas, which in turn lead to selfish behaviour.

However, we limit this proposition to a *dysfunctional* variant of the power motive. Prior research has shown that humans desire power for various purposes^{23,36,37}. We theorize that those high in the dysfunctional variant of the power motive desire power as a means to perceived superiority in an authoritarian or materialistic sense. In contrast, if an individual desires power in order to pursue a greater good, we refer to that as a *functional* power motive and include it in this study only for comparison. We do not expect the functional power motive to relate to selfishness because its other-related purpose should compensate negative effects of being motivated by power³⁸.

Previous research on motives and cooperation in social dilemmas. We believe that the present work contributes to the literature on cooperation in social dilemmas by using an alternative approach to measuring motives. For half a century now, a rich body of literature has accumulated^{14,39,40} showing that cooperation in social dilemmas can be predicted from *social value orientation* (SVO) both in the laboratory⁴¹ as well as in several field studies^{42,43}. Researchers measure SVO by asking participants to split money between themselves and a fictional stranger (called decomposed games approach). Researchers then derive a score classifying participants as prosocial, individualistic, competitive, or unclassifiable^{44,45}. This approach infers participants' social preferences indirectly from their choices.

Here we use a different approach based on self-reported motives⁴⁶. We consider two separate motive variants at once (functional affiliation motive, dysfunctional power motive). While the functional affiliation motive emphasizes on a concern for others' interests, the dysfunctional power motive focuses on self-enhancement through devaluation of others. Examining both motives simultaneously may help gauging each motive's relative importance for cooperation in social dilemmas. Furthermore, being able to distinguish these motive variants from related motive variants (dysfunctional affiliation motive, functional power motive) may sharpen our understanding of the boundaries of each motive variant. Previous research has already predicted cooperation from personality^{14,47} (which can be seen as closely related to motives⁴⁸) and from values or motives^{21,46,49–51}. However, the relative importance of multiple motives for cooperation in social dilemmas seems less clear (but see refs^{21,46}).

Previous research on motives and leadership. Furthermore, we hope that the present study contributes to the literature on the role of motives for leadership. Previous research has focused primarily on *implicit* (i.e., subconsciously activated) motives. Supplementary Table S1 provides an overview of all studies on the role of implicit affiliation and power motives for leadership or leader outcomes that we are aware of ($k = 26$ samples, $n = 2,495$ participants). This overview suggests that it is difficult to draw any overall conclusions from these studies as a whole. Whereas an early study indicated that a *low* implicit affiliation motive might be beneficial in leaders⁵², almost all of the other studies yielded contradictory^{53–56} or inconsistent^{36,57–61} results—including a reanalysis of data from the original sample⁶². Supplementary Table S1 also suggests that it is difficult to draw conclusions about the role of the implicit power motive for effective leadership. As presented in Columns 11 and 12 of Supplementary Table S1, most studies examined a specific variant of the implicit power motive (15 samples^{36,52,55,56,59,62–66}) or its combination with other motives (4 samples^{54,57,58}). While doing so, researchers used a total of 10 different operationalizations^{36,52,55,57,59,63,64,66} (cf. Columns 11 and 12 in Supplementary Table S1). This degree of heterogeneity makes it difficult to interpret individual studies or to integrate findings across multiple studies. It may also explain why motives were only included in 2 of the 15 reviews and meta-analyses on the role of individual differences for leadership since 2011⁶⁷. Both reviews did not systematically synthesize empirical research^{28,68} and one did not distinguish between functional and dysfunctional variants of motives⁶⁸.

Here we focus on *explicit* (i.e., consciously accessible) affiliation and power motives and include *implicit* motives only as control variables in one sample. We can thereby examine if theoretical assumptions about the role of affiliation and power motives for leadership^{23,26,28} apply to the *explicit* motivational system. Dual motive theory postulates that implicit and explicit motives belong to two separate motivational systems^{69–71}. While implicit motives are assumed to energize *operant* behaviour which is spontaneously enacted and driven by affect, explicit motives are expected to influence *respondent* behaviour which is subject of conscious thought and deliberation⁷².

Leadership roles may often require individuals to think through decisions and to consider factors that are imposed from outside. Both of these processes are theorized to be influenced by the explicit motivational system⁷². A large body of literature supports the importance of explicit variables⁷³ and explicit motives⁷⁴ for job performance⁷⁵ and effective leadership^{62,76}. However, we know of no measures that enable the assessment of variants of explicit affiliation and power motives as they are conceptualized in the literature^{23,26–28,77} (for existing scales, see refs. ^{37,78,79}). For this reason, we developed short self-report scales that measure functional and dysfunctional variants of affiliation and power motives. These scales provide an opportunity to study the importance of explicit motive variants for leadership outcomes.

Gender differences in motives and cooperation. Finally, we expect that men show more selfish behaviour than women and that this can be partially attributed to gender differences in motives. We know from meta-analyses that men, on average, have lower moral sensitivity than women⁸⁰, exhibit weaker deontological inclinations⁸¹, and behave more selfishly in resource dilemmas⁸². Even though there are detailed theoretical accounts for gender differences and similarities in general^{83,84}, scholars still call for more research on the specific factors that give rise to gender differences⁸⁵.

We draw on sociocultural theory⁸⁴ as an explanation for gender differences in motives. Based on historical division of labour by gender, women's assignment to the role of child care might have contributed to the development of a higher functional affiliation motive in women because of the congruency between this motive and the responsibilities associated with child care⁸⁶. Previous findings from large-sampled studies appear to support this idea. Women place, on average, higher importance than men on values that are directed towards the well-being of others such as benevolence, universalism⁸⁷, and social values in general^{88,89}. Other research shows that socio-cultural factors also underlie gender differences in competition⁹⁰ which may be intended to preserve the gender hierarchy^{91,92}. This might have contributed to the development of a higher dysfunctional power motive in men who, on average, place more value on control than women⁹³, respond more strongly to intergroup conflicts⁹⁴, and experience competition more positively⁹⁵. Based on our propositions about the roles of the functional affiliation motive and the dysfunctional power motive for cooperation, we assume that gender differences in these motives translate to gender differences in cooperation.

The present study. To test our propositions, we observed groups during a game of *Settlers of Catan*. In this game, players need to make efficient use of resources in order to populate an uninhabited island. All players manage their own population. Players can compete or cooperate at any given time. We choose the *Oil Springs* iteration of this game, which creates a resource dilemma by providing the option to use oil. Using oil allows players to extend their empire, but gradually destroys the island and its resources⁹⁶. We set financial incentives that were intended to activate motives related to both cooperation and selfishness. Participants knew that after the game, a coin toss determined whether they received a payment based on group performance (intended to activate motives related to cooperation) or on individual performance (intended to activate motives related to selfishness). A total of 201 individuals participated in groups of 3 to 4 players who hardly knew each other before meeting in the laboratory.

During the game, which lasted about 75 min, we videotaped the whole conversation. Communication about a dilemma often increases cooperation rates¹³. As a measure of verbal encouragement of cooperation, we count and aggregate all statements that favor either cooperation (e.g., “let us avoid using oil”) or selfishness (e.g., “everyone should look out for themselves”, inverse coded).

Results

Motives relate to cooperation in social dilemmas. Figure 1a shows that the functional affiliation motive is positively related to encouragement of cooperation, $\beta = 0.25$, $P = 0.0009$, whereas the dysfunctional power motive is not significantly related to encouragement of cooperation, $\beta = -0.14$, $P = 0.054$.

We also count how many oil spills an individual causes during the game. Oil spills typically inflict lasting damage to one or more group members. We find that the functional affiliation motive is related to fewer occurrences of oil spills, $\beta = -0.25$, $P = 0.0009$, whereas the dysfunctional power motive is related to more oil spills, $\beta = 0.23$, $P = 0.0013$ (Fig. 1b).

As another test of both propositions, we conducted a field survey. We recruited 961 individuals online ($M_{\text{age}} = 31$ y, $s.d. = 12$) who read six business scenarios. Each scenario describes a social dilemma requiring a decision between personal benefits and preventing harm to society or the environment. Figure 1c shows that the functional affiliation motive is negatively related to selfish business decisions, $\beta = -0.20$, $P < 0.0001$, whereas the dysfunctional power motive is positively related to selfish business decisions, $\beta = 0.44$, $P < 0.0001$. Many respondents of this study are in leadership positions ($n = 257$) or have other kinds of work experience ($n = 446$). For all subsamples, similar results emerge (see Supplementary Table S2). This suggests that with regard to selfish business decisions, our results seem to generalize across different occupational statuses.

In both studies, all relationships involving the functional affiliation motive remain significant after we account for important control variables (e.g., personality traits, other motives, reasoning ability, Supplementary Fig. S1, Supplementary Table S3). This indicates that the functional affiliation motive accounts for aspects of human behaviour in the investigated social dilemmas beyond the predictive validity of established traits. Furthermore, we include a dysfunctional variant of the affiliation motive and a functional variant of the power motive in all analyses but find no substantial relationships to the outcomes so far described (Supplementary Table S3). Full correlation matrices are provided in Supplementary Datasets 1 and 2.

Gender differences in motives and cooperation. Next, we examine gender differences in cooperation and in the motives underlying it. If such analyses highlight strengths that are—on average—associated with

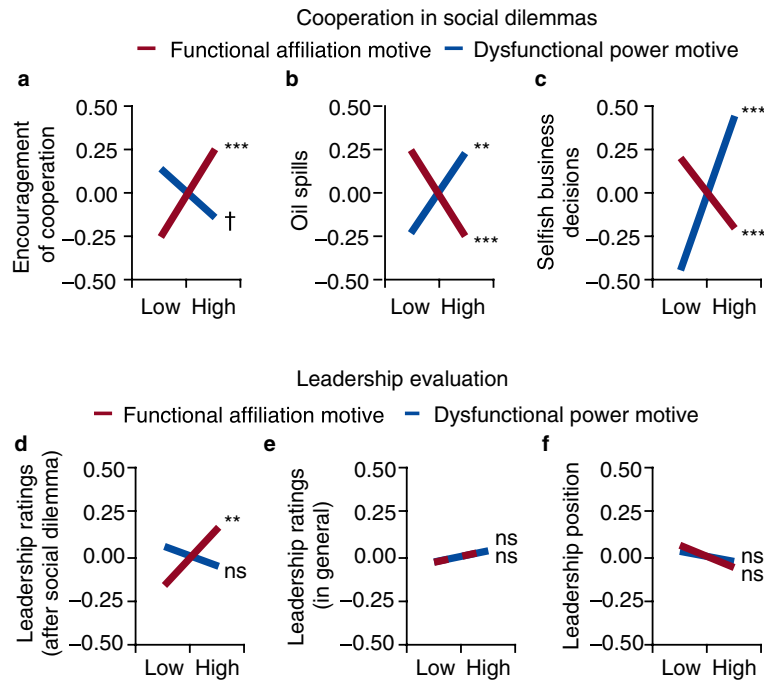


Figure 1. Motives relate to cooperation in social dilemmas but have little impact on leadership evaluations. (a,b) Motives relate to behaviour during a game of *Settlers of Catan* ($n = 201$). Conversations are videotaped and any statements favoring cooperation (positive values) or selfishness (negative values) are counted, log-transformed, aggregated using equal weights, and then aggregated over two independent observers ($r = 0.71$). By using oil, players cause oil spills, which damage the fictitious island of Catan. We count how many oil spills each player causes. (c) Motives relate to selfish decisions in a field survey ($n = 960$). Respondents read six business scenarios each posing a social dilemma. (d) After the game of *Settlers of Catan*, players rate each other on transformational leadership. (e) In the field survey, 739 peers rate the general leadership competence of 486 respondents. (f) Respondents state whether they hold a professional leadership position. All values on y axes are z-standardized. Lines represent slopes from multiple regression analysis while controlling for a dysfunctional affiliation motive and a functional power motive (Supplementary Tables S3, S4). Low/high ± 1 s.d. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.10$, two-sided t -tests. ns, not significant.

the female gender, a finding like this might contribute towards deconstructing male leadership stereotypes^{97,98}. Figure 2 shows that women encourage cooperation more consistently than men, $d = 0.40$, $P = 0.006$, whereas men are more than 500% more likely to cause an oil spill (39 of 45 are caused by men), $d = -0.67$, $P < 0.0001$, and make business decisions that are more selfish, $d = -0.47$, $P < 0.0001$ (see Supplementary Table S5 for more detail). These gender differences in cooperative behaviour can partially be attributed to gender differences in motives. More specifically, women report a stronger functional affiliation motive than men in both the laboratory ($d = 0.64$) and field study ($d = 0.39$) whereas men report a stronger dysfunctional power motive than women ($ds = -0.49$ and -0.32 , respectively). Gender differences in these motives are largely consistent across different occupational statuses (Fig. 3, Supplementary Table S5) with five of six effect sizes reaching statistical significance. Indirect effects of gender on cooperation via motives are reported in Supplementary Table S6.

Motives hardly affect leadership evaluation. How do motives relate to leadership evaluations from others? On the one hand, those motivated to cooperate should receive *positive* evaluations given that they cooperate more and cooperation tends to be effective in our studies (e.g., cooperation enables high group performance in the game of *Settlers of Catan*, Supplementary Information, Section 4; Supplementary Fig. S3). On the other hand, those motivated to cooperate might receive *negative* evaluations because people hold stereotypes about leadership that expect leaders to be dominant rather than cooperative^{97,99,100} (Supplementary Information, Section 7). One way to express dominance in social dilemmas is through selfish behaviour such as causing oil spills.

We assess three types of leadership evaluation (Fig. 1d–f). In the laboratory, each player rates all other group members in transformational leadership after the game of *Settlers of Catan*, resulting in a total of 582 ratings (each ratee receives 2 to 3 ratings from group members—depending on group size—which we then average within each ratee). In an attempt to make sure that ratings are based specifically on behaviour during the game, we control for baseline ratings from before the game. We find that individuals with a strong functional affiliation motive tend to receive slightly higher ratings from their group members, $\beta = 0.17$, $P = 0.0024$ (Fig. 1d). In contrast, the dysfunctional power motive is unrelated to ratings of transformational leadership, $\beta = -0.06$, $P = 0.28$, despite its positive relationship to the occurrence of oil spills. In the field survey, 739 peers rate the general leadership competence of 486 respondents ($M = 1.52$ peer-ratings per respondent, s.d. = 0.75). Here we obtain null results for

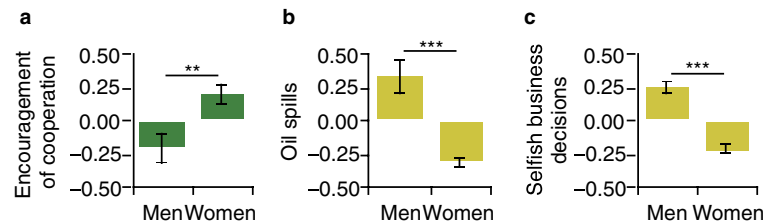


Figure 2. On average, women cooperate more than men. (a) Women ($n = 103$) encourage cooperation more consistently than men ($n = 98$), $d = 0.40$. (b) Men are 583% more likely to cause an oil spill than women (39 vs. 6), $d = -0.67$. (c) Men's ($n = 448$) decisions in business scenarios are more selfish than women's ($n = 512$), $d = -0.47$. Bars represent z-standardized means ± 1 s.e.m. Raw values are presented in Supplementary Table S5. *** $P < 0.001$, ** $P < 0.01$, two-sided t -tests.

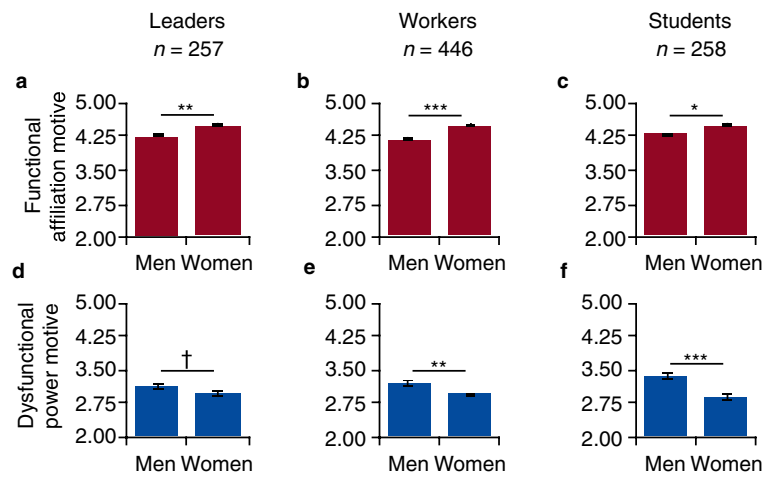


Figure 3. Across occupational groups, women report a stronger functional affiliation motive and men report a stronger dysfunctional power motive. (a,d) Respondents state whether they currently hold a professional leadership position or, if they are not working anymore, held one in the past. (b,e) Respondents without leadership position but with work experience (3 mo to 52 y, mean = 9 y, s.d. = 11). (c,f) Respondents are mostly students or homemakers (mean age = 22 y, s.d. = 5). All data are from the field survey ($n = 961$). Scales range from 1 “does not apply at all” to 6 “fully applies.” Bars represent means ± 1 s.e.m. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, † $P < 0.10$, two-sided t -tests.

both motives (Fig. 1e). Neither the functional affiliation motive, $\beta = 0.03$, $P = 0.47$, nor the dysfunctional power motive, $\beta = 0.03$, $P = 0.50$, matter (despite their relationship with selfish business decisions). Instead, a functional power motive is important, $\beta = 0.25$, $P < 0.0001$ (Supplementary Table S4). We obtain similar results with respect to the occupancy of a professional leadership position (Fig. 1f). Again, neither the functional affiliation motive, $\beta = -0.06$, $P = 0.074$, nor the dysfunctional power motive, $\beta = -0.03$, $P = 0.40$, matter. Instead, a functional power motive is important, $\beta = 0.21$, $P < 0.0001$. In line with raters' partial indifference towards selfishness, women do not, on average, receive more positive ratings than men after the game of Settlers of Catan, $d = -0.06$, $P = 0.67$, despite behaving more cooperatively on average (Supplementary Table S5).

Why do raters not always appreciate cooperativeness? After the game of Settlers of Catan, players also rate if they think that other players assumed a leadership role. We find that selfish players (those who cause oil spills) tend to be perceived as assuming a leadership role, $\beta = 0.14$, $P = 0.015$. Being perceived as assuming a leadership role, in turn, relates positively to ratings of transformational leadership, $\beta = 0.48$, $P < 0.0001$, so that a small indirect effect of oil spills to transformational leadership (via assumed leadership role) emerges, $\beta = 0.069$, $z = 2.35$, $P = 0.019$, 95% CI [0.007, 0.147] (Fig. 4b). This indirect effect somewhat offsets negative evaluations of oil spills, $\beta = -0.17$, $P = 0.0002$, so that, overall, no strong or significant relationship exists between oil spills and ratings of transformational leadership, $\beta = -0.10$, $P = 0.057$ (Fig. 4a), despite the havoc that oil spills are causing.

Exploratory analyses indicate that only some raters show this pattern of evaluation. More specifically, it is those with *high* awareness of gender-based discrimination (conditional effects for +1 s.d.) who do not seem to particularly appreciate cooperators. Instead, they tend to rate oil-spill causing group members—of which 84% are male—as assuming a leadership role, $\beta = 0.20$, $P < 0.0001$ (Fig. 4f), which results in a small positive indirect effect on transformational leadership, $\beta = 0.061$, s.e.m. = 0.019, 95% CI [0.028, 0.103] (Fig. 4d). In contrast, those with *low* awareness of gender-based discrimination (conditional effects for -1 s.d.) do not rate oil-spill causing group members as assuming a leadership role, $\beta = 0.00$, $P = 0.92$ (Fig. 4e), so that we find no significant indirect effect on transformational leadership, $\beta = 0.002$, s.e.m. = 0.035, 95% CI [-0.071, 0.065] (Fig. 4d). Instead, these

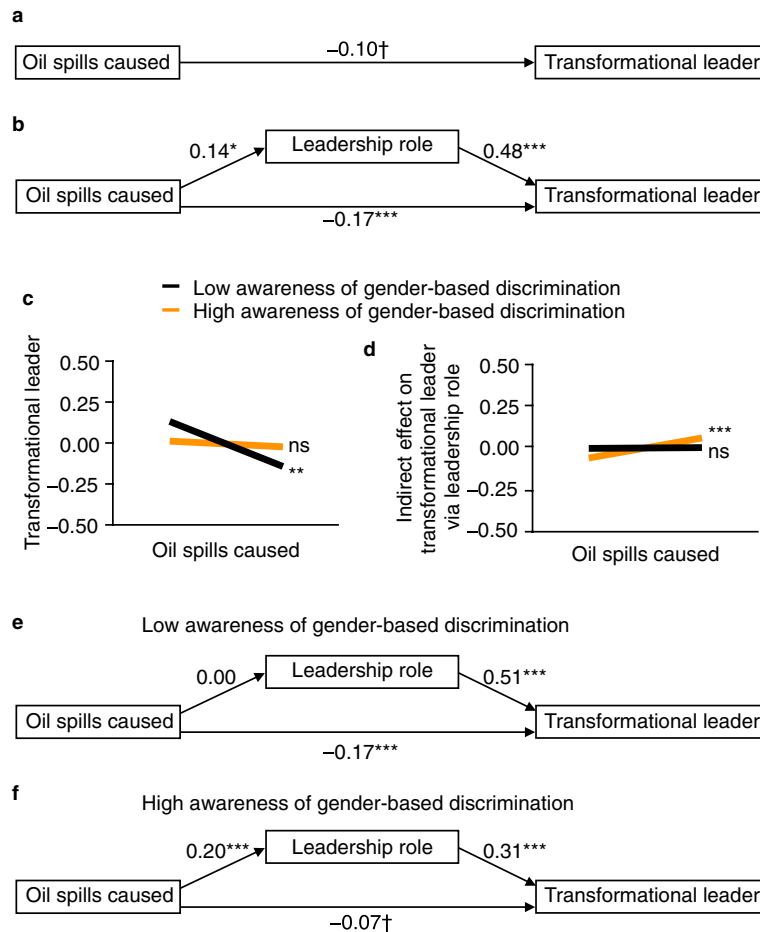


Figure 4. Perceivers think that selfish players assume a leadership role, which relates positively to their ratings of transformational leadership. **(a)** Overall, no substantial relationship exists between the number of oil spills an individual causes and the ratings of transformational leadership that individual receives from his/her group members after the game of *Settlers of Catan*, even though oil spills are very harmful. **(b)** Perceivers think that selfish players assume a leadership role, which in turn relates positively to leadership ratings and offsets more negative evaluations of oil spills ($n = 201$ players). **(c–f)** Awareness of gender-based discrimination itself promotes this stereotypical pattern of evaluation ($n = 582$ dyads, displayed are conditional effects). **(c,e)** Only group members with *low* (-1 s.d.) awareness of gender-based discrimination disapprove of (predominantly male) players who cause oil spills. **(d,f)** In contrast, group members with *high* ($+1$ s.d.) awareness of gender-based discrimination rate (predominantly male) players who cause oil spills as assuming a leadership role and, in turn, as transformational leaders. All coefficients are z -standardized. $***P < 0.001$, $**P < 0.01$, $*P < 0.05$, $^\dagger P < 0.10$, two-sided t -tests. ns, not significant.

raters with *low* awareness of gender-based discrimination tend to rate oil-spill causing group members as low on transformational leadership, $\beta = -0.17$, $P = 0.0001$ (total effect = -0.14 , $P = 0.004$) which raters with *high* awareness of gender-based discrimination do not seem to do to a statistically significant extent, $\beta = -0.07$, $P = 0.063$ (total effect = -0.02 , $P = 0.68$; Fig. 4c; Supplementary Information, Section 5; see Supplementary Information, Section 7 for a discussion of these results).

Discussion

The present work examines the role of motives for cooperation and finds that a functional affiliation motive positively relates to cooperation and that a dysfunctional power motive negatively relates to two of three indicators of cooperation. On average, women cooperate more than men and this can be partially attributed to women's higher functional affiliation motive and their lower dysfunctional power motive. All results involving the functional affiliation motive are consistent across dependent variables, across subsamples with different occupational statuses (including leaders), and after accounting for a number of relevant control variables (including motivation to lead, fairness, reasoning, and implicit motives). Results involving the dysfunctional power motive are less consistent after accounting for relevant control variables. When interpreting these findings, readers should keep in mind limitations such as the short length of the motive measures and their suboptimal psychometric properties (discussed below and in Supplementary Information, Section 2).

Motives and cooperation in social dilemmas. We believe that this article adds information to the literature on social dilemmas. The present study uses a different approach to measuring motives than most studies on social dilemmas^{41,45}. This may promote further insights into the nature of the motive variants that are relevant to cooperation in social dilemmas. We find that a concern for others' interests (functional affiliation motive) relates to cooperation independently of a focus on self-enhancement through devaluation of others (dysfunctional power motive). Findings involving the dysfunctional power motive are less robust than those for the functional affiliation motive. These motive variants must not be confused with a striving for harmonious relationships (dysfunctional affiliation motive) or a desire to pursue power for the greater good (functional power motive) which we control in our study, thereby delineating a boundary condition of our propositions¹⁰¹. We believe that this approach clarifies a detail about the specific nature of the motives that relate to cooperation in social dilemmas.

Motives and effective leadership. We assume that the present work contributes to the literature on the role of motives for leadership. A functional affiliation motive seems to be desirable in leaders whereas a dysfunctional power motive tends to be undesirable—at least if one considers cooperation to be essential for effective leadership^{8,12,102–104}. So far, affiliation and power motives were left out of most of the recent reviews and meta-analyses on the role of individual differences for leadership⁶⁷. To the best of our knowledge, there has been no definitive answer which motive variants are desirable in leaders^{23,36,52–66} (cf. Supplementary Table S1 for an overview of previous studies on the role of implicit affiliation and power motives for leadership). Maybe more research is needed that clarifies the roles of affiliation and power motives (both implicit and explicit) for different criteria of effective leadership such as ratings or cooperation. The present study may contribute to this endeavour. It introduces an economic alternative to previous measurements of motive variants and it considers multiple criteria of effective leadership simultaneously. This allows us to examine if general theoretical propositions about the role of motive variants for leadership (i.e., that a power motive is desirable whereas an affiliation motive is not²³) apply to the *explicit* motivational system. The present research finds no evidence that these propositions apply with regard to cooperation or with regard to ratings in a situation requiring cooperation (i.e., in the laboratory study). Hopefully, the introduction of *explicit* measures of motive variants will promote the accumulation of empirical findings that can be aggregated more easily than findings from previous studies using implicit measures (cf. Columns 11–13 in Supplementary Table S1). In future research, it may also be interesting to examine how variants of affiliation and power motives relate to other criteria of effective leadership such as entrepreneurial success⁷⁶ which have not been considered in this study. Furthermore, it is possible that situational factors (such as crises or cultural values) and follower characteristics (such as parenting or empathy) moderate the effectiveness of motives in leaders^{105,106}.

Motives and evaluations of leaders. This work emphasizes how little approval cooperators gain. Approval from peers influences what activities people enjoy and, in turn, what behaviour they perpetuate in the future¹⁰⁷. By examining the link between cooperation and leadership evaluations, this study draws attention to a critical pattern of evaluation—the tendency to approve of antisocial forms of influence such as causing oil spills. This is consistent with previous research which found that overconfidence¹⁰⁸ and norm violations¹⁰⁹ signal status to others.

Being perceived as assuming a leadership role appears to be a central mechanism underlying positive evaluations of selfish behaviour. Understanding this mechanism offers two potential targets for interventions: first, the link between selfishness and perceived leadership emergence and second, the link between perceived leadership emergence and transformational leadership. Scholars have questioned if leadership emergence is always aligned with the effectiveness of a team or an organization¹¹⁰. Leadership emergence—i.e., becoming influential in the eyes of others¹¹¹—should not be equated with effective leadership¹¹², and even less so in the short term^{113–115}. Future research may develop interventions that base on these findings in order to weaken the link between leadership emergence and ratings of leader effectiveness, potentially reducing negative consequences from so called over-emergence¹¹⁶ due to selfish behaviour.

Selfish behaviours such as causing oil spills are in line with masculine leadership stereotypes^{97,100} and are enacted more often by men in this study. Given that mere awareness of gender-based discrimination does not preclude raters from patterns of evaluation that are in line with masculine leadership stereotypes^{97,100}, we call for the development of specific interventions that help individuals reflect on their evaluation of selfish vs. cooperative behaviour in leaders. We also suggest that organizations lead by example and publicly convey their appreciation of cooperative leaders and their disapproval of selfishness¹¹⁷. Such messages are expected to attract cooperative individuals into leadership positions^{118–121}, which has been found to increase overall levels of cooperation¹⁰² and may reduce discrimination against leaders who do not fit a masculine leadership stereotype^{91,92,122}. While previous research has suggested to foster a *general* power motive among women¹²³, we see no indication in our data to prioritize such an approach (given that men reported a stronger *dysfunctional* power motive whereas both genders reported similar levels on the *functional* power motive). Finally, we do not wish to say that cooperative individuals will necessarily be the most successful leaders in all circumstances. It may be possible that cooperative leaders generate short-term costs with regard to, for instance, negotiation outcomes¹²⁴ or firm innovation¹²⁵.

Strengths and weaknesses. This study has some strengths and weaknesses. In the laboratory study, we assessed two behavioural measures as our outcome variables (verbal encouragement of cooperation, number of oil spills caused). This was in response to calls for research that applies multiple methods and includes behavioural measures^{126,127}. Both outcomes did not rely on self-report and occurred at critical junctures during the interaction in groups¹²⁸ (Supplementary Fig. S3, Supplementary Information, Section 4). At the same time, these measures are likely to contain a substantial amount of error variance. In line with this criticism, we only find a small and non-significant relationship between the dysfunctional power motive and encouragement of

cooperation ($\beta = -0.14, P = 0.054$). The other hypothesized relationships with these behavioural outcomes are also only modest in size (β s = $|0.23|$ to $|0.25|$, P s < 0.0014) indicating that large amounts of variance in these outcomes remain unexplained.

Another critical aspect of our study is our approach to measuring functional and dysfunctional variants of affiliation and power motives. We created short self-report scales because we were not aware of any other way that would have allowed us to measure the affiliation motive as it was conceptualized in previous research (i.e., to measure the *dysfunctional* variant as conceptualized in ref.²³ or to measure the *functional* counterpart to the dysfunctional affiliation motive (which might help explain inconsistent findings involving the affiliation motive^{28,36,52–61}). Existing scales for the power motive^{37,78,79} also did not match the original^{23,25,26} conceptualization of the two faces of the power motive. On the one hand, creating new scales might be perceived as a strength of this study because these scales allowed us to test propositions involving specific motive variants. These scales are now available to be used and improved by other researchers. On the other hand, these scales have suboptimal psychometric properties which limit their value. Most importantly, each scale consists of only 4 items which may have contributed to their relatively low reliability ($0.52 < \alpha < 0.75$ in the field survey; $0.54 < ICC < 0.87$ in a pilot study with monthly measurements over a quarter year). In addition, some items have considerable cross-loadings on one or more other motive variants (see Supplementary Table S9). We tested various psychometric characteristics of these scales including unidimensionality, discriminant validity, and their factorial structure which are described in Supplementary Information, Section 2. At best, an α coefficient of 0.53 (for the functional affiliation motive) sets an upper limit for that scale's validity to 0.73 (i.e., to the square-root of reliability) and leads to underestimated effect sizes. At worst, low reliability raises problems concerning the interpretation of a measure. These scales represent a first step in developing measures for functional and dysfunctional variants of affiliation and power motives. We suggest that future research builds upon these scales and creates much longer versions of them with improved psychometric characteristics.

Among the strengths of the present work are the combination of data from laboratory and field assessments (including actual leaders), leadership ratings from different sources (group members, peers), a total number of 1,161 participants (not counting the 739 peers), and the inclusion of a number of relevant control variables (implicit motives, reasoning ability, motivation to lead, fairness, and other personality characteristics).

Conclusion

We conclude that the functional affiliation motive seems to be desirable in leaders. However, those who are high in this motive (among many are women) are not always appreciated for that. Raters sometimes interpret selfish acts as leadership behaviour. Those who are aware of gender-based discrimination are not immune to this pattern of evaluation (but even appear to be at increased risk).

Materials and Methods

Participants. *Laboratory study (Settlers of Catan).* For sample size, we set a target of “200” before we started collecting data. Participants are 201 individuals (103 women) aged $M = 24$ y (s.d. = 6). Most are students (89%) majoring in psychology (51%). Some presently hold or formerly (in their last employment) held a professional leadership position (17%). We recruited participants on campus and through local advertisements. All participants received a variable payment of $M = €6.94$, s.d. = 1.89, in addition to either a fixed amount of €20 (available to all participants except psychology students) or course credit (available to psychology students) for a total duration of approx. 4 h (with an additional €2 or course credit for every 15 min beyond 4 h 10 min). Fresh organic fruits, snacks, as well as hot and cold beverages were available to participants free of charge.

Field survey. Budget (€2,000 for the final wave of recruitment) determined sample size. Respondents were 961 individuals (513 women) aged $M = 31$ y (s.d. = 12). Most of them have work experience (73%) of, on average, 9 y (s.d. = 12). Some presently hold or formerly held (if not working anymore) a professional leadership position (27%). We recruited half of them via an online labour market and the other half through local advertisements and social networks. Respondents received approx. €2.50 for 15–25 min. Respondents recruited 739 peers (439 women) who are either friends/acquaintances (43%) and family/partners (43%) of the respondents, or work together with respondents (14%). In total, we obtained one or more peer ratings for 486 of the respondents. Peers were not compensated.

Procedure of the laboratory study (Settlers of Catan). We distributed data collection over two occasions $M = 19$ days (s.d. = 30) apart from each other. At Time 1, we measured all independent variables in an online survey. At Time 2, participants came to the laboratory and interacted with other participants. We informed participants in the beginning of both occasions that they were going to be videotaped at Time 2. All participants provided informed consent online (Time 1) and with their signature (Time 2). We explicitly notified participants before we started recording video. Both cameras and two video lights were clearly visible.

After completing the survey at Time 1, participants automatically received regular emails with personalized invitations for Time 2 through a custom-coded script, until they registered for a particular date. Personalizing invitations in this way allowed us to stratify group composition. We intended that all groups contain 2 male and 2 female individuals. If multiple group members were psychology students, they were not allowed to belong to the same cohort so that most group members would not know each other. This procedure resulted in $n = 45$ complete groups with 4 members each (2 male, 2 female) and $n = 7$ smaller groups with 3 members each in case that one person did not show up. We control for group size in all analyses. The average degree of familiarity between group members was $M = 1.2$ (s.d. = 0.6) on a scale of 1 to 6.

In the laboratory, at Time 2, participants first had a group discussion about the solution to a fictitious rescue scenario (i.e., *The Desert Dilemma*¹²⁹) which lasted up to 15 min ($M = 11:00$ min, $s.d. = 2:55$). Group members could thereby get to know each other. After the discussion, group members rated each other for the first time on transformational leadership and on assuming a leadership role. They provided these ratings on individual workstations separated by dividing walls. However, they could see the backs of all other group members and tags with their first names on it which were attached at various positions. We use these ratings as a baseline measure in all analyses involving transformational leadership and perceived influence, respectively. A short break followed.

After the break, we instructed participants for a second time (the first time was at the end of the online part of this study) about the rules of the Settlers of Catan game and, in particular, about the Oil Springs iteration of this game⁹⁶. We handed over all different pieces of the game to each participant so that they could familiarize themselves with them by themselves at their workstation before sitting down with the others at a table in the center of the room with the game on it. The experimenter assured participants that they could ask about the rules of the game at any time. All questions were answered at all times as long as they were related to the understanding of the game.

In the Settlers of Catan game, all players manage their own population. The goal is to grow one's population on an island that all players share. Players earn so called *victory points* for constructing buildings, long roads, or for sequestering (instead of using) oil. To be able to build anything, players need resources which they obtain over time or by trading with other players. We chose the Oil Springs iteration of this game, which simulates the real-world issues associated with global consumption of fossil fuels⁹⁶. The Oil Springs scenario allows players to drill for oil and utilize it to grow their populations faster. All use of oil is indicated on the board so that all players are aware of it. After each fifth oil that is being used by any one of the players, an oil spill happens. Such a disaster either destroys one of the perimeters of the island and its future capacity to produce resources (approx. 80% likelihood) or causes coastal flooding which destroys all settlements located directly on coasts (approx. 20% likelihood). This creates a social dilemma of the type of a resource dilemma. While a single player benefits from using oil, the whole group suffers from deterioration of future productivity as a result of that player's oil use. The game was over after 10 rounds (40 moves in groups of 4 and 30 moves in groups of 3, $M = 76$ min, $s.d. = 26$). However, we concealed this fact from participants. Not knowing how long the game would last made it impossible for participants to anticipate the extent of future losses of productivity due to oil spills (see Supplementary Information, Section 4 for a discussion of the utility of oil use). Participants received financial incentives based on the results of the game. These incentives were intended to activate motives related to both cooperation and selfishness. Participants knew that after the game, a coin toss determined whether they received a payment based on group performance (intended to activate motives related to cooperation) or on individual performance (intended to activate motives related to selfishness). Previous research has found that activation of motives increases their impact on behaviour⁴⁶. The ambiguity that is introduced by coupling the payment to a coin toss resembles the ambiguity that occurs in real-life social dilemmas. In reality, there is often uncertainty involved whether cooperation pays off. Performance is indicated by the number of victory points a player earns during the game. All victory points exceeding a cutoff of 5 were worth €1 per point ($M = €1.13$, $s.d. = 1.47$). Supplementary Information, Section 3 describes further modifications that we made to the original procedure of the game.

After the game, all players rated each other again on transformational leadership and on assuming a leadership role. Finally, we asked participants whether they would recommend the study to others, to which 99% answered "yes". After completing all questionnaires, we compensated participants and thanked them for their contribution. If they had any questions about the study, we tried to answer them as well as we could. We only requested that they would not share any strategies or ideas with their friends, if those friends might want to participate in the study. All procedures were in line with all relevant ethical regulations described in the Ethics Code of the American Psychological Association. The Technische Universität Darmstadt institutional review board provided guidelines for study procedures. All procedures were approved by the University of Bamberg institutional review board.

Procedure of the field survey. The survey was conducted online. All scales were presented in randomized order. We used 6-point scales if not otherwise indicated. All participants provided informed consent. All scales were answered by the respondents using self-report measures except for the peer ratings of general leadership competence for which respondents' peers rated respondents' general leadership competence. All procedures were approved by the Technische Universität Darmstadt institutional review board.

Measures used in both studies. *Motives.* We define the *functional affiliation motive* as a desire for social interactions that are sincere and considerate, fostering deep and honest relationships with others. A sample item is "I wish that people like me for being sympathetic and cooperative". We define the *dysfunctional power motive* as a drive for possessing and using authority in order to serve one's personal interests. A sample item is "it pleases me to have a lot of power and influence, because there are many people that you need to keep under control". We define the *dysfunctional affiliation motive* as a striving for harmonious relationships with others that is characterized by confirmation seeking and self-effacement. A sample item is "it is very important to me to be accepted by others. Therefore I sometimes say things of which I am not convinced that they are right, but that make me look good". We define the *functional power motive* as a desire for using responsible and benevolent channels of influence. A sample item is "I enjoy to contribute something through my channels of influence that is aligned with the greater good". Supplementary Information, Section 1 provides details on item development and item selection. Supplementary Information, Section 2 describes our findings when evaluating the newly developed scales psychometrically. Supplementary Table S7 lists the final selection of items used in this work.

Control variables. We measured affective *motivation to lead* with 9 items¹³⁰. We used 6-point scales in the laboratory study and for the first $n = 203$ participants in the field survey (to keep response scales consistent across measures) but changed to the original 5-point format for the next 758 participants in the field survey (to be able to provide unpaid respondents with norm-based feedback on their motivation to lead as an incentive for participation). We assessed *personality* using a short version of the Big Five Inventory with a total of 10 items¹³¹ as well as the 3-item fairness facet of the Honesty-Humility factor¹³². Reliabilities, descriptive statistics, and intercorrelations of all variables used in each study are available in Supplementary Datasets 1 and 2 as well as at <https://osf.io/yt4qh/>. The datasets analyzed during the current study are available in the Open Science Framework repository, <https://osf.io/yt4qh/>, except for information that could compromise research participant privacy.

Demographic information. We asked participants whether they currently hold a professional leadership position, or, if they are not working at the moment (e.g., because they are retired), whether they held a leadership position at some point in the past.

Measures used in the laboratory study (Settlers of Catan). *Verbal statements encouraging cooperation.* We videotaped the whole conversation during the game. We count all statements that favor either cooperation or selfishness. This count reflects (i) statements about cooperative/selfish strategies (e.g., “we should share the resources that everyone needs” vs. “I think it is best if everyone does their own thing”) and (ii) more general statements expressing a positive/negative attitude towards the group (e.g., “great, now everyone has more than 5 victory points” vs. “I don’t care what happens when I cause an oil spill”). We count all statements that (i) initiate a conversation about a topic related to cooperation, (ii) support such an initiative, or (iii) reject such an initiative (*reverse coded*, i.e., counting toward the other category). As support or rejection, we count only instances where a person makes an active statement. We do not count one word answers, nodding, or shaking one’s head.

For both statements encouraging cooperation and statements encouraging selfishness, we log-transform count values to reduce the weight of statements that are repetitions of a player’s position relative to statements that reveal a player’s position for the first time. Agreement over two trained raters is $r = 0.79$, $P < 0.0001$ for statements encouraging cooperation and $r = 0.78$, $P < 0.0001$ for statements encouraging selfishness. Next, we z -standardize statements encouraging cooperation ($M_{\text{raw}} = 4.8$, $s.d. = 5.9$) and statements encouraging selfishness ($M_{\text{raw}} = 1.6$, $s.d. = 2.6$) separately. Given that statements encouraging selfishness are more rare, we assume that they have a higher weight per statement in the conversation. By standardizing both types of statements separately before aggregating them, we assign an equal weight to both indices. When aggregating both indices, we assign a negative sign to statements encouraging selfishness. Inter-rater agreement is $r = 0.73$, $P < 0.0001$. Finally, we aggregate the resulting aggregates from both raters. Without log transformation in the beginning, the final aggregates would have had higher kurtosis (9.34 vs. 0.95, $s.e.m. = 0.34$). All count values are available on <https://osf.io/yt4qh/>.

Oil spills. During the game, the experimenter noted all moves on a custom-made form (available at <https://osf.io/yt4qh/>). Any inconsistencies in the record (the occurrence of an oil spill was logged at two different places) were resolved by replaying the game on video.

Ratings. Participants rated each other immediately after the game of Settlers of Catan “with respect to the behaviour of [each group member, referred to by their first names] during the whole experiment”. To ensure that all ratings are only based on behaviour during the game (instead of being based on non-behavioural information¹³³), we control for baseline ratings that we measured in the beginning of the experiment after a short group discussion. There, we asked respondents to provide all answers “with respect to the behaviour of [each group member, referred to by their first names] during the group discussion”.

Transformational leadership describes the extent to which followers feel inspired and supported by a leader to work towards common goals. We assessed transformational leadership with a German language adaptation¹³⁴ of the Multifactor Leadership Questionnaire¹³⁵. This adaptation measures 6 facets of transformational leadership. For each facet, we chose the item with the highest loading on one of the two factors representing transformational leadership¹³⁴, resulting in a total of 6 items answered on 5-point scales.

We measured *assumed leadership role* with 2 items answered on 5-point scales reading (i) “to what extent has [first name] shown leadership behaviour?” and (ii) “[during the group discussion/the whole experiment], [first name] has assumed a leadership role”.

Awareness of gender-based discrimination. We measured this variable with 3 items from the *denial of continuing discrimination* subscale of the *modern sexism scale*¹³⁶. Of its 5 items, we chose the ones that are either general or work-related, e.g., “women often miss out on good jobs due to sexual discrimination”. We coded the scale so that high values indicate high awareness of gender-based discrimination (and, thus, low modern sexism).

Control variables. We measured an *implicit affiliation motive* and an *implicit power motive* using the approach of the picture story exercise. We showed respondents a picture for 10 s and then asked them to come up with a story surrounding the depicted situation within 4 min per picture. We used 3 pictures—*women in laboratory*¹³⁷, *mad scientist*¹³⁸, and *nightclub scene*¹³⁸. Respondents’ stories were then coded for motive imagery by a trained coder using Winter’s coding system for running text. For example, if a character in one of the stories attempts to influence another character, that particular sentence of that particular story is coded as power imagery. Activity inhibition is coded by counting how often the word *not* is used¹³⁹. We correct for word count using regression analysis.

We measured an *achievement motive* with 4 items from a German questionnaire (the business focused inventory of personality¹⁴⁰). We chose items 22, 85, 159, and 172, because these items had the highest factor loadings

of all items that are phrased general enough for our purpose, e.g., “even after a very good performance, I still seek improvement”.

We measured *reasoning ability* with the short version of the Hagen Matrices Test¹⁴¹. This version consists of 6 3×3 matrices with 8 response options each. Each matrix needs to be completed within 2 min.

Measures used in the field survey. *Selfish business decisions.* Respondents read six detailed descriptions of hypothetical business scenarios¹⁴². All scenarios involve social dilemmas. Each decision requires balancing personal benefits against expected harm to society, the environment, or legal liability. Respondents indicated on 6-point scales how likely it was that they would make a selfish decision.

General leadership competence. Peers rated respondents on 3 items measuring their leadership competence in general, e.g., “the person that I am rating is/would make a good leader”. We asked respondents to nominate peers who know them very well. Peers indicated that they know respondents well, $M = 5.5$ (s.d. = 0.9) on a scale of 1 to 6. Family members (29%) gave the highest ratings ($r = 0.15$, $P < 0.0001$) whereas friends (38%, $r = -0.09$, $P = 0.043$) and acquaintances (6%, $r = -0.13$, $P = 0.0018$) gave the lowest ratings. We collapsed ratings across these different types of peers because we wanted to include all respondents regardless of external circumstances that might have influenced peer nomination (e.g., occupational status, availability of friends and family members, or partnership status).

Data Availability

Data are available for this paper at <https://osf.io/yt4qh/>.

References

- Hardin, G. The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
- The cooperative human. *Nat. Hum. Behav.* **2**, 427–428 <https://www.nature.com/articles/s41562-018-0389-1> (2018).
- Pais, L. & dos Santos, N. R. In *The Wiley Blackwell Handbook of the Psychology of Training, Development, and Performance Improvement*. (eds Kraiger, K., Passmore, J., dos Santos, N. R. & Malvezzi, S.) 278–302 (Wiley-Blackwell, Chichester, 2014).
- Kerr, N. L. Motivation losses in small groups: a social dilemma analysis. *J. Pers. Soc. Psychol.* **45**, 819–828 (1983).
- Hardy, C. L. Nice guys finish first: the competitive altruism hypothesis. *Pers. Soc. Psychol. B.* **32**, 1402–1413 (2006).
- Son Hing, L. S., Bobocel, D. R., Zanna, M. P. & McBride, M. V. Authoritarian dynamics and unethical decision making: high social dominance orientation leaders and high right-wing authoritarianism followers. *J. Pers. Soc. Psychol.* **92**, 67–81 (2007).
- de Cremer, D. & van Dijk, E. When and why leaders put themselves first: leader behaviour in resource allocations as a function of feeling entitled. *Eur. J. Soc. Psychol.* **35**, 553–563 (2005).
- Harrell, A. & Simpson, B. The dynamics of prosocial leadership: power and influence in collective action groups. *Soc. Forces* **94**, 1283–1308 (2016).
- Nauta, A., de Dreu, C. K. W. & van der Vaart, T. Social value orientation, organizational goal concerns and interdepartmental problem-solving behavior. *J. Organ. Behav.* **23**, 199–213 (2002).
- Milinski, M., Semmann, D. & Krambeck, H.-J. Reputation helps solve the ‘tragedy of the commons’. *Nature* **415**, 424–426 (2002).
- Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577 (1998).
- Treviño, L. K., den Nieuwenboer, N. A. & Kish-Gephart, J. J. (Un)ethical behavior in organizations. *Annu. Rev. Psychol.* **65**, 635–660 (2014).
- Dawes, R. M. Social dilemmas. *Annu. Rev. Psychol.* **31**, 169–193 (1980).
- van Lange, P. A. M., Joireman, J., Parks, C. D. & van Dijk, E. The psychology of social dilemmas: a review. *Organ. Behav. Hum. Dec.* **120**, 125–141 (2013).
- Pruitt, D. G. & Kimmel, M. J. Twenty years of experimental gaming: critique, synthesis, and suggestions for the future. *Annu. Rev. Psychol.* **28**, 363–392 (1977).
- De Waal, F. B. M. Putting the altruism back into altruism: the evolution of empathy. *Annu. Rev. Psychol.* **59**, 279–300 (2008).
- Zaki, J. Empathy: A motivated account. *Psychol. Bull.* **140**, 1608–1647 (2014).
- Klimecki, O. M., Mayer, S. V., Jusyte, A., Scheeff, J. & Schönberg, M. Empathy promotes altruistic behavior in economic interactions. *Sci. Rep.-UK* **6**, 31961 (2016).
- Haruno, M. & Frith, C. D. Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nat. Neurosci.* **13**, 160–161 (2010).
- Batson, C. D. & Ahmad, N. Empathy-induced altruism in a prisoner’s dilemma II: what if the target of empathy has defected? *Eur. J. Soc. Psychol.* **31**, 25–36 (2001).
- Chierchia, G., Pariansen-Lesemann, F. H., Snower, D., Vogel, M. & Singer, T. Caring cooperators and powerful punishers: differential effects of induced care and power motivation on different types of economic decision making. *Sci. Rep.-UK* **7**, 11068 (2017).
- Edele, A., Dziobek, I. & Keller, M. Explaining altruistic sharing in the dictator game: the role of affective empathy, cognitive empathy, and justice sensitivity. *Learn. Individ. Differ.* **24**, 96–102 (2013).
- McClelland, D. C. & Burnham, D. H. Power is the great motivator. *Harvard Bus. Rev.* **25**, 159–166 (1976).
- Chierchia, G. *et al.* Navigating motivation: a semantic and subjective atlas of 7 motives. Preprint at, <https://doi.org/10.31234/osf.io/wjhmq> (2018).
- Winter, D. G. *Power Motivation in Thought and Action* (unpublished doctoral dissertation, Harvard University, Cambridge, 1967).
- McClelland, D. C. The two faces of power. *J. Int. Aff.* **24**, 29–47 (1970).
- Weinberger, J., Cotler, T. & Fishman, D. In *Implicit motives*. (eds Schultheiss, O. & Brunstein, J.) 71–89 (Oxford University Press, New York, 2010).
- Spangler, W. D., Tikhomirov, A., Sotak, K. L. & Palrecha, R. Leader motive profiles in eight types of organizations. *Leadership Quart.* **25**, 1080–1094 (2014).
- Keltner, D., Gruenfeld, D. H. & Anderson, C. Power, approach, and inhibition. *Psychol. Rev.* **110**, 265–284 (2003).
- Whitson, J. A. *et al.* The blind leading: power reduces awareness of constraints. *J. Exp. Soc. Psychol.* **49**, 579–582 (2013).
- Anderson, C. & Galinsky, A. D. Power, optimism, and risk-taking. *Eur. J. Soc. Psychol.* **36**, 511–536 (2006).
- Fast, N. J., Sivanathan, N., Mayer, N. D. & Galinsky, A. D. Power and overconfident decision-making. *Organ. Behav. Hum. Dec.* **177**, 249–260 (2012).
- Galinsky, A. D., Magee, J. C., Inesi, M. E. & Gruenfeld, D. H. Power and perspectives not taken. *Psychol. Sci.* **17**, 1068–1074 (2006).
- See, K. E., Morrison, E. W., Rothman, N. B. & Soll, J. B. The detrimental effects of power on confidence, advice taking, and accuracy. *Organ. Behav. Hum. Dec.* **116**, 272–285 (2011).

35. Hildreth, J. A. D. & Anderson, C. Failure at the top: how power undermines collaborative performance. *J. Pers. Soc. Psychol.* **110**, 261–286 (2016).
36. Magee, J. C. & Langner, C. A. How personalized and socialized power motivation facilitate antisocial and prosocial decision-making. *J. Res. Pers.* **42**, 1547–1559 (2008).
37. Wang, F. & Sun, X. Absolute power leads to absolute corruption? Impact of power on corruption depending on the concepts of power one holds. *Eur. J. Soc. Psychol.* **46**, 77–89 (2016).
38. Rus, D., van Knippenberg, D. & Wisse, B. Leader power and self-serving behavior: the moderating role of accountability. *Leadership Quart.* **23**, 13–26 (2012).
39. Kopelman, S., Weber, J. M. & Messick, D. M. In *The Drama of the Commons*. (eds Ostrom, E., Dietz, T., Dolšák, N., Stern, P. C., Stonich, S. & Weber, E. U.) 113–156 (National Academy Press, Washington, 2002).
40. Bogaert, S., Boone, C. & Declerck, C. Social value orientation and cooperation in social dilemmas: a review and conceptual model. *Brit. J. Soc. Psychol.* **47**, 453–480 (2008).
41. Balliet, D., Parks, C. & Joireman, J. Social value orientation and cooperation in social dilemmas: a meta-analysis. *Group Process. Interg.* **12**, 533–547 (2009).
42. Van Vugt, M., Van Lange, P. A. M. & Meertens, R. M. Commuting by car or public transportation? A social dilemma analysis of travel mode judgements. *Eur. J. Soc. Psychol.* **26**, 373–395 (1996).
43. Van Lange, P. A. M., Bekkers, R., Schuyt, T. N. M. & Van Vugt, M. From games to giving: social value orientation predicts donations to noble causes. *Basic Appl. Soc. Psych.* **29**, 375–384 (2007).
44. Murphy, R. O., Ackermann, K. A. & Handgraaf, M. J. J. Measuring social value orientation. *Judgm. Decis. Mak.* **6**, 771–781 (2011).
45. Murphy, R. O. & Ackermann, K. A. Social value orientation: theoretical and measurement issues in the study of social preferences. *Pers. Soc. Psychol. Rev.* **18**, 13–41 (2014).
46. Sagiv, L., Sverdlik, N. & Schwarz, N. To compete or to cooperate? Values' impact on perception and action in social dilemma games. *Eur. J. Soc. Psychol.* **41**, 64–77 (2011).
47. Zhao, K. & Smillie, L. D. The role of interpersonal traits in social decision making: exploring sources of behavioral heterogeneity in economic games. *Pers. Soc. Psychol. Rev.* **19**, 277–302 (2015).
48. McCabe, K. O. & Fleeson, W. Are traits useful? Explaining trait manifestations as tools in the pursuit of goals. *J. Pers. Soc. Psychol.* **110**, 287–301 (2016).
49. De Cremer, D. & Leonardelli, G. J. Cooperation in social dilemmas and the need to belong: the moderating effect of group size. *Group Dyn.-Theor. Res.* **7**, 168–174 (2003).
50. Volk, S., Thöni, C. & Ruigrok, W. Personality, personal values and cooperation preferences in public goods games: a longitudinal study. *Pers. Individ. Differ.* **50**, 810–815 (2011).
51. Buchan, N. R., Jeong, S. S. & Ward, A. K. Local reasons to give globally: identity extension and global cooperation. *Sci. Rep.-UK* **7**, 15527 (2017).
52. McClelland, D. C. & Boyatzis, R. E. Leadership motive pattern and long-term success in management. *J. Appl. Psychol.* **67**, 737–743 (1982).
53. Sorrentino, R. M. & Field, N. Emergent leadership over time: the functional value of positive motivation. *J. Pers. Soc. Psychol.* **50**, 1091–1099 (1986).
54. Winter, D. G. Power, affiliation, and war: three tests of a motivational model. *J. Pers. Soc. Psychol.* **65**, 532–545 (1993).
55. Steinmann, B., Dörr, S. L., Schultheiss, O. C. & Maier, G. W. Implicit motives and leadership performance revisited: what constitutes the leadership motive pattern? *Motiv. Emotion* **39**, 167–174 (2015).
56. Steinmann, B., Ötting, S. K. & Maier, G. W. Need for affiliation as a motivational add-on for leadership behaviors and managerial success. *Front. Psychol.* **7**, 1972 (2016).
57. Cornelius, E. T. & Lane, F. B. The power motive and managerial success in a professionally oriented service industry organization. *J. Appl. Psychol.* **69**, 32–39 (1984).
58. Winter, D. G. Leader appeal, leader performance, and the motive profiles of leaders and followers: a study of American presidents and elections. *J. Pers. Soc. Psychol.* **52**, 196–202 (1987).
59. Jacobs, R. L. & McClelland, D. C. Moving up the corporate ladder: a longitudinal study of the leadership motive pattern and managerial success in women and men. *Consult. Psychol. J.: Pract. Res.* **46**, 32–41 (1994).
60. Kirkpatrick, S. A., Wofford, J. C. & Baum, J. R. Measuring motive imagery contained in the vision statement. *Leadership Quart.* **13**, 139–150 (2002).
61. Delbecq, A., House, R. J., de Luque, M. S. & Quigley, N. R. Implicit motives, leadership, and follower outcomes: an empirical test of CEOs. *J. Leadersh. Org. Stud.* **20**, 7–24 (2013).
62. Howard, A. In *New Developments in Goal Setting and Task Performance*. (eds Locke, E. A. & G. Latham, P.) 246–261 (Routledge, New York, 2013).
63. Spangler, W. D. & House, R. J. Presidential effectiveness and the leadership motive profile. *J. Pers. Soc. Psychol.* **60**, 439–455 (1991).
64. Winter, D. G. A motivational model of leadership: predicting long-term management success from TAT measures of power motivation and responsibility. *Leadership Quart.* **2**, 67–80 (1991).
65. Schultheiss, O. C. & Brunstein, J. C. Inhibited power motivation and persuasive communication: a lens model analysis. *J. Pers.* **70**, 553–582 (2002).
66. De Hoogh, A. H. B. *et al.* Leader motives, charismatic leadership, and subordinates' work attitude in the profit and voluntary sector. *Leadership Quart.* **16**, 17–38 (2005).
67. Zaccaro, S. J., Green, J. P., Dubrow, S. & Kolze, M. Leader individual differences, situational parameters, and leadership outcomes: a comprehensive review and integration. *Leadership Quart.* **29**, 2–43 (2018).
68. Antonakis, J. In *The Sage Handbook of Leadership*. (eds Bryman, A., Collinson, D., Grint, K., Jackson, B. & Uhl-Bien, M.) 269–285 (Sage, Thousand Oaks, 2011).
69. McClelland, D. C., Koestner, R. & Weinberger, J. How do self-attributed and implicit motives differ? *Psychol. Rev.* **96**, 690–702 (1989).
70. Köllner, M. G. & Schultheiss, O. C. Meta-analytic evidence of low convergence between implicit and explicit measures of the needs for achievement, affiliation, and power. *Front. Psychol.* **5**, 826 (2014).
71. Hagemeyer, B., Dufner, M. & Denissen, J. J. Double dissociation between implicit and explicit affiliative motives: a closer look at socializing behavior in dyadic interactions. *J. Res. Pers.* **65**, 89–93 (2016).
72. Brunstein, J. C. In *Motivation and Action*. (eds Heckhausen, J. & Heckhausen, H.) 369–405 (Springer, Cham, 2018).
73. Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L. & Banaji, M. R. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* **97**, 17–41 (2009).
74. Sagiv, L., Roccas, S., Cieciuch, J. & Schwartz, S. H. Personal values in human life. *Nat. Hum. Behav.* **1**, 630–639 (2017).
75. Cerasoli, C. P., Nicklin, J. M. & Ford, M. T. Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. *Psychol. Bull.* **140**, 980–1008 (2014).
76. Collins, C. J., Hanges, P. J. & Locke, E. A. The relationship of achievement motivation to entrepreneurial behavior: a meta-analysis. *Hum. Perform.* **17**, 95–117 (2004).
77. Baker, C. R. Defining and measuring affiliation motivation. *Eur. J. Soc. Psychol.* **9**, 97–99 (1979).

78. Schönbrodt, F. D., Gerstenberg, F. X. & An, I. R. T. analysis of motive questionnaires: the Unified Motive Scales. *J. Res. Pers.* **46**, 725–742 (2012).
79. Suessenbach, F., Loughnan, S., Schönbrodt, F. & Moore, A. B. The dominance, prestige, and leadership account of social power motives. *Eur. J. Pers.* **33**, 7–33 (2019).
80. You, D., Maeda, Y. & Bebeau, M. J. Gender differences in moral sensitivity: a meta-analysis. *Ethics Behav.* **21**, 263–282 (2011).
81. Friesdorf, R., Conway, P. & Gawronski, B. Gender differences in responses to moral dilemmas: a process dissociation analysis. *Pers. Soc. Psychol. B.* **41**, 696–713 (2015).
82. Balliet, D., Li, N. P., Macfarlan, S. J. & Van Vugt, M. Sex differences in cooperation: a meta-analytic review of social dilemmas. *Psychol. Bull.* **137**, 881–909 (2011).
83. Eagly, A. H. & Wood, W. The nature-nurture debates: 25 years of challenges in understanding the psychology of gender. *Perspect. Psychol. Sci.* **8**, 340–357 (2013).
84. Hyde, J. S. Gender similarities and differences. *Annu. Rev. Psychol.* **65**, 373–398 (2014).
85. Zell, E., Krizan, Z. & Teeter, S. R. Evaluating gender similarities and differences using metasynthesis. *Am. Psychol.* **70**, 10–20 (2015).
86. Diekmann, A. B. & Eagly, A. H. In *Handbook of Motivation Science*. (eds Shah, J. Y. & Gardner, W. L.) 434–447 (Guilford Press, New York, 2008).
87. Schwartz, S. H. & Rubel, T. Sex differences in value priorities: cross-cultural and multimethod studies. *J. Pers. Soc. Psychol.* **89**, 1010–1028 (2005).
88. Su, R., Rounds, J. & Armstrong, P. I. Men and things, women and people: a meta-analysis of sex differences in interests. *Psychol. Bull.* **135**, 859–884 (2009).
89. Gouveia, V. V., Vione, K. C., Milfont, T. L. & Fischer, R. Patterns of value change during the life span: some evidence from a functional approach to values. *Pers. Soc. Psychol. B.* **41**, 1276–1290 (2015).
90. Booth, A. & Nolen, P. Choosing to compete: how different are girls and boys? *J. Econ. Behav. Organ.* **81**, 542–555 (2012).
91. Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E. & Nauts, S. Status incongruity and backlash effects: defending the gender hierarchy motivates prejudice against female leaders. *J. Exp. Soc. Psychol.* **48**, 165–179 (2012).
92. Williams, M. J. & Tiedens, L. Z. The subtle suspension of backlash: a meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychol. Bull.* **142**, 165–197 (2016).
93. Locke, K. D. & Heller, S. Communal and agentic interpersonal and intergroup motives predict preferences for status versus power. *Pers. Soc. Psychol. B.* **43**, 71–86 (2017).
94. Van Vugt, M., De Cremer, D. & Janssen, D. P. Gender differences in cooperation and competition: the male-warrior hypothesis. *Psychol. Sci.* **18**, 19–23 (2007).
95. Kivikangas, J. M., Kätsyri, J., Järvelä, S. & Ravaja, N. Gender differences in emotional responses to cooperative and competitive game play. *Plos One* **9**, e100318 (2014).
96. Griswold, A. High-stakes game of oil use. *Proc. Natl Acad. Sci. USA* **110**, 2685 (2013).
97. Koenig, A. M., Eagly, A. H., Mitchell, A. A. & Ristikari, T. Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychol. Bull.* **137**, 616–642 (2011).
98. Science for all. *Nature* **495**, 5 <https://doi.org/10.1038/495021a> (2013).
99. Ensari, N., Riggio, R. E., Christian, J. & Carslaw, G. Who emerges as a leader? Meta-analyses of individual differences as predictors of leadership emergence. *Pers. Individ. Differ.* **51**, 532–536 (2011).
100. Epitropaki, O. & Martin, R. Implicit leadership theories in applied settings: factor structure, generalizability, and stability over time. *J. Appl. Psychol.* **89**, 293–310 (2004).
101. Busse, C., Kach, A. P. & Wagner, S. M. Boundary conditions: what they are, how to explore them, why we need them, and when to consider them. *Organ. Res. Methods* **20**, 574–609 (2017).
102. Gartzia, L. & van Knippenberg, D. Too masculine, too bad: effects of communion on leaders' promotion of cooperation. *Gr. Organ. Manag.* **41**, 458–490 (2016).
103. Peterson, S. J., Galvin, B. M. & Lange, D. CEO servant leadership: exploring executive characteristics and firm performance. *Pers. Psychol.* **65**, 565–596 (2012).
104. Ou, A. Y., Waldman, D. A. & Peterson, S. J. Do humble CEOs matter? An examination of CEO humility and firm outcomes. *J. Manage.* **44**, 1147–1173 (2018).
105. Harms, P. D., Wood, D., Landay, K., Lester, P. B. & Vogelgesang Lester, G. Autocratic leaders and authoritarian followers revisited: a review and agenda for the future. *Leadership Quart.* **29**, 105–122 (2018).
106. Keller Hansbrough, T. In *Emerging Challenges in Business, Optimization, Technology, and Industry* (eds Tadj, L. & Garg, A. K.) 15–20 (Springer, Cambridge, 2018).
107. Thorndike, E. L. Animal intelligence: an experimental study of the associative processes in animals. *Psychol. Rev.: Monogr. S.* **2**, i (1898).
108. Anderson, C., Brion, S., Moore, D. A. & Kennedy, J. A. A status-enhancement account of overconfidence. *J. Pers. Soc. Psychol.* **103**, 718–735 (2012).
109. Van Kleef, G. A., Homan, A. C., Finkenauer, C., Gündemir, S. & Stamkou, E. Breaking the rules to rise to power: how norm violators gain power in the eyes of others. *Soc. Psychol. Pers. Sci.* **2**, 500–507 (2011).
110. Kaiser, R. B., Hogan, R. & Craig, S. B. Leadership and the fate of organizations. *Am. Psychol.* **63**, 96–110 (2008).
111. Acton, B. P., Foti, R. J., Lord, R. G. & Gladfelder, J. A. Putting emergence back in leadership emergence: a dynamic, multilevel, process-oriented framework. *Leadership Quart.* **30**, 145–164 (2019).
112. Morgeson, F. P. et al. Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Pers. Psychol.* **60**, 1029–1049 (2007).
113. Grijalva, E., Harms, P. D., Newman, D. A., Gaddis, B. H. & Fraley, R. C. Narcissism and leadership: a meta-analytic review of linear and nonlinear relationships. *Pers. Psychol.* **68**, 1–47 (2015).
114. Leckelt, M., Küfner, A. C. P., Nestler, S. & Back, M. D. Behavioral processes underlying the decline of narcissists' popularity over time. *J. Pers. Soc. Psychol.* **109**, 856–871 (2015).
115. Kalish, Y. & Luria, G. Leadership emergence over time in short-lived groups: integrating expectations states theory with temporal person-perception and self-serving bias. *J. Appl. Psychol.* **101**, 1474–1486 (2016).
116. Lanaj, K. & Hollenbeck, J. R. Leadership over-emergence in self-managing teams: the role of gender and countervailing biases. *Acad. Manage. J.* **58**, 1476–1494 (2015).
117. Watts, L. L., Ness, A. M., Steele, L. M. & Mumford, M. D. Learning from stories of leadership: how reading about personalized and socialized politicians impacts performance on an ethical decision-making simulation. *Leadership Quart.* **29**, 276–294 (2018).
118. Kennedy, J. A. & Kray, L. J. Who is willing to sacrifice ethical values for money and social status? Gender differences in reactions to ethical compromises. *Soc. Psychol. Pers. Sci.* **5**, 52–59 (2014).
119. Drescher, A. & Schultheiss, O. C. Meta-analytic evidence for higher implicit affiliation and intimacy motivation scores in women, compared to men. *J. Res. Pers.* **64**, 1–10 (2016).
120. Gino, F., Wilmut, C. A. & Brooks, A. W. Compared to men, women view professional advancement as equally attainable, but less desirable. *Proc. Natl Acad. Sci. USA* **112**, 12354–12359 (2015).
121. Peters, K., Ryan, M. K. & Haslam, S. A. Marines, medics, and machismo: lack of fit with masculine occupational stereotypes discourages men's participation. *Brit. J. Psychol.* **106**, 635–655 (2015).

122. Horvath, L. K. & Sczesny, S. Reducing women's lack of fit with leadership positions? Effects of the wording of job advertisements. *Eur. J. Work Organ. Psy.* **25**, 316–328 (2016).
123. Schuh, S. C. *et al.* Gender differences in leadership role occupancy: the mediating role of power motivation. *J. Bus. Ethics* **120**, 363–379 (2014).
124. Mazei, J. *et al.* A meta-analysis on gender differences in negotiation outcomes and their moderators. *Psychol. Bull.* **141**, 85–104 (2015).
125. Tang, Y., Li, J. & Yang, H. What I see, what I do: how executive hubris affects firm innovation. *J. Manage.* **41**, 1698–1723 (2015).
126. Baumeister, R. F., Vohs, K. D. & Funder, D. C. Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspect. Psychol. Sci.* **2**, 396–403 (2007).
127. Back, M. D. & Vazire, S. The social consequences of personality: six suggestions for future research. *Eur. J. Pers.* **29**, 296–307 (2015).
128. Furr, R. M. Personality psychology as a truly behavioural science. *Eur. J. Pers.* **23**, 369–401 (2009).
129. Craig, J. M. & Sherif, C. W. The effectiveness of men and women in problem-solving groups as a function of group gender composition. *Sex Roles* **14**, 453–466 (1986).
130. Felfe, J., Elprana, G., Gatzka, M. & Stiehl, S. *FÜMO: Hamburger Führungsmotivationsinventar*. (Hogrefe, Göttingen, 2012).
131. Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. A short scale for assessing the big five dimensions of personality: 10 item big five inventory (BFI-10). *Methods, Data, Analyses* **7**, 233–249 (2013).
132. Ashton, M. C. & Lee, K. The HEXACO-60: A short measure of the major dimensions of personality. *J. Pers. Assess.* **91**, 340–345 (2009).
133. Olivola, C. Y., Funk, F. & Todorov, A. Social attributions from faces bias human choices. *Trends Cogn. Sci.* **18**, 566–570 (2014).
134. Felfe, J. Validierung einer deutschen version des “Multifactor Leadership Questionnaire” (MLQ Form 5 x Short) von Bass und Avolio (1995). *Zeitschrift für Arbeits- und Organisationspsychologie* **50**, 61–78 (2006).
135. Bass, B. M. & Avolio, B. *MLQ Multifactor Leadership Questionnaire: technical report* (Mind Garden, Redwood City, 1995).
136. Swim, J. K., Aikin, K. J., Hall, W. S. & Hunter, B. A. Sexism and racism: old-fashioned and modern prejudices. *J. Pers. Soc. Psychol.* **68**, 199–214 (1995).
137. Smith, C. P. *Motivation and Personality: Handbook of Thematic Content Analysis*. (Cambridge Univ. Press, Cambridge, 1992).
138. Winter, D. G. *The Power Motive*. (The Free Press, New York, 1973).
139. McClelland, D. C. *Power: The Inner Experience*. (Irvington, New York, 1975).
140. Hossiep, R., Paschen, M. & Mühlhaus, O. *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung*. (Hogrefe, Göttingen, 2003).
141. Heydasch, T., Haubrich, J. & Renner, K.-H. The short version of the Hagen Matrices Test (HMT-S). *Methods, Data, Analyses* **7**, 183–208 (2013).
142. Ashton, M. C. & Lee, K. The prediction of Honesty-Humility-related criteria by the HEXACO and Five-Factor Models of personality. *J. Res. Pers.* **42**, 1216–1228 (2008).

Acknowledgements

We thank M. Georg for contributing to item development. M. Gros, K. Gut, H. Hering, and D. Makoschey coded study variables.

Author Contributions

C.W. designed the studies and planned the paper with N.K., C.W. collected and analyzed the data. C.W. wrote the manuscript with N.K.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-45931-4>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Supplementary Information

Motives relate to cooperation in social dilemmas

but have an inconsistent association with leadership evaluation

Christian Wolff*^{1,2} & Nina Keith²

¹University of Bamberg

²Technische Universität Darmstadt

*christian.wolff@uni-bamberg.de

| Study | Year | Overlapping samples | n | Sector/context | Sample | Motive measure | Dependent variable(s) | Zero-order correlation | | Motive combination/variants | | |
|---|------|---------------------|-----|---|--|---|---|------------------------|--------------------|---|---|--|
| | | | | | | | | Power motive | Affiliation motive | Analytical approach | Label and formula | Central finding(s) |
| McClelland & Boyatzis, Sample 1 ⁴² | 1982 | No | 92 | Telecommunication | Male managers with engineering responsibilities | Picture Story Exercise (6 pictures) | <i>Objective data</i> Leader career success (after 8 and 16 y) | | | Configuration | Leadership motive pattern: $nPow \geq 45$, $nPow \geq nAff$, $AI > \text{median}$, $AI_{RAW} \geq 2$ | Leadership motive pattern not related to career success |
| McClelland & Boyatzis, Sample 2 ⁴² | 1982 | No | 144 | | Male managers without engineering responsibilities | Picture Story Exercise (6 pictures) | <i>Objective data</i> Leader career success (after 8 and 16 y) | | | Configuration | Leadership motive pattern: $nPow \geq 45$, $nPow \geq nAff$, $AI > \text{median}$, $AI_{RAW} \geq 2$ | Leadership motive pattern positive for career success |
| Cornelius & Lane, Sample 1 ⁴³ | 1984 | No | 18 | Education | Curriculum directors | Picture Story Exercise (6 pictures) | <i>Objective data</i> Administrative efficiency Center size <i>Ratings by subordinates</i> Employee satisfaction Team spirit Organizational clarity | | | Difference values | Leadership motive pattern: $nPow - nAff$ | -0.42* 0.59** -0.36† -0.20 -0.18 |
| Cornelius & Lane, Sample 2 ⁴³ | 1984 | No | 21 | | Center managers | Picture Story Exercise (6 pictures) | <i>Objective data</i> Administrative efficiency Center size <i>Ratings by subordinates</i> Employee satisfaction Team spirit Organizational clarity | | | Difference values | Leadership motive pattern: $nPow - nAff$ | -0.09 0.23 0.04 -0.26 -0.21 |
| Sorrentino & Field ⁴⁴ | 1986 | No | 48 | Laboratory (5 sessions, 1.5 to 2 h/ session, groups of 4) | Male psychology students | Descriptive sentences were used to elicit stories (4 sentences) | <i>Ratings by group members</i> Task leadership Socioemotional leadership Leadership emergence 1st choice Leadership emergence 1st & 2nd c. | | | Interactions via dichotomized variables | Two-way interaction between $nAff$ and $nAch$ | Interaction between $nAff$ and $nAch$ positive for 2 of 5 outcomes |

| | | | | | | | | | | | | |
|--------------------------------|------|--|-------|----------------------|--|-------------------------------------|---|---|--|---|---|--|
| Winter ⁴⁵ | 1987 | No | 14-31 | Government/ politics | US presidents | Speech (first inaugural address) | <i>Objective data/coded variables</i> Vote percentage Margin of victory Reelected Reelected (all instances) % vote for party's House candidates Court/cabinet rejections Percentage vetoes overridden Adjusted midterm House loss War entry War avoidance Arms limitation Consensus of greatness Great decisions cited | -0.04 -0.07 0.06 0.27 0.13 -0.19 0.01 -0.23 0.52** 0.34† -0.05 0.40* 0.51** | | Difference values | <i>nPow</i> – <i>nAff</i> | 0.10 0.05 -0.05 0.16 0.20 -0.20 -0.04 0.03 0.36† 0.16 -0.55* 0.35 0.27 |
| Spangler & House ⁴⁶ | 1991 | Yes (Winter, 1987 ⁴⁵) | 29-39 | Government/politics | US presidents | Speech (first inaugural address) | <i>Objective data/coded variables</i> War entry War avoidance Consensus of greatness Great decisions Mean greatness Social performance Economic performance International relations performance | 0.52** 0.33 0.40* 0.51** 0.26 0.17 0.00 0.27 | 0.17 0.23 0.09 0.30 -0.18 -0.20 0.07 0.00 | Configuration and interactions (both multiplicative and dichotomized) in the same multiple regression model | Syndrome: <i>nPow</i> ≥ 45, <i>nPow</i> ≥ <i>nAff</i> , AI ≥ median; two-way interaction between <i>nPow</i> and AI | Syndrome and interaction positive for 0 of 5 outcomes |
| House et al. ⁴⁷ | 1991 | Yes (Winter, 1987 ⁴⁵) | 31 | Government/politics | US presidents | Speech (first inaugural address) | <i>Objective data/coded variables</i> Charisma Direct action Subjective performance International relations performance Economic performance Social performance | | | Multiple regression analysis | Simultaneous main effects of <i>nPow</i> , <i>nAff</i> , <i>nAch</i> , AI, charisma, crises, and age | Power positive for 5 of 6 outcomes |
| Winter ⁴⁸ | 1991 | Yes (McClelland & Boyatzis, 1982 ⁴²) | 141 | Telecommunication | Male managers without engineering responsibilities | Picture Story Exercise (6 pictures) | <i>Objective data</i> Leader career success (after 16 y) | | | Direct coding of responsibility; configural approach | Responsible power: <i>nPow</i> ≥ 50, responsibility ≥ 45; responsible power motivation pattern: <i>nPow</i> > 45, <i>nPow</i> ≥ <i>nAff</i> , responsibility ≥ 45 | Responsible power and responsible power motivation pattern positive for career success |

| | | | | | | | | | | | | |
|---|------|----|-----|---------------------------|--|---|---|-----------|-----------|--|---|--|
| Winter ⁴⁹ | 1993 | No | 58 | Government/politics | British government | Sovereign's speech/ speech from the throne | <i>Objective data/coded variables</i> War entry (1 year prior) War entry (2 years prior) War entry (3 years prior) | | | Difference values | $nPow - nAff$ | 0.30* 0.21 -0.05 |
| Jacobs & McClelland, Sample 1 ⁵⁰ | 1994 | No | 229 | Telecommunication | Entry-level managers who stayed in the company | Picture Story Exercise (2 pictures) | <i>Objective data</i> Leader career success (after 12 y) | <i>ns</i> | <i>ns</i> | Direct coding of power themes; configural approach | Standard leadership motive pattern: $nPow \geq 45$, $nPow \geq nAff$, $AI \geq$ median; modified leadership motive pattern: $nPow \geq 45$, $nPow \geq nAff$; power themes: resourceful vs. reactive vs. helpless power | No significant relationships for motive patterns or variants |
| Jacobs & McClelland, Sample 2 ⁵⁰ | 1994 | No | 56 | Telecommunication | Entry-level managers who left the company | Picture Story Exercise (5 pictures) | <i>Objective data</i> Leader career success (after 12 y) | 0.34** | -0.30* | Direct coding of power themes; configural approach | Standard leadership motive pattern: $nPow \geq 45$, $nPow \geq nAff$, $AI \geq$ median; modified leadership motive pattern: $nPow \geq 45$, $nPow \geq nAff$; power themes: resourceful vs. reactive vs. helpless power | Modified leadership motive pattern positive for career success; successful male managers used reactive power themes while successful female managers used resourceful power themes |
| Langner & Winter, Sample 1 ⁵¹ | 2001 | No | 67 | Political crises | Documents by government officials | Official public statements and letters | <i>Objective data</i> Net concessions during political crises | -0.22† | 0.23† | | | |
| Kirkpatrick et al., Sample 1 ⁵² | 2002 | No | 269 | Architectural woodworking | Entrepreneurs in the role of the CEO | Vision statements | <i>Objective data</i> Venture growth (sales, employment, and profit over a period of 2 y, controlling for previous growth) | 0.11* | 0.07 | | | |

| | | | | | | | | | | | | |
|--|------|----|----|---|-----------------------------------|---|--|---------------------------------|--------------------|--|--|---|
| Kirkpatrick et al., Sample 2 ⁵² | 2002 | No | 82 | Federal engineering services | Supervisory managers | Vision statements | <i>Ratings by subordinates</i> Manager performance <i>Ratings by manager</i> Unit performance | -0.06 -0.16 [†] | 0.19* 0.21* | | | |
| Schultheiss & Brunstein ⁵³ | 2002 | No | 68 | Mixed | Students and employees | Picture Story Exercise (6 pictures) | <i>Ratings by observers</i> Persuasiveness | -0.01 | | Interactions in multiple regression analysis | Inhibited power motive: two-way interaction between <i>nPow</i> and AI | Inhibited power motive positive for persuasiveness |
| De Hoogh et al. ⁵⁴ | 2005 | No | 73 | Wide range of for-profit and non-profit organizations | CEOs | Semistructured interviews about the CEOs' role and their functioning as a manager (45 to 60 min) | <i>Ratings by subordinates</i> Charismatic leadership Organizational commitment | 0.28* -0.10 | -0.18 -0.10 | Interactions in hierarchical regression analysis | Two-way interaction between <i>nPow</i> and responsibility and three-way interaction between <i>nPow</i> , responsibility, and organization type | Irresponsible power positive for charismatic leadership in for-profit organizations |
| Magee & Langner, Sample 1 ⁶ | 2008 | No | 90 | Laboratory experiment (political conflict) | Students and university employees | Participants' drafts of a letter to Premier N. S. Krushchev on behalf of US President J. F. Kennedy | <i>Self-reported decisions</i> Advised escalation of conflict | | 0.00 | Direct coding of motive variants (Winter, 1973 ⁶⁰) | Personalized power: hope for power; socialized power: fear of power (Winter, 1973 ⁶⁰) | Personalized power positive for escalation of conflict |
| Magee & Langner, Sample 2 ⁶ | 2008 | No | 69 | Laboratory experiment (healthcare) | Students and university employees | Fifteen written personal strivings in participants' everyday lives | <i>Self-reported decisions</i> Recommendation for approving a beneficial (but risky) drug | | 0.04 | Direct coding of motive variants (Winter, 1973 ⁶⁰) | Personalized power: hope for power; socialized power: fear of power (Winter, 1973 ⁶⁰) | Socialized power positive for drug approval |

| | | | | | | | | | | | | |
|--------------------------------|------|--|-----|-----------------------------|--|---|---|---|--|---------------|--|---|
| Kazén & Kuhl ⁵⁵ | 2011 | No | 382 | Mixed | Executive managers | Operant Motive Test (15 pictures) | <i>Self-reported well-being</i> Well-being Stress | 0.01 -0.03 | | | | |
| Delbecq et al. ⁵⁶ | 2013 | No | 28 | Technology (Silicon Valley) | CEOs | Semistructured interviews about concerns, beliefs, values, opinions, and management philosophies (45 to 60 min) | <i>Ratings by subordinates</i> Participative leadership Instrumental leadership Charismatic leadership Follower motivation Exceptional performance Teamwork | 0.08 -0.25 -0.17 -0.24 -0.16 -0.04 | 0.13 0.22 0.28 0.27 0.12 0.12 | | | |
| Howard, Sample 1 ⁵⁷ | 2013 | Yes (McClelland & Boyatzis, 1982 ⁴²) | 101 | Telecommunication | Male managers with engineering responsibilities | Picture Story Exercise (4 to 6 pictures) | <i>Objective data</i> Leader career success (after 25 y) Leader career success (after 25 y), predictors measured at year 8 | 0.08 -0.04 | -0.09 0.06 | Configuration | Leadership motive pattern: $nPow \geq 45$, $nPow \geq nAff$, $AI \geq$ median; leader motive without AI: $nPow \geq 45$, $nPow \geq nAff$ | Both leadership motive patterns not related to career success at year 25 |
| Howard, Sample 2 ⁵⁷ | 2013 | Yes (McClelland & Boyatzis, 1982 ⁴²) | 174 | Telecommunication | Male managers without engineering responsibilities | Picture Story Exercise (4 to 6 pictures) | <i>Objective data</i> Leader career success (after 25 y) Leader career success (after 25 y), predictors measured at year 8 | 0.03 0.01 | -0.25** 0.03 | Configuration | Leadership motive pattern: $nPow \geq 45$, $nPow \geq nAff$, $AI \geq$ median; leader motive without AI: $nPow \geq 45$, $nPow \geq nAff$ | Both leadership motive patterns not related to career success at year 25 |
| Howard, Sample 3 ⁵⁷ | 2013 | Yes (Jacobs & McClelland, 1994 ⁵⁰) | 111 | Telecommunication | Managers who left the organization | Picture Story Exercise (4 to 6 pictures) | <i>Ratings by manager</i> Leader career success (after 25 y) Salary (after 25 y) | 0.00 -0.01 | 0.04 -0.03 | Configuration | Leadership motive pattern: $nPow \geq 45$, $nPow \geq nAff$, $AI \geq$ median; leader motive without AI: $nPow \geq 45$, $nPow \geq nAff$ | Both leadership motive patterns not related to career success and salary at year 25 |

| | | | | | | | | | | | | |
|--------------------------------|------|----|----|-------|----------|--|--|---|---|--|--|--|
| Steinmann et al. ⁵⁸ | 2015 | No | 70 | Mixed | Managers | Picture Story Exercise (6 pictures, 5 min/picture) | <i>Ratings by manager</i> Goal attainment of the team Developments in income | -0.11 0.06 | 0.09 -0.11 | Interactions in hierarchical regression analysis | Compassionate leadership profile: two-way interactions between <i>nPow</i> , <i>nAff</i> , and <i>AI</i> and three-way interaction between <i>nPow</i> , <i>nAff</i> , and <i>AI</i> | Three-way interaction between <i>nPow</i> , <i>nAff</i> , and <i>AI</i> positive for 2 of 2 outcomes |
| Steinmann et al. ⁵⁹ | 2016 | No | 70 | Mixed | Managers | Picture Story Exercise (6 pictures, 4 min/picture) | <i>Ratings by subordinates</i> Transformational leadership Passive leadership Concern for followers' needs Job satisfaction Satisfaction with the leader In-role performance Organizational citizenship behavior <i>Ratings by manager</i> Developments in income | 0.09 0.03 0.11 0.01 0.09 -0.03 -0.18 -0.11 | 0.14 -0.04 0.25* 0.14 0.06 0.12 0.15 -0.18 | Interactions in hierarchical regression analysis | Two-way interactions between <i>nPow</i> , <i>nAff</i> , <i>nAch</i> , and <i>AI</i> and three-way interaction between <i>nPow</i> , <i>nAff</i> , and <i>nAch</i> | Three-way interaction between <i>nPow</i> , <i>nAff</i> , and <i>nAch</i> positive for 4 of 6 outcomes |

Table S1. Previous studies on the relationship of implicit power and affiliation motives, their combinations, and/or their variants with leadership and leader outcomes ($k = 26$, $n = 2,495$ participants). Notes: We excluded studies that (i) did not provide enough information (McClelland & Burnham, 1976¹; Burnham, 1997⁶¹; Lukić, 2015⁶²; Winter, 2018⁶³), (ii) were not specific for leadership (Jenkins, 1994⁶⁴; Winter et al., 1998⁶⁵), or (iii) were unavailable to us (Winter, 1979⁶⁶). CEO = chief executive officer, *nPow* = need for power/power motive, *nAff* = need for affiliation/affiliation motive, *nAch* = need for achievement/achievement motive, and *AI* = activity inhibition. This overview only includes information about the achievement motive if it was central to a study that also included the power and/or affiliation motive. The lists of dependent variables are not exhaustive. ** $P < 0.01$, * $P < 0.05$, † $P < 0.10$.

| Predictor | Leaders (n = 257) | | | Workers (n = 446) | | | Students (n = 258) | | |
|----------------------------------|-------------------|----------|----------|-------------------|----------|----------|--------------------|----------|----------|
| | β | <i>t</i> | <i>P</i> | β | <i>t</i> | <i>P</i> | β | <i>t</i> | <i>P</i> |
| Functional affiliation motive | -0.18 | -3.10 | 0.002 | -0.19 | -4.04 | 0.000 | -0.25 | -3.86 | 0.000 |
| Dysfunctional affiliation motive | -0.08 | -1.34 | 0.182 | 0.10 | 2.07 | 0.039 | 0.03 | 0.44 | 0.661 |
| Functional power motive | -0.10 | -1.62 | 0.107 | -0.07 | -1.42 | 0.157 | 0.09 | 1.25 | 0.212 |
| Dysfunctional power motive | 0.60 | 11.03 | 0.000 | 0.38 | 8.27 | 0.000 | 0.35 | 5.36 | 0.000 |

Table S2. Motives relate to selfish business decisions across different occupational statuses. Notes: *Leaders* state that they currently hold a professional leadership position or, if they are not working anymore, held one in the past. *Workers* report having work experience (but no leadership position). *Students* are either students or homemakers.

| Predictor | Encouragement of cooperation (<i>n</i> = 201) | | | Oil spills (<i>n</i> = 201) | | | Selfish business decisions (<i>n</i> = 960) | | |
|---|--|----------|----------|------------------------------|----------|----------|--|----------|----------|
| | β | <i>t</i> | <i>P</i> | β | <i>t</i> | <i>P</i> | β | <i>t</i> | <i>P</i> |
| <i>Simple model</i> | | | | | | | | | |
| Group size (3 vs. 4) | -0.03 | -0.50 | 0.620 | 0.16 | 2.37 | 0.019 | | | |
| Functional affiliation motive | 0.25 | 3.37 | 0.001 | -0.25 | -3.38 | 0.001 | -0.20 | -6.30 | 0.000 |
| Dysfunctional affiliation motive | -0.00 | -0.06 | 0.955 | -0.01 | -0.09 | 0.927 | 0.03 | 1.03 | 0.301 |
| Functional power motive | 0.09 | 1.08 | 0.283 | 0.01 | 0.17 | 0.863 | -0.06 | -1.60 | 0.111 |
| Dysfunctional power motive | -0.14 | -1.94 | 0.054 | 0.23 | 3.26 | 0.001 | 0.44 | 14.32 | 0.000 |
| <i>Extended model</i> | | | | | | | | | |
| Group size (3 vs. 4) | -0.02 | -0.22 | 0.824 | 0.13 | 1.89 | 0.060 | | | |
| Neuroticism | 0.11 | 1.43 | 0.154 | -0.02 | -0.30 | 0.767 | -0.10 | -3.33 | 0.001 |
| Extraversion | 0.08 | 0.93 | 0.354 | 0.11 | 1.46 | 0.147 | 0.01 | 0.22 | 0.830 |
| Openness | 0.10 | 1.45 | 0.150 | -0.05 | -0.71 | 0.478 | -0.13 | -4.89 | 0.000 |
| Agreeableness | -0.06 | -0.75 | 0.455 | 0.00 | 0.06 | 0.955 | -0.06 | -2.06 | 0.039 |
| Conscientiousness | -0.04 | -0.43 | 0.665 | -0.19 | -2.32 | 0.022 | -0.01 | -0.31 | 0.756 |
| Fairness | 0.14 | 1.83 | 0.068 | -0.13 | -1.67 | 0.097 | -0.31 | -10.67 | 0.000 |
| Reasoning ability | -0.04 | -0.54 | 0.588 | 0.03 | 0.49 | 0.627 | | | |
| Achievement motive | 0.10 | 1.19 | 0.236 | -0.05 | -0.56 | 0.578 | | | |
| Motivation to lead | -0.03 | -0.31 | 0.760 | -0.03 | -0.34 | 0.735 | -0.02 | -0.63 | 0.526 |
| Implicit affiliation motive | -0.08 | -1.09 | 0.277 | -0.03 | -0.36 | 0.720 | | | |
| Implicit power motive | -0.04 | -0.50 | 0.620 | 0.11 | 1.57 | 0.118 | | | |
| Activity inhibition | 0.13 | 1.77 | 0.078 | 0.00 | 0.02 | 0.984 | | | |
| Affiliation × power | -0.08 | -1.07 | 0.289 | -0.04 | -0.52 | 0.603 | | | |
| Affiliation × activity inhibition | -0.00 | -0.04 | 0.969 | 0.02 | 0.30 | 0.765 | | | |
| Power × activity inhibition | 0.09 | 1.19 | 0.237 | -0.04 | -0.59 | 0.556 | | | |
| Affiliation × power × activity inhibition | -0.16 | -1.87 | 0.063 | -0.05 | -0.54 | 0.588 | | | |
| Functional affiliation motive | 0.23 | 2.74 | 0.007 | -0.18 | -2.22 | 0.028 | -0.12 | -3.73 | 0.000 |
| Dysfunctional affiliation motive | -0.03 | -0.32 | 0.751 | -0.04 | -0.40 | 0.688 | 0.05 | 1.42 | 0.156 |
| Functional power motive | 0.10 | 1.03 | 0.306 | 0.02 | 0.17 | 0.865 | -0.03 | -0.86 | 0.389 |
| Dysfunctional power motive | -0.11 | -1.18 | 0.241 | 0.13 | 1.40 | 0.164 | 0.33 | 10.00 | 0.000 |

Table S3. Motives relate to cooperation in social dilemmas. Notes: *Group size* is coded so that 3 persons = 3 (*n* = 7) and 4 persons = 4 (*n* = 45).

| Predictor | Leadership ratings (after social dilemma, $n = 201$) | | | Leadership ratings (in general, $n = 486$) | | | Leadership position ($n = 961$) | | |
|---|---|-------|-------|---|-------|-------|-----------------------------------|-------|-------|
| | β | t | P | β | t | P | β | t | P |
| <i>Simple model</i> | | | | | | | | | |
| Baseline values | 0.70 | 14.02 | 0.000 | | | | | | |
| Group size (3 vs. 4) | 0.10 | 1.94 | 0.054 | | | | | | |
| Functional affiliation motive | 0.17 | 3.08 | 0.002 | 0.03 | 0.73 | 0.467 | -0.06 | -1.79 | 0.074 |
| Dysfunctional affiliation motive | -0.09 | -1.57 | 0.118 | -0.09 | -1.80 | 0.072 | -0.08 | -2.10 | 0.036 |
| Functional power motive | 0.02 | 0.26 | 0.799 | 0.25 | 4.80 | 0.000 | 0.21 | 5.44 | 0.000 |
| Dysfunctional power motive | -0.06 | -1.08 | 0.283 | 0.03 | 0.68 | 0.498 | -0.03 | -0.84 | 0.404 |
| <i>Extended model</i> | | | | | | | | | |
| Baseline values | 0.68 | 12.74 | 0.000 | | | | | | |
| Group size (3 vs. 4) | 0.08 | 1.57 | 0.118 | | | | | | |
| Neuroticism | -0.02 | -0.31 | 0.755 | 0.04 | 0.90 | 0.369 | -0.05 | -1.33 | 0.183 |
| Extraversion | 0.02 | 0.37 | 0.715 | 0.08 | 1.64 | 0.102 | -0.02 | -0.65 | 0.513 |
| Openness | -0.01 | -0.21 | 0.834 | -0.07 | -1.62 | 0.106 | 0.04 | 1.24 | 0.216 |
| Agreeableness | 0.08 | 1.32 | 0.190 | 0.00 | 0.03 | 0.980 | 0.08 | 2.25 | 0.025 |
| Conscientiousness | -0.12 | -1.84 | 0.068 | 0.14 | 3.05 | 0.002 | 0.12 | 3.45 | 0.001 |
| Fairness | 0.04 | 0.69 | 0.494 | 0.06 | 1.32 | 0.187 | -0.02 | -0.57 | 0.567 |
| Reasoning ability | 0.08 | 1.46 | 0.147 | | | | | | |
| Achievement motive | 0.06 | 0.96 | 0.336 | | | | | | |
| Motivation to lead | 0.08 | 1.07 | 0.285 | 0.16 | 2.79 | 0.005 | 0.26 | 6.28 | 0.000 |
| Implicit affiliation motive | -0.01 | -0.20 | 0.841 | | | | | | |
| Implicit power motive | -0.00 | -0.08 | 0.939 | | | | | | |
| Activity inhibition | 0.04 | 0.76 | 0.447 | | | | | | |
| Affiliation \times power | 0.07 | 1.22 | 0.226 | | | | | | |
| Affiliation \times activity inhibition | -0.05 | -0.89 | 0.375 | | | | | | |
| Power \times activity inhibition | 0.03 | 0.50 | 0.620 | | | | | | |
| Affiliation \times power \times activity inhibition | -0.01 | -0.19 | 0.848 | | | | | | |
| Functional affiliation motive | 0.14 | 2.30 | 0.022 | 0.02 | 0.32 | 0.748 | -0.09 | -2.34 | 0.019 |
| Dysfunctional affiliation motive | -0.10 | -1.45 | 0.150 | -0.04 | -0.70 | 0.484 | -0.02 | -0.41 | 0.679 |
| Functional power motive | -0.04 | -0.63 | 0.533 | 0.16 | 2.83 | 0.005 | 0.09 | 2.09 | 0.037 |
| Dysfunctional power motive | -0.05 | -0.72 | 0.471 | 0.02 | 0.29 | 0.772 | -0.08 | -2.10 | 0.036 |

Table S4. Motives relate to leadership evaluation. Notes: *Group size* is coded so that 3 persons = 3 ($n = 7$) and 4 persons = 4 ($n = 45$). *Leadership ratings (after social dilemma)* are made on a scale measuring transformational leadership. We control for baseline values from before the game so that ratings are only based on behaviour during the game. For *leadership position*, we calculated standard multiple regression analysis to facilitate comparison of results. Binomial logistic regression analysis yields virtually identical results.

| Dependent variable | Sample | Men | | | Women | | | <i>d</i> | <i>t</i> | <i>P</i> |
|---|--------------------------------------|----------|-------|------|----------|-------|------|----------|----------|----------|
| | | <i>n</i> | Mean | s.d. | <i>n</i> | Mean | s.d. | | | |
| Motive | | | | | | | | | | |
| Functional affiliation motive | Laboratory study (Settlers of Catan) | 98 | 4.33 | 0.66 | 103 | 4.73 | 0.57 | 0.64 | 4.56 | 0.000 |
| | Field survey | 448 | 4.24 | 0.71 | 513 | 4.50 | 0.61 | 0.39 | 5.96 | 0.000 |
| | Subgroup: leaders | 132 | 4.28 | 0.76 | 125 | 4.51 | 0.65 | 0.32 | 2.68 | 0.008 |
| | Subgroup: workers | 189 | 4.19 | 0.66 | 257 | 4.49 | 0.60 | 0.48 | 5.05 | 0.000 |
| | Subgroup: students | 127 | 4.29 | 0.72 | 131 | 4.51 | 0.60 | 0.33 | 2.58 | 0.010 |
| Dysfunctional power motive | Laboratory study (Settlers of Catan) | 98 | 3.10 | 0.96 | 103 | 2.65 | 0.90 | -0.49 | -3.48 | 0.001 |
| | Field survey | 448 | 3.23 | 0.92 | 513 | 2.93 | 0.95 | -0.32 | -4.93 | 0.000 |
| | Subgroup: leaders | 132 | 3.16 | 0.95 | 125 | 2.95 | 1.02 | -0.21 | -1.67 | 0.097 |
| | Subgroup: workers | 189 | 3.19 | 0.91 | 257 | 2.94 | 0.95 | -0.27 | -2.80 | 0.005 |
| | Subgroup: students | 127 | 3.36 | 0.89 | 131 | 2.90 | 0.88 | -0.52 | -4.23 | 0.000 |
| Cooperation in social dilemmas | | | | | | | | | | |
| Encouragement of cooperation | Laboratory study (Settlers of Catan) | 98 | -0.20 | 1.12 | 103 | 0.19 | 0.84 | 0.40 | 2.80 | 0.006 |
| Statements pro cooperation | | 98 | 5.36 | 6.99 | 103 | 4.22 | 4.54 | -0.20 | -1.37 | 0.173 |
| Statements pro selfishness | | 98 | 2.26 | 2.91 | 103 | 1.05 | 2.06 | -0.48 | -3.38 | 0.001 |
| Oil spills | | 98 | 0.40 | 0.68 | 103 | 0.06 | 0.24 | -0.67 | -4.66 | 0.000 |
| Selfish business decisions | Field survey | 448 | 2.64 | 1.04 | 512 | 2.18 | 0.91 | -0.47 | -7.22 | 0.000 |
| Leadership evaluation | | | | | | | | | | |
| Leadership ratings (after social dilemma) | Laboratory study (Settlers of Catan) | 98 | -0.08 | 0.39 | 103 | -0.10 | 0.36 | -0.06 | -0.43 | 0.667 |
| Leadership ratings (in general) | Field survey | 191 | 4.51 | 1.01 | 295 | 4.69 | 0.93 | 0.19 | 2.00 | 0.046 |
| Leadership position | | 448 | 0.29 | 0.46 | 513 | 0.24 | 0.43 | -0.11 | -1.77 | 0.076 |

Table S5. Motives, cooperation, and leadership evaluations by gender. Notes: Positive values of *d* and *t* indicate higher scores for women compared to men. *Leaders* state that they currently hold a professional leadership position or, if they are not working anymore, held one in the past. *Workers* report having work experience (but no leadership position). *Students* are either students or homemakers. *Leadership ratings (after social dilemma)* are difference values (ratings after dilemma minus baseline ratings). We measured *encouragement of cooperation* by counting all statements favoring cooperation (positive values) or selfishness (negative values). Count values are then log-transformed, aggregated using equal weights, and then aggregated over two independent observers ($r = 0.71$). Given that statements encouraging selfishness are more rare, we assume that they have a higher weight per statement in the conversation. By standardizing both types of statements separately before aggregating them, we assign an equal weight to both indices. We do not control for group size in any of the values reported in this table.

| Dependent variable | Via | n | β | z | P | 95% CI | |
|---|-------------------------------|-----|---------|-------|-------|--------|--------|
| | | | | | | Lower | Upper |
| <i>Cooperation in social dilemmas</i> | | | | | | | |
| Encouragement of cooperation | Functional affiliation motive | 201 | 0.067 | 2.34 | 0.019 | 0.016 | 0.143 |
| | Dysfunctional power motive | 201 | 0.029 | 1.46 | 0.144 | -0.003 | 0.082 |
| Oil spills | Functional affiliation motive | 201 | -0.055 | -2.08 | 0.037 | -0.123 | -0.015 |
| | Dysfunctional power motive | 201 | -0.047 | -2.13 | 0.034 | -0.113 | -0.013 |
| Selfish business decisions | Functional affiliation motive | 960 | -0.033 | -3.95 | 0.000 | -0.053 | -0.017 |
| | Dysfunctional power motive | 960 | -0.066 | -4.58 | 0.000 | -0.096 | -0.039 |
| <i>Leadership evaluation</i> | | | | | | | |
| Leadership ratings (after social dilemma) | Functional affiliation motive | 201 | 0.060 | 2.71 | 0.007 | 0.023 | 0.121 |
| | Dysfunctional power motive | 201 | 0.017 | 1.21 | 0.228 | -0.006 | 0.055 |
| Leadership ratings (in general) | Functional affiliation motive | 486 | 0.002 | 0.32 | 0.750 | -0.011 | 0.020 |
| | Dysfunctional power motive | 486 | -0.009 | -0.99 | 0.324 | -0.033 | 0.008 |
| Leadership position | Functional affiliation motive | 961 | -0.023 | -1.33 | 0.185 | -0.062 | 0.010 |
| | Dysfunctional power motive | 961 | 0.011 | 0.84 | 0.399 | -0.014 | 0.041 |

Table S6. Indirect effects of gender via motives on cooperation and evaluations. Notes: Positive values of β and z indicate that women score higher on the dependent variable due to their average level on the mediating motive. We control for a dysfunctional affiliation motive and a functional power motive in all analyses. Additionally, we control for group size in the laboratory study. CI, bias-corrected confidence interval (10,000 bootstrap samples).

| Original German item | English version |
|---|--|
| Instruction | |
| Bitte kreuzen Sie an, inwieweit die Aussagen im Arbeitskontext auf Sie zutreffen. | Please check to what extent these statements apply to you in the context of work. |
| Functional affiliation motive | |
| Ich wünsche mir, für meine verständnisvolle und kooperative Art gemocht zu werden. | I wish that people like me for being sympathetic and cooperative. |
| Bei Entscheidungen, die ich gegen den Willen anderer treffen muss, achte ich sehr darauf, nicht in eine Außenseiterposition zu geraten. | When I have to make decisions against the will of others, I pay close attention not to put myself in the position of an outsider. |
| Ich genieße es, mit anderen Menschen konstruktiv ein gemeinsames Ziel zu verfolgen. | I enjoy to constructively pursue a common goal with other people. |
| Gerade bei unpopulären Entscheidungen finde ich es besonders wichtig, viel Verständnis für diejenigen aufzubringen, die von solchen Entscheidungen betroffen sind. | Especially when making unpopular decisions, I find it particularly important to be appreciative of those who are affected by these decisions. |
| Dysfunctional affiliation motive | |
| Ich vermeide es um jeden Preis, Konflikte auszutragen, die das harmonische Miteinander in der Gruppe gefährden. | I avoid at all costs to engage in conflicts that jeopardize harmonious togetherness within the group. |
| Ich mache mir oft Sorgen, von anderen weniger gemocht zu werden, weil ich etwas Falsches sage. In solchen Momenten schweige ich lieber, als dass ich es riskiere, mit meiner Meinung anzuecken. | I often worry that others like me less for saying something wrong. In these moments I rather fall silent than risk to offend with my opinion. |
| Es ist mir wichtiger, Konflikte konstruktiv anzugehen, anstatt sie unter den Teppich zu kehren, nur um die Harmonie aufrecht zu erhalten. (reverse coded) | It is more important to me to approach conflicts constructively rather than sweeping them under the rug only to maintain harmony. (reverse coded) |
| Es ist mir sehr wichtig, von anderen akzeptiert zu werden. Deshalb sage ich manchmal Dinge, von deren Richtigkeit ich zwar nicht überzeugt bin, aber durch die ich gut dastehe. | It is very important to me to be accepted by others. Therefore I sometimes say things of which I am not convinced that they are right, but that make me look good. |
| Functional power motive | |
| Es stellt mich zufrieden, andere Menschen so in ihren Handlungen und Einstellungen zu beeinflussen, dass sie ungeahnte Fähigkeiten entdecken und herausfordernde Aufgaben bewältigen können. | It satisfies me to influence others in their actions and attitudes so that they discover unexpected capabilities and accomplish challenging tasks. |
| Es bereitet mir Freude, Verantwortung für eine übergeordnete Sache zu übernehmen, auch wenn das bedeuten kann, Rückschläge zu erfahren und Fehler eingestehen zu müssen. | It pleases me to take responsibility for a greater cause, even if that might involve experiencing setbacks and admitting mistakes. |
| Ich mag es, kontroverse Standpunkte zu vertreten, aber nur, solange es auf angemessene Art und Weise geschieht. | I like advancing controversial views, but only if it happens in an appropriate way. |
| Ich genieße es, durch die eigenen Einflussmöglichkeiten etwas beitragen zu können, das im Sinne übergeordneter Ziele steht. | I enjoy to contribute something through my channels of influence that is aligned with the greater good. |
| Dysfunctional power motive | |
| Ich genieße es, wenn andere meinen Rat oder meine Anweisungen einholen müssen, bevor sie handeln. | I enjoy it if others have to obtain my advice or instructions before they act. |
| Es gefällt mir, viel Macht und Einfluss zu haben, da es viele Menschen gibt, die man unter Kontrolle halten sollte. | It pleases me to have a lot of power and influence, because there are many people that you need to keep under control. |
| Es ist mir so wichtig, meine persönlichen Ziele zu erreichen, dass ich dafür auch andere Menschen benutzen würde. | It is so important for me to reach my personal goals that I would use other people for it. |
| Es ist ein schönes Gefühl, meinen gesellschaftlichen Status zu demonstrieren. | It is a nice feeling to demonstrate my social status. |

Table S7. Wording of items measuring variants of affiliation and power motives. Notes: Items were translated to English and back translated to German. Discrepancies were resolved through discussion. Response scales range from 1 (*does not at apply at all*) to 6 (*fully applies*).

| Scale | | Heterotrait-monotrait ratio of inter-item correlations | | | |
|-------|----------------------------------|--|-------|------|---|
| | | 1 | 2 | 3 | 4 |
| 1 | Functional affiliation motive | — | | | |
| 2 | Dysfunctional affiliation motive | 0.10 | — | | |
| 3 | Functional power motive | 0.68 | -0.47 | — | |
| 4 | Dysfunctional power motive | -0.01 | 0.36 | 0.28 | — |

Table S8. Discriminant validity between affiliation and power motives. Notes: Heterotrait-monotrait ratios of inter-item correlations compare average inter-item correlations within a scale to the average correlations of the items of that scale with the items of another scale ($n = 960$ to 961).

| Item | Affiliation | | Power | | |
|---|--|---------------|-------------|---------------|-------------|
| | Functional | Dysfunctional | Functional | Dysfunctional | |
| | λ | λ | λ | λ | |
| Functional affiliation motive | | | | | |
| 1 | I wish that people like me for being sympathetic and cooperative. | 0.64 | 0.19 | 0.38 | 0.29 |
| 2 | When I have to make decisions against the will of others, I pay close attention not to put myself in the position of an outsider. | 0.36 | 0.23 | 0.14 | -0.04 |
| 3 | I enjoy to constructively pursue a common goal with other people. | 0.39 | -0.14 | 0.15 | -0.41 |
| 4 | Especially when making unpopular decisions, I find it particularly important to be appreciative of those who are affected by these decisions. | 0.39 | -0.02 | 0.26 | 0.00 |
| Dysfunctional affiliation motive | | | | | |
| 1 | I avoid at all costs to engage in conflicts that jeopardize harmonious togetherness within the group. | 0.21 | 0.54 | -0.20 | 0.03 |
| 2 | I often worry that others like me less for saying something wrong. In these moments I rather fall silent than risk to offend with my opinion. | 0.41 | 0.70 | -0.31 | 0.10 |
| 3 | It is more important to me to approach conflicts constructively rather than sweeping them under the rug only to maintain harmony. (reverse coded) | -0.17 | 0.42 | -0.19 | 0.20 |
| 4 | It is very important to me to be accepted by others. Therefore I sometimes say things of which I am not convinced that they are right, but that make me look good. | -0.06 | 0.67 | -0.17 | 0.25 |
| Functional power motive | | | | | |
| 1 | It satisfies me to influence others in their actions and attitudes so that they discover unexpected capabilities and accomplish challenging tasks. | 0.13 | -0.25 | 0.42 | 0.00 |
| 2 | It pleases me to take responsibility for a greater cause, even if that might involve experiencing setbacks and admitting mistakes. | 0.03 | -0.20 | 0.48 | 0.07 |
| 3 | I like advancing controversial views, but only if it happens in an appropriate way. | 0.58 | -0.42 | 0.49 | 0.12 |
| 4 | I enjoy to contribute something through my channels of influence that is aligned with the greater good. | 0.39 | 0.02 | 0.66 | 0.21 |
| Dysfunctional power motive | | | | | |
| 1 | I enjoy it if others have to obtain my advice or instructions before they act. | 0.02 | 0.12 | 0.17 | 0.64 |
| 2 | It pleases me to have a lot of power and influence, because there are many people that you need to keep under control. | 0.09 | 0.10 | 0.17 | 0.80 |
| 3 | It is so important for me to reach my personal goals that I would use other people for it. | -0.11 | 0.06 | 0.06 | 0.60 |
| 4 | It is a nice feeling to demonstrate my social status. | -0.21 | 0.32 | 0.12 | 0.63 |

Table S9. Four-dimensional measurement model for affiliation and power motives. Notes: Coefficients are standardized factor loadings from exploratory structural equation modeling. Hypothesized primary loadings on target motive variants are bolded. The hypothesized model fits the data well, RMSEA = 0.042, CFI = 0.98, $X^2/df = 2.68$, $X^2(62) = 166.1$, $P < 0.0001$ ($n = 961$).

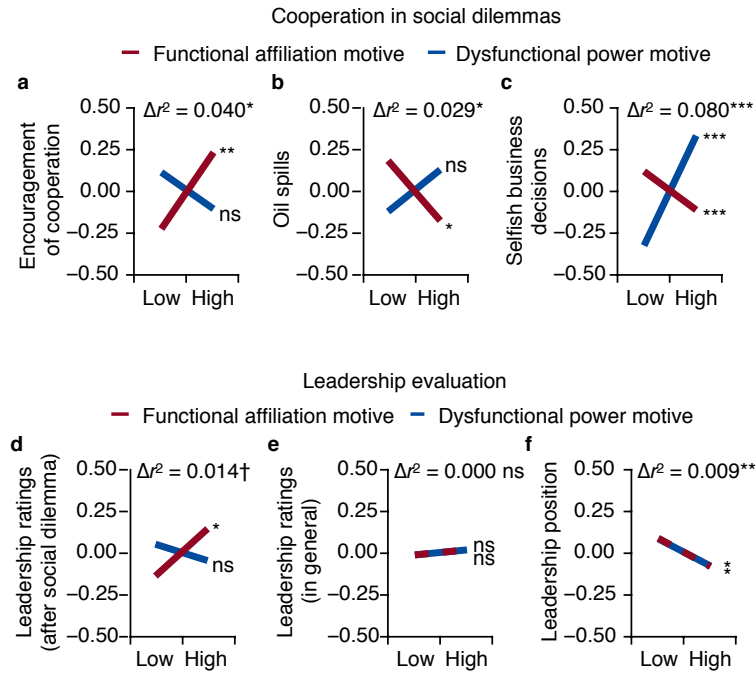


Figure S1. After accounting for an array of established predictors, the functional affiliation motive remains related to cooperation. In both studies, we control for personality (neuroticism, extraversion, openness, agreeableness, conscientiousness, fairness) and motivation to lead. **(a, b, d)** In the laboratory study (Settlers of Catan), we also control for implicit motives (need for affiliation, need for power, activity inhibition, as well as all three two-way interactions and the three-way interaction⁵⁸), an achievement motive, and reasoning ability. **(a, b)** Motives relate to behaviour during a game of Settlers of Catan ($n = 201$). **(c)** Motives relate to selfish business decisions in a field survey ($n = 960$). **(d)** After the game of Settlers of Catan, all players rate each other on transformational leadership. **(e)** In the field survey, 739 peers rate the general leadership competence of 486 respondents. **(f)** Respondents state whether they hold a professional leadership position. See Supplementary Datasets 1 and 2 or <https://osf.io/yt4qh/> for reliabilities, descriptive statistics, and intercorrelations of all variables used in each study. All values on y axes are z -standardized. Lines represent slopes from multiple regression analysis while also controlling for a dysfunctional affiliation motive and a functional power motive. Low/high ± 1 s.d. *** $P < 0.001$, ** $P < 0.01$, † $P < 0.10$, two-sided t -tests. ns, not significant.

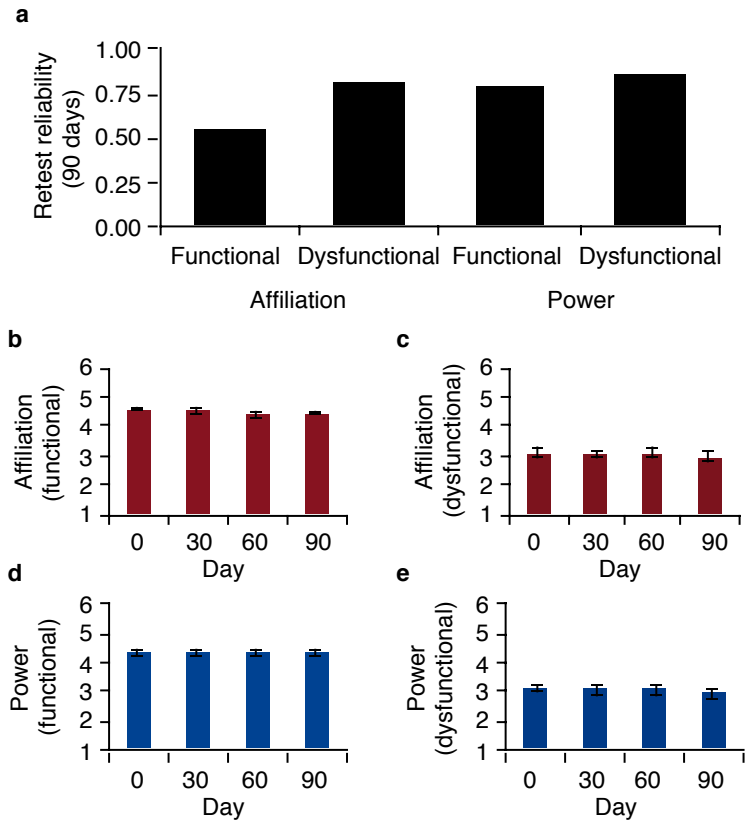


Figure S2. Stability of affiliation and power motives over a period of 90 days. **(a)** Intraclass correlations between the four measurement occasions every 30 days with measurement occasions nested within participants ($n = 35$ participants, 123 individual data points, 12% missing). **(b-e)** Mean values for each motive variant on each measurement occasion ($n = 35$ on day 0, $n = 29$ on day 30, $n = 30$ on day 60, $n = 29$ on day 90). Error bars indicate ± 1 s.e.m.

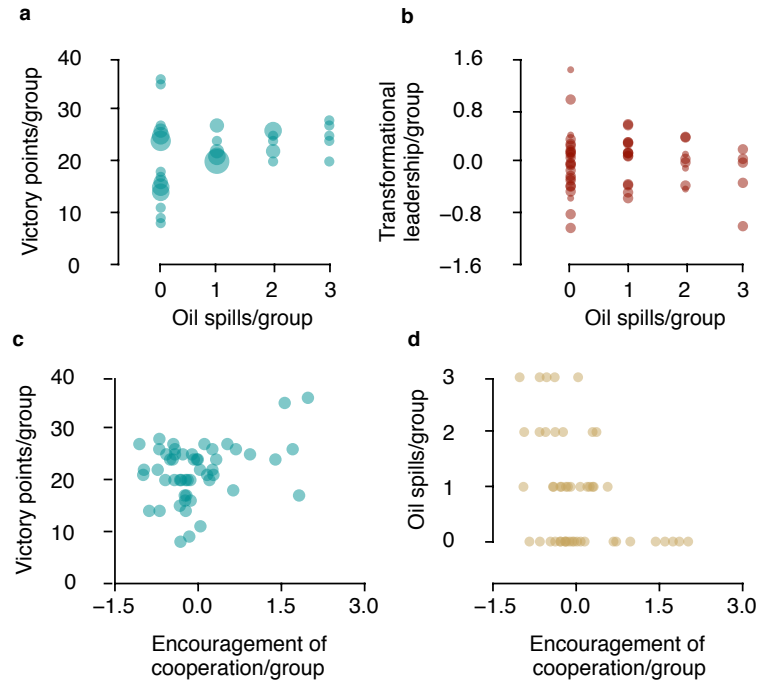


Figure S3. At the group level, high performance depends on encouragement of cooperation but not on oil spills ($n = 52$). **(a)** Group performance does not depend on oil spills ($\beta = 0.09$, $P = 0.46$). **(b)** Oil spills relate negatively to mutual ratings of transformational leadership ($\beta = -0.21$, $P = 0.039$). **(c)** Group performance depends on the average level of encouragement of cooperation in a group ($\beta = 0.34$, $P = 0.0014$). **(d)** In groups with high levels of encouragement of cooperation, fewer oil spills are caused ($\beta = -0.41$, $P = 0.0014$). This means that less fields are destroyed and, in turn, more resources will be available in the future. We use two-sided t -tests and control for group size in all analyses.

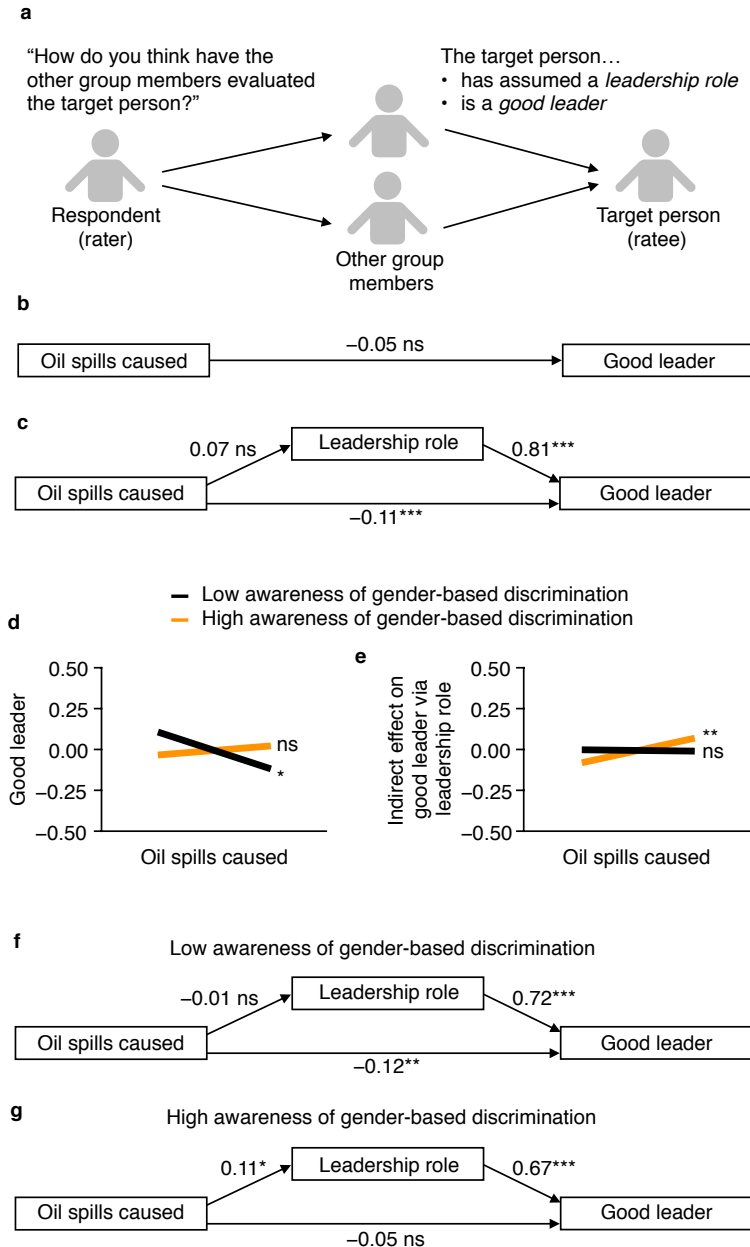


Figure S4. Respondents with high awareness of gender-based discrimination believe that others fail to disapprove of selfish behaviour. They believe that other group members think that (predominantly male) target persons who cause oil spills assume a leadership role and, in turn, are good leaders. **(a)** Schematic explaining the nature of second degree ratings (i.e., asking respondents to speculate about the thoughts of others⁴¹; Supplementary Information, Section 7). **(b)** Overall, no substantial relationship exists between the number of oil spills a target person causes and the guessed leadership rating that the target person receives from his/her group members after the game of Settlers of Catan. **(c)** Respondents hardly believe that group members evaluate those who cause oil spills as assuming a leadership role ($n = 201$ players). **(d-g)**, However, respondents' awareness of gender-based discrimination itself does promote the belief that other group members show a stereotypical pattern of evaluation ($n = 582$ triads, displayed are conditional effects). **(d, f)** Respondents with *low* (-1 s.d.) awareness of gender-based discrimination believe that others disapprove of (predominantly male) players who cause oil spills. **(e, g)** In contrast, respondents with *high* ($+1$ s.d.) awareness of gender-based discrimination believe that other group members evaluate (predominantly male) players who cause oil spills as assuming a leadership role, and in turn as good leaders. All coefficients are z -standardized. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, two-sided t -tests. ns, not significant.

1. Construction and selection of questionnaire items to measure affiliation and power motives

1.1 Item development

As far as we know, no questionnaire exists that allows a separate assessment of functional and dysfunctional variants of affiliation and power motives. When conceptualizing the two variants of the affiliation motive, our most relevant source were McClelland and Burnham¹. For the power motive, there is an array of theoretical arguments about its duality²⁻⁵. We considered ideas of McClelland and Burnham¹, Winter³, and Magee and Langner⁶ when we developed the initial item pool. We arrived at an initial item pool of 35 items (functional affiliation motive: 6 items, dysfunctional affiliation motive: 7 items, functional power motive: 9 items, dysfunctional power motive: 13 items). These numbers reflect the preponderance of the power motive in the literature as compared to the affiliation motive. We tolerated complex phrasing of an item if we deemed it useful for distinguishing the functional variant of a motive from its dysfunctional variant.

1.2 Sample for item selection

We then reduced the number of items from 35 to 16. For item selection, we used responses of the first $n = 201$ respondents of the field survey plus an additional $n = 111$ participants of the laboratory study (resulting in $n = 312$ individuals, 53% female, $M_{\text{age}} = 26$ y, s.d. = 14). After determining the final set of 16 items, only those items were presented to the remaining $n = 758$ respondents of the field survey.

1.3 Criteria for item selection

When we reduced the number of items from 35 to 16, we attempted to maximize the following criteria: (i) high correlation between an item and its scale, (ii) relatively lower correlation between an item and the other scales, particularly those with either the same motive (but a different variant) or the same variant (but a different motive), and (iii) meaningful coverage of the construct of interest. For example, criterion (ii) led us to drop the item “if someone is well disposed to me I like to reward that with little somethings or favors” as an indicator for the dysfunctional affiliation motive given that it was substantially related with the dysfunctional power motive. As another example, criterion (iii) influenced our decision to drop the item “as a member of a group, I like representing it at public events” given that it did not unambiguously fit our definition of the functional variant of the affiliation motive. Supplementary Table S7 shows a list of the 16 items that we finally selected.

2. Psychometric evaluation of scales to measure affiliation and power motives

2.1 Unidimensionality of each motive variant

Unidimensionality is present if all questionnaire items that are selected as indicators for a particular construct covary only along one single dimension. This implies that no subgroups of indicators can be identified that covary along another dimension. One then infers that the observed indicators covary only because they measure the same underlying construct, i.e., the one that they are intended to measure. In this case, residual variance of each item is idiosyncratic⁷. Researchers interpret unidimensionality as an indication that they can aggregate single questionnaire items into a joint measurement model. When unidimensionality is present, they assume that this measurement model meaningfully represents the construct of interest⁷. In the framework of generalizability theory, this relates to the item facet of generalizability. Users of a measure want to be able to generalize from the sampled items to the unsampled item space representing the construct of interest.

We evaluated unidimensionality for each of the four motive variants. We applied minimum rank factor analysis⁸ and calculated the ratio of the explained common variance (ECV) of the first factor compared to all remaining factors⁹ using the program FACTOR¹⁰. We interpret ECV as the closeness of a scale to unidimensionality¹¹. We used data from the field survey, because the sample was larger and more diverse as compared to the laboratory study. The ECV for the functional affiliation motive (78%) was slightly lower than the ECV for the dysfunctional affiliation motive (89%), the functional power motive (96%), and the dysfunctional power motive (90%). In all cases, parallel analysis¹² suggested a unidimensional solution.

2.2 Average inter-item correlations within each motive variant

The average correlation of all items of a scale with each other is an important property of a scale. For a construct that is conceptually narrow, high values are desirable whereas lower values are desirable for a broader construct¹³. There is a tradeoff known as bandwidth-fidelity dilemma¹⁴ between redundancy (in the case of high correlations) versus low content saturation (in the case of low correlations) which is indicative of

a higher proportion of measurement error in the items⁷. For broad constructs such as motives, we do not expect inter-item correlations at the top end of the recommended 0.15 to 0.50 range¹³. When keeping the number of items constant, average inter-item correlations directly transform to Cronbach's alpha. In the field survey, we find low inter-item correlations for the functional affiliation motive (0.23, $\alpha = 0.53$), intermediary values for the dysfunctional affiliation motive (0.34, $\alpha = 0.67$) as well as the functional power motive (0.33, $\alpha = 0.66$), and substantial inter-item correlations for the dysfunctional power motive (0.41, $\alpha = 0.74$).

2.3 Discriminant validity between motive variants

We intended that each of the four motive variants measures a construct that is theoretically different from the other three motive variants. This needs to be reflected in relationships between the constructs that are lower than the relationships of the indicators within each construct. As a direct test of discriminant validity, we calculated the heterotrait-monotrait ratio of correlations¹⁵. This ratio compares the average correlations of items within a particular scale to the average correlations of the items of that scale with items of another scale. As presented in Supplementary Table S8, analyses of data from the field survey support discriminant validity of the four motive variants towards each other with all ratios being substantially lower than 1. The functional affiliation motive and the functional power motive have the highest overlap (0.68). We find lower ratios (0.011 to 0.471) for the remaining pairwise comparisons.

2.4 Measurement model

In order to test the hypothesized factorial structure when considering all four scales simultaneously, we perform exploratory structural equation modeling (ESEM). We choose ESEM over confirmatory factor analysis, because we expect many items to be related to more than one motive variant. In particular, we expect that some items not only capture variance of their focal motive variant, but also reflect variance of adjacent motive variants (i.e., the same motive or the same functionality). It is common that single items are imperfect indicators of a single construct^{16, 17}. ESEM is suited to multidimensional questionnaires that contain imperfect indicators. ESEM integrates exploratory factor analysis and confirmatory factor analysis. Through factor rotation, ESEM allows researchers to target cross-loadings as close to zero as possible¹⁸.

We use the robust weighted least square estimator implemented in *Mplus* 7.3 to account for the ordinal response scale of the items. We collapse across extreme response categories so that each response category contains at least 5% of the responses to increase stability of estimation¹⁹. We evaluate model fit via root mean square error of approximation (RMSEA) and comparative fit index (CFI). Values lower than 0.06 for RMSEA and higher than 0.95 for CFI indicate good model fit.

Results show that a four-dimensional ESEM fits the data well, RMSEA = 0.042, CFI = 0.98 with a X^2/df of 2.68, $X^2(62) = 166.1$, $P < 0.0001$. As presented in Supplementary Table S9, primary factor loadings on the focal motive variant are all substantial, $\lambda_{\text{median}} = 0.57$, range = 0.36 to 0.80, $t_s > 9.95$, $P < 0.0001$ whereas most cross-loadings on the adjacent motive variants are small, $\lambda_{\text{median}} = 0.17$, range = 0.001 to 0.581.

2.5 Retest-reliability of each motive variant

We assessed retest-reliability of all scales to evaluate whether the scales capture stable interindividual differences in affiliation and power motives or whether they fluctuate heavily over time. If affiliation and power motives exhibit a considerable degree of stability, then it is more likely that they persistently influence individuals' choices and behaviours. We recruited $n = 35$ participants (80% female, $M_{\text{age}} = 26$ y, s.d. = 7, 91% students) who completed the scales for affiliation and power motives four times within a period of 3 months (30-32 days between each measurement occasion) resulting in a total number of 123 (of 140) data points (12% missing). We conducted multilevel analysis with the four measurement occasions (level 1) nested within participants (level 2). In this analysis, the intraclass correlation coefficient (ICC) of the null model reflects the amount of variance that can be attributed to individual participants in relation to the total amount of variance observed over the period of 3 months (higher ICCs indicate higher stability over time). Stabilities are relatively high with ICCs ranging between 0.55 and 0.86 (functional affiliation motive: 0.55, dysfunctional affiliation motive: 0.81, functional power motive: 0.79, dysfunctional power motive: 0.86, Supplementary Fig. S2).

3. Modifications to the procedure of the game of Settlers of Catan

3.1 Standardization

We wanted to be able to compare individuals' behaviour during the game between all groups. For this purpose, we took multiple steps to standardize the procedure of the game. First, instead of actual dice, we used chips with printed numbers on them. Instead of rolling the dice, the player whose turn it was turned the chip. The numbers on the chips substituted the numbers on a dice. We determined the order of the numbers on the chips at random and then modified the numbers itself so that each number (from 2 to 12, representing two dices with

6 numbers each) occurred in the frequency that would be expected on average. Given that participants did not know how long the game lasted, they were unable to exploit this fact. We divided the chips into multiple piles and provided extra chips (exceeding the actual duration of the game) so that participants could not predict the duration of the game from the height of the pile of chips. We prepared another set of chips to determine type and location of an oil spill. We allowed a maximum of 3 oil spills (Supplementary Fig. S3). Moreover, we simplified the set of development cards so that only knight cards and monopoly cards were available. Both occurred in identical proportions (so that each type had a probability of 50%). Developmental cards were provided in the same order in each group.

3.2 Information sheet with hints

We provided all participants with a document containing the most important information, including advice about useful strategies for playing the game. This step was designed to advance all participants to a similar level with regard to their understanding of basic principles of the game.

3.3 Goal of the game

The original rules of the game define that the game always has exactly one winner (i.e., the person who first accumulates 12 victory points). We eliminated this rule. Instead, we informed participants that the game ended after a fixed amount of rounds which we kept secret. Furthermore, we stated that the goal of the game is to populate the island as well as possible. We pointed out that participants could freely choose if they would like to support the other members of their group or even obstruct them. We suggested that there was no one “right” strategy. We informed participants that they received a payoff that was either based on their individual victory points or on the group average (decided by a coin flip after the game). On the information sheet, we advised participants that it might be profitable to collaborate with the other group members.

4. Utility of cooperation in the game of Settlers of Catan

The interpretation of the main results of this paper—i.e., how motives shape cooperation and how motives affect leadership evaluations—somewhat depends on the utility of cooperation under the specific circumstances of our studies. Is it better to cooperate or to act selfishly in the situations that we examine in our studies? In this section, we discuss this question with respect to the laboratory study. In the field survey, we believe that most people agree that good leaders avoid the kinds of selfish business decisions measured there (see ref. 20 for the wording of the scenarios that we used).

4.1 Utility of cooperation for the individual

As a proxy for the utility of cooperation in the laboratory study, we use the number of victory points which participants earned in the game of Settlers of Catan. Victory points, in turn, determine the financial payoff to participants. Given that an actual coin toss after the game decided whether we payed participants based on their own victory points or on the average number of victory points of all group members, we calculated an expected value for the individual payoff that averages both types of payoff.

We find that causing an oil spill is actually profitable for the individual who causes the oil spill, $\beta = 0.21$, $t = 3.12$, $P = 0.0021$. Oil spills are a direct consequence of the use of oil, which buys resources and buildings so that victory points are earned faster than without using oil. Translated to raw values, this means that each oil spill earns the individual who causes it an average of €0.40 (s.e.m. = 0.13). Such individual benefits associated with oil spills are to be expected, given that the underlying oil use offers powerful short-term advantages in the game that are meant to be similar to the real-world phenomena that are modeled in the game²¹. It is characteristic for one-shot public goods games in general that acting selfishly results in a higher payoff for the selfish individual.

4.2 Utility of cooperation for the group

In contrast to benefits for the individual, causing oil spills does not increase the sum of the victory points of all members of a group, $\beta = 0.09$, $t = 0.74$, $P = 0.46$ ($n = 52$ groups, Supplementary Fig. S3). On a descriptive level, the data show that the two highest performing groups are able to achieve such performances without causing a single oil spill, indicating there are ways to success beyond oil. When analyzing individual data ($n = 201$), we find a non-significant negative relationship between oil spills and the average number of victory points of the other group members, $\beta = -0.08$, $t = 1.15$, $P = 0.25$.

Furthermore, victory points do not reflect the future loss of resources that is due to the permanent damage from oil spills as defined by the rules of the game. As participants did not know how long the game lasts, they could not objectively predict how severely this future loss of resources would affect them and their group members. Anecdotal evidence from observing the groups suggests that group members never reacted positively

whenever one of the players decided to cause an oil spill. It seems as if the other group members understood that they had to pay the price for the behaviour of the selfish individual. In fact, oil spills relate negatively to the group mean of ratings of transformational leadership, $\beta = -0.21$, $t = 2.13$, $P = 0.039$ ($n = 52$, Supplementary Fig. S3).

In contrast to oil spills, verbal encouragement of cooperation was an unambiguously positive behaviour. Supplementary Fig. S3 shows that the group mean of verbal encouragement of cooperation positively relates to group performance both in terms of victory points, $\beta = 0.34$, $t = 3.39$, $P = 0.0014$, as well as in terms of future availability of resources, $\beta = 0.41$, $t = 3.38$, $P = 0.0014$ (or, in other words, in terms of avoidance of oil spills).

In summary, these findings suggest that oil spills yield small benefits for an individual in terms of financial profit. For the whole group, however, oil spills do not increase financial profit (based on aggregate victory points) but decrease future availability of resources (which unfolds more and more, the longer the games lasts). Two groups were able to achieve outstanding performance without causing a single oil spill. Verbal encouragement of cooperation is effective both with regard to financial profit and future availability of resources.

5. Awareness of gender-based discrimination moderates evaluation of oil-spill causing players

5.1 Analytical approach

This section provides more details for the analysis presented in Fig. 4c-f, i.e., whether awareness of gender-based discrimination links to systematic differences in the tendency to rate oil-spill causing players as assuming a leadership role and, in turn, as transformational leaders. We use the process macro for SPSS (v. 2.16.1) for mediation and moderation analyses. In all multiple regression analyses, we include the same control variables as throughout the manuscript (group size: 3 or 4 players, baseline values of all ratings). In all analyses of moderated mediation, we specify the least restricted models which include interaction terms for all three paths that are displayed in Fig. 4b, e, f.

5.2 Interaction effects

In addition to the findings already reported in the main text, here we only report interaction terms. In an unmediated moderation model predicting ratings of *transformational leadership* by the number of oil spills a ratee causes, the interaction with a respondent's awareness of gender-based discrimination is $\beta = 0.06$, $t = 1.98$, $P = 0.048$ (Fig. 4c). In a mediated moderation model predicting ratings of *assumed leadership role* by the number of oil spills a ratee causes, the interaction with rater's awareness of gender-based discrimination is $\beta = 0.10$, $t = 2.95$, $P = 0.003$. When predicting ratings of *transformational leadership* by *assumed leadership role*, the interaction with a respondent's awareness of gender-based discrimination is $\beta = -0.10$, $t = -3.24$, $P = 0.0013$. When predicting ratings of *transformational leadership* by the number of oil spills a ratee causes, the interaction with a respondent's awareness of gender-based discrimination is $\beta = 0.05$, $t = 1.75$, $P = 0.081$. The resulting differences in conditional regression weights are displayed in Fig. 4e, f.

6. Data analysis using multilevel analysis

The data from the laboratory study possess a multilevel structure in that (i) individual participants are nested within groups and (ii) ratings are nested within ratees. Multilevel analysis might therefore appear to be the method of choice for data analysis. In this section, we first provide reasons why we do not choose multilevel analysis in our particular cases. Second, we still report results using multilevel analysis for the interested reader. Our findings converge with those using standard regression analysis, even though some coefficients are slightly smaller.

6.1 Modeling participants as being nested within groups using multilevel analysis

6.1.1 Assumed necessity of multilevel analysis. In a typical scenario for multilevel analysis, being member in a particular group (e.g., a class at school) acts as a third variable that—over an extended period of time—potentially influences multiple variables of all members of that group (e.g., a particular teacher influences motivation and performance of all students of one class). Furthermore, group membership is often not randomly distributed (e.g., students of one particular school often share similar socio-demographic backgrounds). Group membership can thereby cause or represent spurious correlations between variables on the level of the individual members of a group which would otherwise be unrelated. Due to this kind of impact of group membership, it is often appropriate to account for group membership using multilevel analysis.

In our case, however, we do not expect this impact of group membership to be present. Group membership was temporarily determined so that it was only effective during the time of interaction in the laboratory. Group members did not know each other before meeting in the laboratory (the average degree of familiarity between group members was $M = 1.2$, $s.d. = 0.6$, on a scale of 1 to 6). Furthermore, our software composed groups basically at random by offering available time slots to participants via automated emails. In these circumstances, we interpret all similarities within groups as being a consequence of the effect of group members' characteristics on group interaction (group members' characteristics determine who influences others and who accepts being influenced). We do not interpret similarities within groups as being a consequence of third variables at the group level. Whereas multilevel analysis partials out similarities within groups from analyses on the level of individuals, we believe that in this particular case, similarities within groups constitute valid portions of variance that are to be explained by independently measured characteristics of participants (such as their motives). In our case, participants' characteristics (and their effect on group interaction) are the predominant factor determining behaviour (considering that there are no discernible third variables at the group level). We prefer to interpret similarities within groups as being the consequence of group members' characteristics rather than being the consequence of arbitrary third variables associated with group membership. From this perspective, standard regression analysis serves our purpose better than multilevel analysis. For the interested reader, we still report the results using multilevel analysis. Our findings converge with those using standard regression analysis.

6.1.2 Null models for all dependent variables from the laboratory study using multilevel analysis. First, we report null models for all three dependent variables of the laboratory study. For each dependent variable, null models allow to determine whether the variability between groups constitutes a significant portion of the total variability. This indicates that members of a group are more similar to each other compared to members of other groups. The intraclass correlation coefficient (ICC) is a measure of similarity within groups. As throughout the manuscript, we include group size (3 vs. 4) as a control variable. For encouragement of cooperation, we find an ICC of 0.31 based on a significant amount of variance between groups, $\sigma^2 = 0.31$, Wald's $Z = 3.16$, $P = 0.0016$. For oil spills, we find an ICC of 0.00 indicating that we cannot systematically attribute portions of variance in oil spills to group membership. For ratings of transformational leadership after the game (controlling for baseline values), we find an ICC of 0.09 based on a not significant amount of variance between groups, $\sigma^2 = 0.04$, Wald's $Z = 1.28$, $P = 0.20$.

6.1.3 Relationship between motives and encouragement of cooperation using multilevel analysis. Second, based on these ICCs, we include affiliation and power motives as predictors on level 1 with encouragement of cooperation as the dependent variable. Similar to the findings displayed in Fig. 1a, the functional affiliation motive positively relates to encouragement of cooperation, $\beta = 0.23$, $t = 3.64$, $P = 0.0004$, whereas the dysfunctional power motive negatively relates to encouragement of cooperation, $\beta = -0.17$, $t = -2.87$, $P = 0.005$. This finding indicates that multilevel analysis yields an identical pattern of results in this case as compared to standard multiple regression analysis.

6.2 Modeling ratings as being nested within ratees using multilevel analysis

6.2.1 Assumed necessity of multilevel analysis. For the analyses reported in Fig. 4c-f, Supplementary Fig. S4, and Supplementary Information, Section 5, one might assume that this is a case to apply multilevel analysis because raters are nested within ratees (each ratee receives ratings from the other 2-3 individuals who are in the same group as the ratee). In contrast to standard regression analysis, multilevel analysis accounts for dependencies between units of analysis that are due to membership in a higher level entity (e.g., multiple ratings belonging to a single ratee). By accounting for dependencies in individual observations, multilevel analysis avoids underestimation of standard errors. It achieves this by estimating a separate model for each higher level entity (here: for each ratee).

However, we argue that in the present case, this complexity is not necessary and might even reduce statistical power. In our opinion, multilevel analysis is not necessary in this case because all observations used to estimate our parameter of interest—the prediction of leadership ratings by the interaction between a respondent's awareness of gender-based discrimination and the number of oil spills a target person causes—are actually independent from each other once the main effect of oil spills is included in the regression model. This is because awareness of gender-based discrimination was measured *before* raters were assigned to ratees. Ratees therefore cannot have influenced raters' awareness of gender-based discrimination.

6.2.2 Assumed impact of multilevel analysis on statistical power. In the present case of analysis, we furthermore argue that multilevel analysis reduces statistical power because it controls for the mean level on the dependent variable of each ratee. From this follows that the total amount of error variance is increased by introducing error variance from group composition (via group composition, raters are assigned to ratees) which is a type of sampling error. For instance, imagine a group with a ratee causing multiple oil spills and three group members who—by chance—all have high awareness of gender-based discrimination and who all (as a consequence of

their awareness of gender-based discrimination) rate that particular target person as being influential. In comparison to the other ratees, multilevel analysis would model the extremity of these ratings as being a mere effect of the ratee for whom multilevel analysis estimates a unit-specific error term (whereas standard regression analysis treats these ratings as valid portions of variance). At this point, multilevel analysis models valid portions of variance (which could be used for comparing respondents rating *this* ratee to respondents rating *another* ratee) as error variance. From this follows that a unit contributes towards decreasing statistical power if both of the following conditions are met: (i) the distribution of the variable of interest (here: awareness of gender-based discrimination) deviates from the whole sample (due to sampling error that occurred during group composition where raters are assigned to ratees) and (ii) the unit mean on the dependent variable (of one particular ratee) deviates from the grand mean on the dependent variable (over all ratees) so that a unit-specific error term is modeled. Multilevel analysis does not compare each rater to all other raters who participated in the study, but only to those raters who rate the same ratee. Compared to the whole sample, raters in the same group can be particularly similar or dissimilar to one another due to random error that occurred during group composition. Error variance from random composition of groups is thereby introduced when applying multilevel analysis (which accounts for the hierarchical structure of the data) as compared to standard regression analysis (which ignores the hierarchical structure of the data). Increased error variance, in turn, reduces statistical power so that we expect statistical power to be reduced when using multilevel analysis. Even though we have argued that we believe multilevel analysis to be neither necessary nor beneficial (in terms of statistical power) in our particular case, we still report the results using multilevel analysis for the interested reader. Our findings converge with those using standard regression analysis, even though some coefficients are slightly smaller.

6.2.3 Analytical approach and null models. First, we report null models for both types of ratings displayed in Fig. 4, i.e., the dependent variable *transformational leadership* as well as the mediator variable *assumed leadership role*. As throughout the membership, we control for group size (3 vs. 4) and baseline ratings. For *transformational leadership*, we find an ICC of 0.13 based on a significant amount of variance between ratees, $\sigma^2 = 0.08$, Wald's $Z = 2.69$, $P = 0.007$. For *assumed leadership role*, we find an even higher ICC of 0.35 based on a substantial amount of variance between ratees, $\sigma^2 = 0.26$, Wald's $Z = 5.87$, $P < 0.0001$.

6.2.4 Examining the moderating role of awareness of gender-based discrimination on the evaluation of oil-spill causing players using multilevel analysis. Second, based on these ICCs, we include (i) respondent's awareness of gender-based discrimination, (ii) the number of oil spills a ratee causes, and (iii) the interaction between both as predictors on level 1 with *transformational leadership* as the dependent variable. Very similar to the finding from standard regression analysis displayed in Fig. 4c and reported in Supplementary Information, Section 5, the interaction term explains a significant share in transformational leadership, $\beta = 0.06$, $t = 1.97$, $P = 0.049$, indicating that respondents with higher awareness of gender-based discrimination evaluate oil-spill causing group members more positively as compared to respondents with lower awareness of gender-based discrimination. We repeat this analysis with *assumed leadership role* as the dependent variable. The interaction term again explains a significant share in *assumed leadership role*, $\beta = 0.08$, $t = 2.30$, $P = 0.022$ which, however, appears to be slightly lower than the finding from standard regression analysis displayed in the first paths of Fig. 4e, f and reported in Supplementary Information, Section 5. These findings indicate that multilevel analysis yields a similar pattern of results with respect to the moderating effects of awareness of gender-based discrimination as compared to standard regression analysis.

7. Potential explanations for the moderating role of awareness of gender-based discrimination on evaluation of oil-spill causing players

7.1 Evaluation of selfish behaviour in Western cultures

In a completely cooperative world, people would clearly disapprove of selfish behaviour (such as causing an oil spill). However, under the norm of self-interest which prevails in Western cultures²², there is no unanimous disapproval of selfish behaviour in leaders. In contrast, selfish behaviour can even be interpreted as a signal for status^{23, 24}. Most people tend to believe that leaders are masculine²⁵, dominant²⁶, and sometimes tyrannical²⁷. People-oriented leaders are perceived as less effective in task performance²⁸, even though this is not necessarily the case²⁹. Trustworthy leaders are perceived as less dominant³⁰. Dominant individuals are perceived as competent³¹. Competent individuals tend to be perceived as less warm³². Some people even endorse selfish leaders³³.

7.2 How awareness about stereotypes may inadvertently reproduce stereotypes

We identify three explanations why awareness of gender-based discrimination affects leadership ratings. First, people who are aware of gender-based discrimination know at least some stereotypes about leadership because

these stereotypes are very pervasive²⁵. Leadership stereotypes can be seen as societal standards. Being aware of a societal standard often correlates with internalizing it as a personal standard³⁴, which can lead to behavioural change^{35, 36}. Second, awareness of gender-based discrimination reflects some kind of *extrapersonal knowledge* which has been shown to influence automatic mental associations³⁷, which in turn predict behaviour, particularly in socially sensitive domains³⁸. Third, awareness of gender-based discrimination might elicit ratings that are conform with leadership prototypes, even if such prototypes contradict the personal opinion of the rater^{39, 40}.

7.3 How awareness may shape what we think what others think

7.3.1 Theoretical background. It is important to understand why awareness of gender-based discrimination might promote stereotypical patterns of evaluation. Such an understanding might be used to develop remedies that change stereotypical patterns of evaluation. Therefore, we test the explanations outlined above by asking raters about their beliefs about ratings from others. If the explanations above apply, then individuals with high awareness of gender-based discrimination should believe that other people evaluate selfish behaviour in the same way as they do, i.e., that other people interpret oil spills as a form of leadership behaviour by which a target person assumes a leadership role, which in turn raises other people's leadership evaluations of that target person (see Supplementary Fig. S4 for a schematic).

7.3.2 Method. We measured *respondents' beliefs about their group members' ratings* by asking respondents to speculate about the ratings that the other two group members provided about the fourth group member (Supplementary Fig. S4). This type of question is based on Keynes' guessing game (also known as the *beauty contest game*⁴¹). We asked respondents to "please estimate how the other group members presumably evaluated [first name of the fourth group member]. Try to take the perspective of the other group members." Supplementary Fig. S4 displays the two items that we used. We motivated respondents to respond accurately by offering a small incentive based on the actual accuracy of their guess (up to €0.20 per rating), which was calculated and paid at the end of the study.

7.3.3 Assumptions about others' evaluation of oil-spill causing players. In the first step, we analyze whether a respondent—irrespective of his/her awareness of gender-based discrimination—believes that a target person who causes oil spills is seen as a good leader by the other two group members and if this relationship is mediated by respondent's belief that group members think that the oil-spill causing target person assumes a leadership role ($n = 201$, aggregated within target persons).

The standardized total effect of a target person's oil spills on guessed *good leader* ratings is not significant, $\beta = -0.05$, $t = -0.95$, $P = 0.34$ (Supplementary Fig. S4). The path from target person's oil spills to guessed *assumed leadership role* is not significant either, $\beta = 0.07$, $t = 1.23$, $P = 0.22$. The path from guessed *assumed leadership role* to guessed *good leader* ratings is strong, $\beta = 0.81$, $t = 21.08$, $P < 0.0001$. The indirect path from target person's oil spills to guessed *good leader* ratings via *assumed leadership role* is not significant, $\beta = 0.059$, $z = 1.22$, $P = 0.22$, 95% CI [-0.026, 0.160]. The direct path from target person's oil spills to guessed *good leader* ratings is negative and significant, $\beta = -0.11$, $t = -3.55$, $P = 0.0005$ (Supplementary Fig. S4). These analyses indicate that respondents (irrespective of their awareness of gender-based discrimination) do not generally believe that an oil-spill causing player receives positive ratings from the other group members.

7.3.4 Moderating role of awareness of gender-based discrimination on assumptions about others' evaluation of oil-spill causing players. In the second step, we analyze whether respondent's awareness of gender-based discrimination moderates respondent's beliefs about the other two group members' opinion of the target person (Supplementary Fig. S4) in a similar way as it does for ratings from respondents themselves (Supplementary Information, Section 5). We use the disaggregated data set for these analyses with each individual triad being a data point, resulting in $n = 582$ triads (by *triad* we refer to data points that reflect a respondent's belief about his/her group members' opinion of a target person, Supplementary Fig. S4). We find that respondent's awareness of gender-based discrimination interacts with the number of oil spills a target person causes in predicting respondent's belief whether the other group members rate the target person as being a good leader, $\beta = 0.07$, $t = 2.14$, $P = 0.033$. The conditional effect of a target person's oil spills on guessed *good leader* ratings is negative for respondents with low (-1 s.d.) awareness of gender-based discrimination, $\beta = -0.11$, $t = -2.16$, $P = 0.032$, whereas it is not significant for respondents with high (+1 s.d.) awareness of gender-based discrimination, $\beta = 0.03$, $t = 0.62$, $P = 0.53$ (Supplementary Fig. S4). In this analysis, the main effects of oil spills, $\beta = -0.04$, $t = -1.19$, $P = 0.24$, and awareness of gender-based discrimination, $\beta = -0.00$, $t = -0.04$, $P = 0.97$, are not significant. These findings suggest that only respondents with *low* awareness of gender-based discrimination believe that their group members disapprove of a target person who causes oil spills. Respondents with *high* awareness of gender-based discrimination, on the other hand, believe that their group members are neutral towards a (most likely male) target person who causes oil spills.

We also analyze whether the moderating effect of respondent's awareness of gender-based discrimination is reflected in a positive indirect effect via respondent's belief that group members think that a target person

who causes oil spills assumes a leadership role (moderated mediation, Supplementary Fig. S4). We use the least restricted model in which includes interaction terms for all three paths (reflected in different conditional regression coefficients for all paths, see bottom panels of Supplementary Fig. S4). The effect of the interaction between respondent's awareness of gender-based discrimination and target person's oil spills on guessed *assumed leadership role* ratings is marginally significant, $\beta = 0.06$, $t = 1.80$, $P = 0.072$. The effect of the interaction between respondent's awareness of gender-based discrimination and guessed *assumed leadership role* ratings on guessed *good leader* ratings is not significant, $\beta = -0.03$, $t = -1.26$, $P = 0.21$. The effect of the interaction between respondent's awareness of gender-based discrimination and target person's oil spills on guessed *good leader* ratings is not significant either, $\beta = 0.03$, $t = 1.27$, $P = 0.17$. The conditional indirect effect of target person's oil spills on guessed *good leader* ratings via guessed *assumed leadership role* ratings is not significant for respondents with low (-1 s.d.) awareness of gender-based discrimination, $\beta = -0.004$, s.e.m. = 0.037, 95% CI [-0.077, 0.072] whereas it is significant for respondents with high (+1 s.d.) awareness of gender-based discrimination, $\beta = 0.075$, s.e.m. = 0.028, 95% CI [0.025, 0.134] (Supplementary Fig. S4).

More specifically, respondents with low (-1 s.d.) awareness of gender-based discrimination do not believe that their group members think that those who cause oil spills assume a leadership role, $\beta = -0.01$, $t = -0.11$, $P = 0.92$. Instead, they believe that their group members think that those who cause oil spills are bad leaders, $\beta = -0.12$, $t = -3.14$, $P = 0.0018$ (Supplementary Fig. S4). Respondents with high (+1 s.d.) awareness of gender-based discrimination, in contrast, believe that their group members think that those who cause oil spills *do* assume as leadership role, $\beta = 0.11$, $t = 2.53$, $P = 0.012$, and do not believe (to a statistically significant extent) that their group members think that those who cause oil spills are bad leaders, $\beta = -0.05$, $t = -1.63$, $P = 0.104$ (Supplementary Fig. S4). Altogether, these findings suggest that awareness of gender-based discrimination is accompanied by the belief that others make evaluations somewhat in favor of those (predominantly male) individuals who cause oil spills. In short, awareness of gender-based discrimination links to the belief that others make stereotypical evaluations.

7.3.5 Moderating role of awareness of gender-based discrimination on assumptions about others' evaluation of oil-spill causing players using multilevel analysis. Even though we have argued in Supplementary Information, Section 6 that we believe that multilevel analysis is neither necessary nor beneficial (in terms of statistical power) in our particular case, we still report the results using multilevel analysis for the interested reader. The interaction effects that were almost not significant using standard regression analysis are not significant using multilevel analysis. As we argue above, this difference between multilevel analysis and standard regression analysis is particularly present for dependent variables with high ICCs.

First, we report null models for both types of ratings displayed in Supplementary Fig. S5, i.e., the dependent variable guessed *good leader* ratings as well as the mediator variable guessed *assumed leadership role*. For guessed *good leader* ratings, we find an ICC of 0.25 based on a substantial amount of variance between rates, $\sigma^2 = 0.18$, Wald's $Z = 4.68$, $P < 0.0001$. For guessed *assumed leadership role*, we find an ICC of 0.36 also based on a substantial amount of variance between rates, $\sigma^2 = 0.26$, Wald's $Z = 5.96$, $P < 0.0001$.

Second, based on these ICCs, we include (i) respondent's awareness of gender-based discrimination, (ii) the number of oil spills a ratee causes, and (iii) the interaction between both as predictors on level 1 with guessed *good leader* ratings as the dependent variable. Similar to the finding from standard regression analysis displayed in Supplementary Fig. S4 and reported above, the interaction term explains a marginally significant share in guessed *good leader* ratings, $\beta = 0.06$, $t = 1.88$, $P = 0.061$, indicating that respondents with high awareness of gender-based discrimination believe that their group members think that oil-spill causing group members are good leaders. We repeat this analysis with guessed *assumed leadership role* as the dependent variable. The interaction term explains a share in guessed *assumed leadership role* that is not statistically significant, $\beta = 0.04$, $t = 1.19$, $P = 0.24$, and which is somewhat lower than the finding from standard regression analysis displayed in the first paths of the bottom two panels of Supplementary Fig. S4.

These findings indicate that multilevel analysis yields a relatively similar pattern of results as compared to standard regression analysis with respect to the moderating effect of awareness of gender-based discrimination on the relationship between a target person's oil spills and respondent's belief about the ratings from the other group members. However, the interaction effects on the mediating mechanism that were almost not significant using standard regression analysis are not significant using multilevel analysis. As we argue above, this difference between multilevel analysis and standard regression analysis is particularly present for dependent variables with high ICCs.

In summary, the results using standard regression analysis are in line with our explanations why awareness of gender-based discrimination affects leadership evaluations. Individuals with high awareness of gender-based discrimination believe that other people interpret selfish behaviour as a form of leadership behaviour, enabling one to assume a leadership role, which in turn is characteristic of good leaders. In short, awareness of gender-based discrimination links to the belief that others do not disapprove of selfish behaviour, such as causing oil spills.

7.4 Conclusion

These results highlight potential risks from increasing awareness of gender-based discrimination. Instead of reducing discrimination, the opposite can happen. Therefore, it seems helpful if out-dated stereotypes about leadership will not be reproduced more than necessary, but rather be substituted by new messages that contradict old-fashioned views of leadership. We recommend that organizations appreciate cooperativeness and pay close attention to motives that shape cooperation in and beyond social dilemmas.

References

1. McClelland, D. C. & Burnham, D. H. Power is the great motivator. *Harvard Bus. Rev.* **25**, 159-166 (1976).
2. Wang, F. & Sun, X. Absolute power leads to absolute corruption? Impact of power on corruption depending on the concepts of power one holds. *Eur. J. Soc. Psychol.* **46**, 77-89 (2016).
3. Winter, D. G. Taming power: generative historical consciousness. *Am. Psychol.* **71**, 160-174 (2016).
4. Sturm, R. E. & Antonakis, J. Interpersonal power: a review, critique, and research agenda. *J. Manag.* **41**, 136-163 (2015).
5. McClelland, D. C. The two faces of power. *J. Int. Aff.* **24**, 29-47 (1970).
6. Magee, J. C. & Langner, C. A. How personalized and socialized power motivation facilitate antisocial and prosocial decision-making. *J. Res. Pers.* **42**, 1547-1559 (2008).
7. John, O. P. & Benet-Martínez, V. in *Handbook of Research Methods in Social and Personality Psychology* (eds Reis, H. T. & Judd, C. M.) 339-369 (Cambridge Univ. Press, New York, 2000).
8. Ten Berge, J. M. F. & Kiers, H. A. L. A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika* **56**, 309-315 (1991).
9. Ten Berge, J. M. F. & Sočan, G. The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika* **69**, 613-625 (2004).
10. Lorenzo-Seva, U. & Ferrando, P. J. FACTOR 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Appl. Psych. Meas.* **37**, 497-498 (2013).
11. Sijtsma, K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* **74**, 107-120 (2009).
12. Timmerman, M. E. & Lorenzo-Seva, U. Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychol. Methods* **16**, 209-220 (2011).
13. Clark, L. A. & Watson, D. Constructing validity: basic issues in objective scale development. *Psychol. Assessment* **7**, 309-319 (1995).
14. Cronbach, L. J. & Gleser, G. C. *Psychological Tests and Personnel Decisions* (Univ. of Illinois Press, Oxford, 1965).
15. Henseler, J., Ringle, C. M. & Sarstedt, M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Market. Sci.* **43**, 115-135 (2015).
16. Marsh, H. W. *et al.* A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychol. Assessment* **22**, 471-491 (2010).
17. Morin, A. J. S., Arens, A. K. & Marsh, H. W. A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct. Equ. Modeling* **23**, 116-139 (2016).
18. Morin, A. J. S., Arens, A. K., Tran, A. & Caci, H. Exploring sources of construct-relevant multidimensionality in psychiatric measurement: a tutorial and illustration using the composite scale of morningness. *Int. J. Method Psych.* **25**, 277-288 (2016).
19. Brown, M. B. & Benedetti, J. K. On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika* **42**, 347-355 (1977).
20. Ashton, M. C. & Lee, K. The prediction of Honesty-Humility-related criteria by the HEXACO and Five-Factor Models of personality. *J. Res. Pers.* **42**, 1216-1228 (2008).
21. Griswold, A. High-stakes game of oil use. *Proc. Natl Acad. Sci. USA* **110**, 2685 (2013).
22. Miller, D. T. The norm of self-interest. *Am. Psychol.* **54**, 1053-1060 (1999).
23. Anderson, C., Brion, S., Moore, D. A. & Kennedy, J. A. A status-enhancement account of overconfidence. *J. Pers. Soc. Psychol.* **103**, 718-735 (2012).
24. Van Kleef, G. A., Homan, A. C., Finkenauer, C., Gündemir, S. & Stamkou, E. Breaking the rules to rise to power: how norm violators gain power in the eyes of others. *Soc. Psychol. Pers. Sci.* **2**, 500-507 (2011).
25. Koenig, A. M., Eagly, A. H., Mitchell, A. A. & Ristikari, T. Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychol. Bull.* **137**, 616-642 (2011).
26. Ensari, N., Riggio, R. E., Christian, J. & Carlsaw, G. Who emerges as a leader? Meta-analyses of individual differences as predictors of leadership emergence. *Pers. Individ. Differ.* **51**, 532-536 (2011).
27. Epitropaki, O. & Martin, R. Implicit leadership theories in applied settings: factor structure, generalizability, and stability over time. *J. Appl. Psychol.* **89**, 293-310 (2004).

28. Gartzia, L. & Baniandrés, J. Are people-oriented leaders perceived as less effective in task performance? Surprising results from two experimental studies. *J. Bus. Res.* **69**, 508-516 (2016).
29. Zhang, Y., Waldman, D. A., Han, Y.-L. & Li, X.-B. Paradoxical leader behaviors in people management: antecedents and consequences. *Acad. Manage. J.* **58**, 538-566 (2015).
30. Fruhen, L. S., Watkins, C. D. & Jones, B. C. Perceptions of facial dominance, trustworthiness and attractiveness predict managerial pay awards in experimental tasks. *Leadership Quart.* **26**, 1005-1016 (2015).
31. Anderson, C. & Kilduff, G. J. Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *J. Pers. Soc. Psychol.* **96**, 491-503 (2009).
32. Fiske, S. T., Dupree, C. H., Nicolas, G. & Swencionis, J. K. Status, power, and intergroup relations: the personal is the societal. *Curr. Opin. Psychol.* **11**, 44-48 (2016).
33. Keller Hansbrough, T. in *Emerging Challenges in Business, Optimization, Technology, and Industry* (eds Tadj, L. & Garg, A. K.) 15-20 (Springer, Cambridge, 2018).
34. Heinberg, L. J., Thompson, J. K. & Stormer, S. Development and validation of the Sociocultural Attitudes Towards Appearance Questionnaire. *Int. J. Eat. Disorder* **17**, 81-89 (1995).
35. Thompson, J. K., van den Berg, P., Roehrig, M., Guarda, A. S. & Heinberg, L. J. The Sociocultural Attitudes Towards Appearance Scale-3 (SATAQ-3): development and validation. *Int. J. Eat. Disorder* **35**, 293-304 (2004).
36. Stice, E. Risk and maintenance factors for eating pathology: a meta-analytic review. *Psychol. Bull.* **128**, 825-848 (2002).
37. Teige-Mocigemba, S., Klauer, K. C. & Sherman, J. W. in *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications* (eds Gawronski, B. & Payne, B. K.) 117-139 (Guilford Press, New York, 2010).
38. Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L. & Banaji, M. R. Understanding and using the Implicit Association Test: III. meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* **97**, 17-41 (2009).
39. Asch, S. E. Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychol. Monogr.-Gen. A.* **70**, 1-70 (1956).
40. Baron, R. S., Vandello, J. A. & Brunzman, B. The forgotten variable in conformity research: impact of task importance on social influence. *J. Pers. Soc. Psychol.* **71**, 915-927 (1996).
41. Nagel, R., Bühren, C. & Frank, B. Inspired and inspiring: Hervé Moulin and the discovery of the beauty contest game. *Math. Soc. Sci.* **90**, 191-207 (2017).
42. McClelland, D. C. & Boyatzis, R. E. Leadership motive pattern and long-term success in management. *J. Appl. Psychol.* **67**, 737-743 (1982).
43. Cornelius, E. T. & Lane, F. B. The power motive and managerial success in a professionally oriented service industry organization. *J. Appl. Psychol.* **69**, 32-39 (1984).
44. Sorrentino, R. M. & Field, N. Emergent leadership over time: the functional value of positive motivation. *J. Pers. Soc. Psychol.* **50**, 1091-1099 (1986).
45. Winter, D. G. Leader appeal, leader performance, and the motive profiles of leaders and followers: a study of American presidents and elections. *J. Pers. Soc. Psychol.* **52**, 196-202 (1987).
46. Spangler, W. D. & House, R. J. Presidential effectiveness and the leadership motive profile. *J. Pers. Soc. Psychol.* **60**, 439-455 (1991).
47. House, R. J., Spangler, W. D. & Woycke, J. Personality and charisma in the U.S. presidency: a psychological theory of leader effectiveness. *Admin. Sci. Quart.* **36**, 364-396 (1991).
48. Winter, D. G. A motivational model of leadership: predicting long-term management success from TAT measures of power motivation and responsibility. *Leadership Quart.* **2**, 67-80 (1991).
49. Winter, D. G. Power, affiliation, and war: three tests of a motivational model. *J. Pers. Soc. Psychol.* **65**, 532-545 (1993).
50. Jacobs, R. L. & McClelland, D. C. Moving up the corporate ladder: a longitudinal study of the leadership motive pattern and managerial success in women and men. *Consult. Psychol. J.: Pract. Res.* **46**, 32-41 (1994).
51. Langner, C. A. & Winter, D. G. The motivational basis of concessions and compromise: archival and laboratory studies. *J. Pers. Soc. Psychol.* **81**, 711-727 (2001).
52. Kirkpatrick, S. A., Wofford, J. C. & Baum, J. R. Measuring motive imagery contained in the vision statement. *Leadership Quart.* **13**, 139-150 (2002).
53. Schultheiss, O. C. & Brunstein, J. C. Inhibited power motivation and persuasive communication: a lens model analysis. *J. Pers.* **70**, 553-582 (2002).
54. De Hoogh, A. H. B. *et al.* Leader motives, charismatic leadership, and subordinates' work attitude in the profit and voluntary sector. *Leadership Quart.* **16**, 17-38 (2005).
55. Kazén, M. & Kuhl, J. Directional discrepancy between implicit and explicit power motives is related to well-being among managers. *Motiv. Emotion* **35**, 317-327 (2011).
56. Delbecq, A., House, R. J., de Luque, M. S. & Quigley, N. R. Implicit motives, leadership, and follower outcomes: an empirical test of CEOs. *J. Leadersh. Org. Stud.* **20**, 7-24 (2013).

57. Howard, A. in *New Developments in Goal Setting and Task Performance* (eds Locke, E. A. & G. Latham, P.) 246-261 (Routledge, New York, 2013).
58. Steinmann, B., Dörr, S. L., Schultheiss, O. C. & Maier, G. W. Implicit motives and leadership performance revisited: what constitutes the leadership motive pattern? *Motiv. Emotion* **39**, 167-174 (2015).
59. Steinmann, B., Ötting, S. K. & Maier, G. W. Need for affiliation as a motivational add-on for leadership behaviors and managerial success. *Front. Psychol.* **7**, 1972 (2016).
60. Winter, D. G. *The Power Motive* (The Free Press, New York, 1973).
61. Burnham, D. H. *Power is still the great motivator—with a difference!* Retrieved from <https://pdfs.semanticscholar.org/9711/c9e46bcc07051806e9e5de78a1a0253bc553.pdf> (1997).
62. Lukié, M. *Wie Motive den Managementenerfolg beeinflussen.* Retrieved from <https://www.managerberater.com/assets/uploads/pdf/Wie-Motive-den-Managementenerfolg-beeinflussen.pdf> (2015).
63. Winter, D. G. What does Trump really want?. *Anal. Soc. Iss. Pub. Pol.* **18**, 155-171 (2018).
64. Jenkins, S. R. Need for power and women's careers over 14 years: structural power, job satisfaction, and motive change. *J. Pers. Soc. Psychol.* **66**, 155-165 (1994).
65. Winter, D. G., John, O. P., Stewart, A. J., Klohnen, E. C. & Duncan, L. E. Traits and motives: toward an integration of two traditions in personality research. *Psychol. Rev.* **105**, 230-250 (1998).
66. Winter, D. G. *Navy leadership and management competencies: convergence among tests, interviews and performance ratings* (McBer and Company, Boston, 1979).