THE ECONOMIC JOURNAL

The Economic Journal, **132** (*October*), 2339–2365 https://doi.org/10.1093/ej/ueac023
Advance Access Publication Date: 1 April 2022

THREAT OF SABOTAGE AS A DRIVER OF COLLECTIVE ACTION*

Kris De Jaegher

A model is presented where the welfare of contributors to a public good can increase when they face an adversary who ex post sabotages their efforts. It is a best response for the adversary to maximally sabotage the smallest effort, thus increasing a defector's marginal product of effort. This creates a kink in the individual contributor's payoff function around the equilibrium effort, which can lock contributors into exerting high effort. For a sufficiently large degree of complementarity between the contributors' efforts, the adversary increases contributors' welfare. This result is robust when departing from several simplifying assumptions of the model.

Why does collective action take place, in spite of the free-rider problem? A reason often proposed in the vast literature on public goods is the presence of institutions,[1] or the presence of a principal who designs incentives (Holmström, 1982). Yet, an aspect that receives limited attention is that beneficiaries of a public good often face adversaries who are made worse off by the public good. Indeed, this is the case for canonical examples of collective action offered by Olson (1965): labour unions face employers, lobbying groups face rival lobbying efforts and cartels face consumer organisations. This paper presents a model where, at first sight counterintuitively, it is the presence of such adversaries that explains collective action.

In our model, a group of contributors can benefit from producing a local public good, but face an adversary who is worse off the more of the public good is produced. To minimise production of the public good, the adversary can ex post sabotage public-good production, by selectively undoing the individual efforts of the contributors. While ex post it is a best response for the adversary to sabotage, it is shown that the adversary by his capacity to sabotage may make matters worse, as sabotage can boost public-good production, and even improve contributor welfare. This result is obtained in spite of the fact that by partly undoing the contributors' efforts, the adversary effectively destroys some of the public good.

To explain the result, we first note that, when contributors produce the public good according to a convex technology, it is a best response for the adversary to direct all sabotage at the contributor who exerts the least effort (the 'weakest link'), as ex post such a strategy reduces the value of the public good to the largest extent. This incentive of the adversary to maximally sabotage the weakest link, increases the marginal product of effort to any contributor who freerides by exerting a lower effort than other contributors, creating a kink in the curve relating (all else equal) the individual contributor's effort to the level of public-good production. For this reason, the

* Corresponding author: Kris De Jaegher, Utrecht University School of Economics, Utrecht University, Kriekenpitplein 21-22, 3508 TC Utrecht, The Netherlands. Email: k.dejaegher@uu.nl

This paper was received on 2 February 2021 and accepted on 25 March 2022. The Editor was Heski Bar-Isaac.

I am grateful to the editor, Heski Bar-Isaac and two anonymous referees for key comments that considerably improved the paper. I thank Joyce Delnoij, Britta Hoyer and Jurjen Kamphorst for their comments on previous versions of the paper. Finally, I thank seminar participants at Paderborn University, Utrecht University and at the Max Planck Institute for Tax Law and Public Finance for their feedback. Any remaining errors are my own.

[1] Examples include money-back guarantees (Dawes *et al.*, 1986), rebate rules (Marks and Croson, 1998), voting (Bierbrauer and Hellwig, 2016) and production targets (Myatt and Wallace, 2009).

presence of an adversary turns the game between the contributors into a coordination game in which they can lock each other into high effort, and obtain higher welfare than in the absence of an adversary. If contributors play the payoff-dominant equilibrium, for a sufficiently large degree of complementarity between the contributors' efforts, limited ability of the adversary to sabotage suffices for public-good production with an adversary to approach the socially optimal production without an adversary.

Our paper relates to three existing literatures. *First*, the paper relates to a disparate literature spread across several scientific disciplines hypothesising that the presence of a common enemy fosters cooperation within groups (in short, the common-enemy hypothesis; for an overview, see De Jaegher, 2021). For instance, in political science, it has been hypothesised that government repression against opponents may backfire, in that it leads to more cooperation among the opponents (for the example of the Arab spring, see Sutton *et al.*, 2014). In economics, a version of the common-enemy hypothesis can be found in the literatures investigating whether war fosters cooperation (for an overview, see Bauer *et al.*, 2016), and whether within-firm team competition (including competition that consists of mere comparison to other teams) increases team productivity (Baer *et al.*, 2010). Common about this literature is the underlying theoretical explanation that the presence of a common enemy changes the psychology, preferences, or information of individuals, rather than their incentives.

This paper fits in a smaller literature that explains the common-enemy hypothesis as a change in incentives. Hirshleifer (1983) considered the private provision of a public good for key public-good production functions, and showed that specifically with the weakest-link production function (i.e., efforts are perfect complements), a Nash equilibrium exists where the players achieve the social optimum. He argued that the weakest-link production function becomes relevant when players face a common threat such as a natural disaster or an adversary. The intuition is that the presence of a common threat makes each player's contribution to the public good critical, in the sense that a large amount of the public good is lost when a single player defects (Hirshleifer, 1983; Harrison and Hirshleifer, 1989; Vicary, 1990).[2,3]

Our paper differs from Hirshleifer's approach by explicitly modelling the common threat as a strategic adversary, as in the model of De Jaegher and Hoyer (2016a; 2019). In De Jaegher and Hoyer, defenders take an all-or-nothing decision to defend an existing public good against attacks.[4] The current paper again differs from this model in considering continuous efforts rather than discrete efforts. With discrete efforts as in De Jaegher and Hoyer, the common-enemy effect consists of the cooperative equilibrium having a larger basin of attraction in the presence of an adversary. With continuous efforts, however, the presence of an adversary increases equilibrium effort within the cooperative equilibrium itself. Furthermore, while in De Jaegher and Hoyer

[2] Based on a similar intuition, it has been hypothesised in evolutionary biology that cooperation is more likely to evolve when organisms face the common enemy of a harsher environment (Mesterton-Gibbons and Dugatkin, 1992; De Jaegher and Hoyer, 2016b).

[3] Some other papers model the common-enemy effect as a change in incentives without using such a criticality argument. In Kovenock and Roberson (2012), the fact that two players play two disjoint Colonel Blotto games against the same adversary gives them an incentive to make transfers to each other. In the contest model of Münster and Staal (2011), the fact that players face an external contest with a third player leaves them with fewer resources for their internal contest. In Hugh-Jones and Zultan (2012), the presence of a common enemy gives players incentives to form a collective reputation.

[4] By a similar reasoning, in the model of strategic network formation by Hoyer and De Jaegher (2012), under the threat of strategic link disruption, players form a pairwise stable circle network, even though they would not have formed a network otherwise. The presence of a strategic disruptor makes each link critical, as removing a link means that the disruptor can cut the network in two halves. This result is not maintained, however, when employing the Nash equilibrium concept (Hoyer and Haller, 2019).

sabotage takes the form of a discrete attack on a single player at a time, in the current paper the adversary has a sabotage budget available that he can spread continuously over the players in any way he desires; the adversary's decision to sabotage is thereby endogenised. Finally, in De Jaegher and Hoyer the function of the players' efforts is to defend an existing public good against sabotage by an adversary, and efforts have no utility beyond defence; while higher intensity of attacks may backfire in making the basin of attraction of joint defence larger, the players are still better off in the absence of the adversary. In the current model, defence consists of overinvesting in effort, but this effort has utility to the players beyond defence; it becomes possible then that players are better off in the presence of an adversary.

Second, the paper relates to a theoretical literature that looks at the effect on the private provision of public goods (Bergstrom *et al.*, 1986) of such factors as the number of players (Olson, 1965; Nöldeke and Peña, 2020), the degree of heterogeneity of the players (Heckathorn, 1993; Ray *et al.*, 2007) or the shape of the production function of the public good (Hirshleifer, 1983; Heckathorn, 1996). Not considered in this literature is the fact that public-good production does not take place in isolation, and that contributors to a public good may face adversaries on whom the public good imposes a negative externality.

Third, our paper relates to a literature that stresses the importance for our understanding of interaction between competitors of allowing for sabotage, on top of the familiar competitive behaviour (for sabotage in tournaments, see, e.g., Charness *et al.*, 2014; for an overview on sabotage in contests, see Chowdhury and Gürtler, 2015). For instance, in contests, the individual contestant may both compete by increasing her own probability of winning, and sabotage by reducing the probability of other contestants winning (Kempa and Rusch, 2019). Our paper focuses on sabotage rather than on competitive behaviour, and differs from existing sabotage literature in several ways. We focus on sabotage of a group rather than of an individual player (which as Chowdhury and Gürtler 2015, p. 152, point out is a lacuna in the literature; Doğan *et al.*, 2019 is a recent exception).[5] The group sabotage we model is not collective sabotage of the entire group, but targeted sabotage of individual contributors within the group. Our focus is not on ex ante sabotage, where, e.g., the adversary raises the costs of the contributors (e.g., Salop and Scheffman, 1983) or diminishes their benefits (e.g., Kempa and Rusch, 2019) before they decide on their efforts, but on ex post sabotage, where the adversary undoes contributors' efforts. Finally, we do not only look at sabotage itself, but also at defence against sabotage, the latter being an understudied topic (Chowdhury and Gürtler, 2015, p. 152).

The paper is structured as follows. Section 1, which can be read independently, illustrates our main results in the simplified setting of strategic independence between contributors' efforts. Section 2 introduces the general model. Section 3 considers the benchmark case without an adversary. Section 4 contains our main results in the presence of a strategic adversary. Section 5 treats the robustness of these results. We end with a discussion in Section 6, in which we apply the model to the managerial practice of inserting elements of competition or conflict into within-firm team production.

[5] Doğan *et al.* (2019) studied a contest with a ratio contest success function (Tullock, 1980) between two two-player groups. Group members with heterogeneous productivities produce a group effort following a linear production function. Each group member can sabotage by reducing the productivity of a corresponding group member in the other group. In the benchmark without sabotage, free-riding takes place in that only the most productive group member contributes. With sabotage, both group members contribute, with the productive group member contributing less and the unproductive group member contributing more than in the benchmark. The latter effect may prevail, in which case sabotage leads to more production in the sabotaged group.

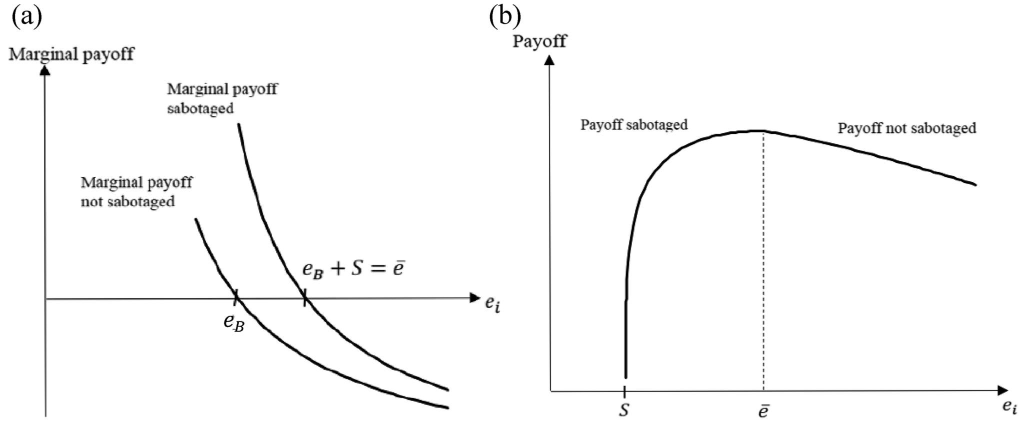

Fig. 1. *Simplified Example: Strategic Independence.*
Notes: In the case of strategic independence, marginal payoffs of a contributor i who is maximally sabotaged and who is not sabotaged (part (a)), and total payoff of such a contributor (part (b)), as a function of individual effort e_i. Any effort on the x axis constitutes a symmetric equilibrium as long as the marginal payoff when sabotaged is non-negative, and the marginal payoff when not sabotaged is non-positive. The payoff-dominant equilibrium effort $\bar{e}$ equals the equilibrium effort without an adversary, plus the sabotage budget S. The total payoff (solid curve) is depicted in part (b) when other players exert effort $\bar{e}$, and has a kink around $\bar{e}$.

1. Simplified Example: Strategic Independence

We treat here, for the case with $n \geq 3$ team members, the simplified case where efforts in the team are strategically independent. We do not further simplify this example by focusing on a team of two contributors, as we will show for the case of strategic independence that the presence of a sabotaging adversary can only increase contributor welfare when there are at least three contributors in the team.

Consider first the game without an adversary. Using symbol ℓ as an index of summation, and symbol i to refer to the typical contributor, let each contributor i of n contributors invest effort $e_i \geq 0$ in team production, where team production generates value $\sum_{\ell=1}^{n}(1/n){e_\ell}^{\alpha}$ to the contributors, with $0 < \alpha < 1$. The team thus produces value according to a convex technology, so that more is produced with a balanced effort profile than with an unbalanced one where some efforts are high and others low. Let the payoff of contributor i equal $(1/n)\sum_{\ell=1}^{n}(1/n){e_\ell}^{\alpha} - c \cdot e_i$. In the first term, each individual team member obtains an equal share of the value generated by team production. The second term reflects the costs of the individual contributor, where constant marginal costs of effort $c > 0$ are assumed.

The marginal payoff of contributor i equals $(\alpha/n^2){e_i}^{\alpha-1} - c$, which is represented as a function of e_i in Figure 1(a) ('Marginal payoff not sabotaged'). Because of the assumption that $0 < \alpha < 1$, this marginal payoff decreases in e_i. Efforts are strategically independent as the marginal payoff does not depend on other contributors' efforts. The symmetric equilibrium effort is obtained when the marginal payoff equals 0 for each contributor, which occurs for effort $e_B = [\alpha/(cn^2)]^{1/(1-\alpha)}$ (where subscript B refers to the benchmark case without an adversary). The socially optimal effort e_B^* (i.e., the effort that maximises the sum of the contributors' payoffs) can be calculated to be $n^{1/(1-\alpha)}$ times larger.

Consider now a modified game with an adversary. In this sequential game, after the contributors have chosen their effort levels, the adversary observes each effort e_i, and can sabotage each contributor i by undoing s_i of her effort, in such a manner that $\sum_{\ell=1}^{n} s_\ell \leq S$, where S is the sabotage budget of the adversary. The payoff of each contributor is now instead $(1/n)\sum_{\ell=1}^{n}(1/n)(e_\ell - s_\ell)^\alpha - c \cdot e_i$ (where we limit ourselves in this section to cases where $e_i \geq s_i$ for each i). The adversary's payoff is a negative function of the value $\sum_{\ell=1}^{n}(1/n)(e_\ell - s_\ell)^\alpha$ produced by the team, so that the adversary always puts $\sum_{\ell=1}^{n} s_\ell = S$.

As an initial way to think of this modified game, consider contributors who expect a fixed pure sabotage strategy from the adversary that is invariable to contributors' efforts. Then the marginal payoff for contributor i equals $(\alpha/n^2)(e_i - s_i)^{\alpha-1} - c$. For instance, if the fixed sabotage strategy is such that contributor i is maximally sabotaged ($s_i = S$), i's marginal payoff equals $(\alpha/n^2)(e_i - S)^{\alpha-1} - c$, which is represented as a function of e_i in Figure 1(a) ('Marginal payoff sabotaged'). Because of the assumption that $0 < \alpha < 1$, it is the case that $(\alpha/n^2)(e_i - S)^{\alpha-1} - c > (\alpha/n^2){e_i}^{\alpha-1} - c$, where $(\alpha/n^2)(e_i - S)^{\alpha-1} - c$ is constructed by shifting the 'Marginal payoff sabotaged' S to the right. As a consequence of the assumption of constant marginal costs, if these costs are sufficiently small, contributor i exerts the effort $e_B + S$ where this marginal payoff equals zero, and perfectly compensates for sabotage. Team production therefore remains exactly the same, but contributor welfare is lower given the larger effort costs.

Yet, in reality the adversary best responds to team members' efforts. After he has observed the efforts, the adversary's best response is to direct the entire sabotage budget at the contributor taking the weakly smallest effort: given that balanced efforts lead to larger team production than unbalanced efforts, and given that the adversary is better off the lower team production, he has an incentive to make the efforts as unbalanced as possible, by maximally undoing the effort of the 'weakest link' in the team.

When setting their effort levels, rational contributors anticipate the adversary's best response. Consider now a candidate symmetric equilibrium where all contributors other than i exert the same effort e. If i now sets at least as much effort as the others ($e_i \geq e$), as she is never sabotaged when increasing effort, her marginal payoff is unchanged compared to the case without an adversary, and continues to equal $(\alpha/n^2){e_i}^{\alpha-1} - c$ ('Marginal payoff not sabotaged' in Figure 1(a)). Yet, if i exerts less effort than others ($e_i < e$), as the sabotage budget is fully targeted at i, her marginal payoff becomes $(\alpha/n^2)(e_i - S)^{\alpha-1} - c$ ('Marginal payoff sabotaged' in Figure 1(a)). It follows that there is a discontinuous increase in i's marginal payoff just below e.

An equilibrium with symmetric effort e now exists when, for individual contributor i, it is the case that $(\alpha/n^2)(e_i - S)^{\alpha-1} - c \geq 0$ for $e_i < e$ and that $(\alpha/n^2){e_i}^{\alpha-1} - c \leq 0$ for $e_i \geq e$. To see why, note that if instead $(\alpha/n^2)(e_i - S)^{\alpha-1} - c < 0$ for $e_i < e$ (so that necessarily $(\alpha/n^2){e_i}^{\alpha-1} - c < 0$), e cannot constitute a symmetric equilibrium as it is a best response for the individual contributor to reduce her effort. If instead $(\alpha/n^2){e_i}^{\alpha-1} - c > 0$ for $e_i \geq e$ (so that necessarily $(\alpha/n^2)(e_i - S)^{\alpha-1} - c > 0$), e cannot constitute a symmetric equilibrium as it is a best response for the individual contributor to increase her effort.

Indicating a candidate equilibrium symmetric effort in Figure 1(a) on the x axis, it is clear that such an effort constitutes an equilibrium whenever the individual contributor's marginal payoff when sabotaged is non-negative at this effort (top curve lies on or above the x axis), and her marginal payoff when not sabotaged is non-positive at this effort (bottom curve lies on or below the x axis). It follows that in the highest-effort, payoff-dominant equilibrium, each contributor exerts effort $\bar{e} = e_B + S$. The total payoff of individual contributor i as a function of her effort e_i when all other contributors set effort $\bar{e}$ is illustrated in Figure 1(b). The part of the payoff function to the right of $\bar{e}$ reflects that contributor i is never sabotaged when exerting more effort than the other contributors. The part of the payoff function to the left of $\bar{e}$ reflects that contributor i is maximally sabotaged when taking less effort than the other contributors. Even though the curve has a negative slope approaching $\bar{e}$ from the right, this does not lead the individual contributor to reduce her effort, and her best response is $\bar{e}$. This is because the fact that the contributor with the lowest effort is sabotaged, creates a kink in the payoff function.

In the payoff-dominant equilibrium, each contributor incurs higher costs than without an adversary, as she exerts S more effort. Net of sabotage, the one contributor who is sabotaged contributes exactly the same effort to team production as without an adversary. Yet, the $(n-1)$ other contributors are not sabotaged, and therefore contribute S more effort to team production, each creating positive externalities for the other contributors. This suggests that the presence of an adversary can increase contributor welfare. This is indeed confirmed in Lemma 1 for $n \geq 3$ (all proofs can be found in Appendix A):

LEMMA 1. *With strategic independence, if all contributors set S more effort than in the absence of an adversary, and if the adversary sabotages a single contributor with the full sabotage budget S, then for teams of at least three contributors, positive sabotage budgets exist such that contributors are better off with these positive sabotage budgets than with a zero sabotage budget.*

Our main focus in the rest of the analysis is to generalise the result in Lemma 1 to a wider set of technologies for team production, including ones where contributors' efforts are strategic complements, or strategic substitutes. For instance, if team production is a routine task on which team members can mostly work separately, the individual contributor may exert less effort the more effort she expects others to exert (strategic substitutes). If team production is a non-routine task then more intense interaction between the team members may be required, with a good result only possible if all team members exert high effort. In this case, efforts may be strategic complements in that the individual contributor exerts more effort the more she expects others to exert (Rahmani *et al.*, 2018).

2. The General Model

At stage 1 of the game, n risk neutral contributors in a team (with n finite) each simultaneously choose an effort level that contributes to the level of team production Q (according to a production function to be defined below), with the effort of typical contributor i denoted e_i (also referred to as effort i), where $e_i \in \mathbb{R}_{\geq 0}$.

At stage 2, after having observed the effort profile set by the individual team members, the adversary sets for each effort e_i a sabotage level $s_i \in \mathbb{R}_{\geq 0}$ that reduces this effort (and therefore reduces team production), such that $\sum_{\ell=1}^{n} s_\ell \leq S$ (where ℓ is used to refer to the index of summation); put otherwise, the adversary chooses any sabotage vector such that the sum of all sabotage levels does not exceed his sabotage budget S.[6] An effort i may also be completely undone by sabotage. In detail, for each contributor i, if the adversary sets a sabotage level s_i, the

[6] All results extend in the following way when the adversary is not able to observe contributors' individual efforts. Team production depends on n attributes, with each contributor exerting effort in contributing to one attribute. The adversary, who observes attributes but not efforts, undoes efforts indirectly by reducing the height of individual attributes. For an example, see the devil's advocacy application in Section 6.

net effort denoted ε_i equals $\varepsilon_i = e_i - s_i$ when $e_i \geq s_i$, and equals $\varepsilon_i = 0$ when $e_i - s_i < 0$ (where e_i is referred to as gross effort whenever the distinction between e_i and ε_i needs emphasising).

We next look at the payoffs of the two types of players. Each contributor i obtains payoff $(Q^\alpha/n) - c \cdot e_i$, where the first term is i's benefit obtained from team production, and the second term is her cost of exerting effort. Here $c > 0$ is the constant marginal cost of exerting effort, assumed to be finite and identical for all contributors; Q^α is the value of team production to the contributors, where $0 < \alpha < 1$, so that additional units of team production create less and less extra value to the contributors.[7] The individual contributor obtains as a benefit a share $(1/n)$ of the value of team production.[8] The adversary obtains a payoff that is a strictly negative function of Q. It is therefore a best response for the adversary at stage 2 to choose a sabotage vector with $\sum_{\ell=1}^{n} s_\ell = S$ that minimises Q.

The level of team production is a function of the net efforts in the team, and equals $Q = [\sum_{\ell=1}^{n}(1/n){\varepsilon_\ell}^{1-\sigma}]^{1/(1-\sigma)}$; team production therefore depends on the effort profile set by the contributors, as well as on the sabotage vector set by the adversary. Here σ is the degree of complementarity[9] between the contributors' net efforts, where we assume that $\sigma \geq 0$.[10] It follows that team production is a generalised mean (or power mean; Bullen, 2003, ch. III; Rahmani *et al.*, 2018) of the contributors' net efforts, ranging from an arithmetic mean ($\sigma = 0$) to a minimum function ($\sigma = +\infty$), so that, for the degrees of complementarity considered, smaller efforts weigh more than larger ones.[11]

Our main focus is on the case $\sigma > 0$, meaning a strictly convex technology for team production. In this case, isoquants over the efforts are strictly convex and the production function is strictly quasiconcave, with as an extreme case the weakest-link production function ($\sigma = +\infty$; perfect complements), in which case $Q = \min(\varepsilon_1, \ldots, \varepsilon_j, \ldots, \varepsilon_n)$. Additionally, we consider the case $\sigma = 0$, meaning a linear production function $Q = (1/n)[\sum_{\ell=1}^{n} \varepsilon_\ell]$, in which case isoquants are linear. For $0 \leq \sigma < (1 - \alpha)$, net efforts are strategic substitutes (in the sense that contributor i's marginal payoff of effort i is smaller the larger any other net effort), and for $\sigma > (1 - \alpha)$, net efforts are strategic complements (in the sense that contributor i's marginal payoff of effort i is larger the larger any other net effort).[12] For $\sigma = (1 - \alpha)$, net efforts are strategically independent, and the simplified example of Section 1 is obtained.

The equilibrium concept employed is the subgame perfect equilibrium (Selten, 1975), where given the efforts that she expects the other contributors to take, the individual contributor sets her effort level anticipating the best response of the adversary at stage 2. In case multiple symmetric subgame perfect equilibria exist, in order to emphasise the possibility of a positive effect on

[7] Alternatively, α is interpreted as a part of the production function, which then has decreasing returns to scale.

[8] To make the public-good nature of team production (Alchian and Demsetz, 1972) concrete, the model can also be framed as a model of public-good production, where Q^α/n is the benefit each individual contributor obtains from the public good, with Q^α the sum of these benefits across contributors. To stress the non-rivalry of the public good, one can alternatively model Q^α as the benefit each individual player gets from the public good, with nQ^α the sum of these benefits. Qualitatively, this does not change our results.

[9] The CES production function adopted (cf. Arrow *et al.*, 1961; Cornes, 1993) ensures that, with symmetric efforts, σ does not affect production; when all efforts are symmetric with the exception of one lower effort, an increase in σ negatively affects production. Then σ measures the extent to which production is reduced when a single effort is decreased, or measures the extent to which each effort is critical.

[10] An alternative CES production function takes the form $Q = [\sum_{\ell=1}^{n} {\varepsilon_\ell}^{1-\sigma}]^{\alpha/(1-\sigma)}$ (e.g., Adams, 2006; Ray *et al.*, 2007). In this case, with symmetric net effort ε, $Q = n^{\alpha/(1-\sigma)}\varepsilon^\alpha$, and production is positively affected by σ. Then σ operates as a *degree of synergy*, measuring to what extent efforts reinforce each other.

[11] With $\sigma < 0$, higher efforts receive more weight, with the best-shot production function as an extreme case ($\sigma = -\infty$). Only a single contributor exerts effort then - a case is of less interest for our focus on collective action.

[12] For $\sigma = 1$, one obtains the Cobb-Douglas production function, which we do not separately consider.

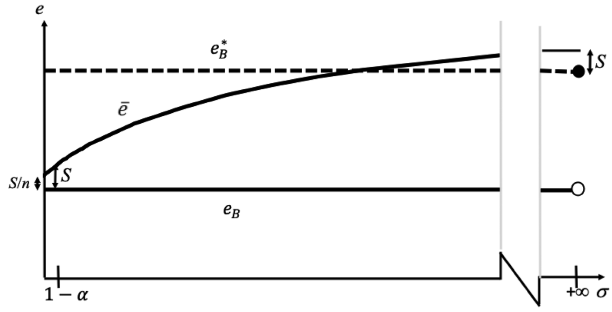

Fig. 2. *Equilibrium Efforts with and without an Adversary.*
Notes: Equilibrium Effort as a Function of the Degree of Complementarity with and without an Adversary σ, socially optimal effort level (e_B^*, dashed line) and payoff-dominant Nash effort level (e_B, solid line) for team production without an adversary. The payoff-dominant Nash effort is not affected by the degree of complementarity, with the exception of the case of perfect complementarity ($\sigma = +\infty$), where $e_B = e_B^*$. The figure represents the case $\alpha = 0.2$, $n = 2$, and $c = 0.1$, in which case $e_B = 0.42$ and $e_B^* = 1$. Additionally, the figure represents as a function of the degree of complementarity the payoff-dominant equilibrium effort $\bar{e}$ with an adversary (sabotage budget S).

contributor welfare of the presence of an adversary, we initially focus on the equilibrium that gives the contributors the largest payoff, before considering a wider set of equilibria in Section 5; a detailed analysis is contained in Section B.3 of Online Appendix B, which shows that our results are maintained as long as contributors coordinate on a sufficiently high effort. For conciseness, we refer to the equilibrium where contributors achieve the largest payoff as the payoff-dominant equilibrium, even though the payoff of the adversary is not taken into account. In the same way, when we refer to the social optimum, we mean the social optimum from the perspective of contributors, where the sum of the contributors' payoffs is maximised. Note that the social optimum is not typically an equilibrium.

3. Benchmark: Game without an Adversary

We start by treating as a benchmark the case without an adversary in Proposition 1, which is illustrated in Figure 2.

PROPOSITION 1. *Consider the game without an adversary. Then, for all considered degrees of complementarity, it is a social optimum for each contributor to exert effort* $e_B^* = [\alpha/(cn)]^{1/(1-\alpha)}$. *Furthermore,*

(*i*) *with perfect complementarity, in the payoff-dominant Nash equilibrium, all contributors exert effort* $e_B^* = [\alpha/(cn)]^{1/(1-\alpha)}$;
(*ii*) *for all considered cases other than perfect complementarity, in the symmetric payoff-dominant Nash equilibrium, each contributor exerts effort* $e_B = [\alpha/(cn^2)]^{1/(1-\alpha)}$, *where* $e_B < e_B^*$.[13]

[13] The qualifier 'payoff dominant' is needed because, for $\sigma \geq 1$, a zero-effort Nash equilibrium additionally exists. The qualifier 'symmetric' is needed because specifically for $\sigma = 0$, the symmetric equilibrium is one of many Nash equilibria where the sum of the contributors' efforts is the same; all of these equilibria are equally payoff dominant. These additional equilibria are characterised in Proposition B3 of Online Appendix B.

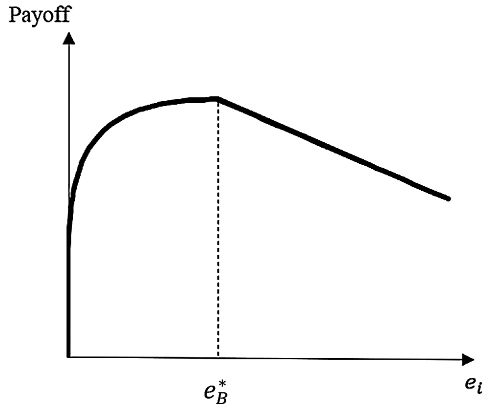

Fig. 3. *Total Payoff without an Adversary: Perfect Complements.*
Notes: In the absence of an adversary, for the case of perfect complements, total payoff of the individual contributor as a function of individual effort e_i, when other contributors are setting the payoff-dominant Nash equilibrium effort e_B^*.

Proposition 1 first looks at the social optimum, which is achieved for a symmetric effort e_B^* that is invariant to the degree of complementarity σ (dashed line in Figure 2). Intuitively, given the symmetric convex technology, symmetric efforts minimise the costs of achieving any given production level; moreover, as the part of the isoquants where efforts are symmetric is invariant to the degree of complementarity, this degree does not affect the socially optimal effort.

Proposition 1 next looks at the payoff-dominant Nash equilibrium, starting with the case of perfect complementarity. With perfect complements, if a contributor exerts lower effort than all other players, the level of production depends only on her own effort, whereas if she exerts higher effort, production depends only on others' efforts. As illustrated in Figure 3, there is therefore a kink in the individual contributor's payoff function around any candidate equilibrium symmetric effort, where the contributor's marginal payoff when exerting higher effort than others equals $-c$. Contributors can therefore lock each other into high effort, and in the payoff-dominant Nash equilibrium with the highest effort contributors achieve the socially optimal effort e_B^* (case illustrated in Figure 3).

With perfect complementarity, contributors can lock each other into high effort because each contributor's effort is critical. At the same time, the degree of complementarity is a natural measure of the extent to which contributors' efforts are critical, as lowering one of the efforts away from a given symmetric effort decreases output by more the larger the degree of complementarity. It is therefore tempting to conclude that the larger the degree of complementarity, the higher the effort that the contributors can achieve. Yet, as Proposition 1 shows, this conclusion is wrong since, with the exception of the case of perfect complementarity, equilibrium effort is invariant to the degree of complementarity (cf. Cornes, 1993). The solid line in Figure 2 represents the payoff-dominant equilibrium effort as a function of the degree of complementarity; as illustrated, this effort is discontinuous at perfect complementarity, jumping from e_B to e_B^*. Intuitively, the set of first-order conditions for the individual contributors characterising the Nash equilibrium only differ from the set of first-order conditions characterising the social optimum by the fact that benefits are divided by n. As the socially optimal effort is invariant to the degree of complementarity, the same applies to the Nash effort, with perfect complementarity as an exception.

4. Results with an Adversary

We now look at the case with an adversary. Applying backward induction, in Lemma 2 below we first look at the best response of the adversary to given effort profiles. Lemma 2 focuses on effort profiles where the smallest effort is larger than the sabotage budget, and shows that in line with the intuition stated for the simplified example in Section 1, it is a best response for the adversary to direct the full sabotage budget at the contributor exerting the lowest effort, i.e., to focus all sabotage on the 'weakest link'. A general version of Lemma 2 is contained in Appendix A, and considers any effort profile (see Lemma A1); this shows that, in general, the adversary prioritises targeting the lowest efforts.

LEMMA 2. *Let each effort be larger than the sabotage budget (i.e., for $i = 1, \ldots n$, it is the case that $e_i > S$). Then the adversary sabotages only the smallest effort, and does so with the full sabotage budget (i.e., if e_1 is the smallest effort then $s_1 = S$).*

Having described the best response of the adversary at stage 2, we are now ready to characterise the payoff-dominant subgame perfect equilibrium. In Lemma 3 below, we first establish the conditions under which the effort in this equilibrium ($\bar{e}$) is higher than the effort in the symmetric payoff dominant Nash equilibrium of the game without an adversary (e_B). These conditions are that neither effort cost c nor sabotage budget S are too high; as the individual contributor's effort may be undone, she will prefer to set zero effort if either of these are too high.

LEMMA 3. *For sufficiently small but positive sabotage budget (S) and contributor effort cost (c), the following applies: more effort is set in the payoff-dominant subgame perfect equilibrium with an adversary than in the symmetric payoff-dominant Nash equilibrium without an adversary ($\bar{e} > e_B$).*

In Proposition 2 below, we now establish by what amount the effort in the payoff-dominant subgame perfect equilibrium with an adversary exceeds the effort in the symmetric payoff-dominant Nash equilibrium of the game without an adversary, and this for each of the considered degrees of complementarity. The results in Proposition 2 are illustrated in Figure 2, which compares the payoff-dominant subgame perfect equilibrium effort with an adversary ($\bar{e}$) to the symmetric payoff-dominant Nash effort (e_B) and the socially optimal effort (e_B^*) without an adversary, as a function of the degree of complementarity.

As we established for the game without an adversary (Section 3), specifically with perfect complements, the payoff function of the individual contributor is characterised by a kink around the equilibrium effort because of the specific shape of the production function in this limit case (see Figure 3). In the game with an adversary, the payoff function of the individual contributor is similarly characterised by a kink (see Figure 1(b)), but this is because of the adversary's strategic behaviour of targeting the weakest link, and the kink is obtained for all considered degrees of complementarity, explaining why high effort is possible in general.

In detail, compared to the simple case of strategic independence in Section 1 ($\sigma = 1 - \alpha$), Proposition 2 shows that, with strategic substitutes ($\sigma < 1 - \alpha$), the payoff-dominant equilibrium effort exceeds the equilibrium effort without an adversary by less than S, and that, with strategic complements, the payoff-dominant equilibrium effort exceeds the equilibrium effort without an adversary by more than S. Intuitively, take as a starting point contributors who exert S more effort than without an adversary. Then, with strategic substitutes, the fact that other contributors exert more effort leads the individual contributor to reduce her effort level, leading to an equilibrium effort that exceeds the effort without an adversary by less than S. Approaching the linear production function, the payoff-dominant equilibrium effort with an adversary is only S/n larger. In contrast, with strategic complements, the fact that others exert more effort leads the individual contributor to exert even more effort, leading to an equilibrium effort that exceeds the effort without an adversary by more than S. Approaching perfect complements, the payoff-dominant equilibrium effort in this way approaches the socially optimal effort in the absence of an adversary (e_B^*), plus S.[14]

PROPOSITION 2. *In the payoff-dominant subgame perfect equilibrium with an adversary,*

(*i*) *with a linear production function ($\sigma = 0$), each contributor exerts S/n more effort than in the symmetric payoff-dominant Nash equilibrium with no adversary ($\bar{e} = e_B + S/n$);*
(*ii*) *with strategic independence, each contributor exerts S more effort than in the symmetric payoff-dominant Nash equilibrium with no adversary; with strategic substitutes, each contributor exerts less than S extra effort, and with strategic complements, more than S extra effort ($\bar{e} \gtreqless e_B + S$ if and only if $\sigma \gtreqless 1 - \alpha$);*
(*iii*) *approaching perfect complementarity, each contributor approaches exerting S more effort than in the social optimum with no adversary, without discontinuity at the case of perfect complements (as $\sigma \to +\infty$, $\bar{e} \to e_B^* + S$).*

In Propositions 3 and 4 below, we investigate to what extent the result that the presence of an adversary increases contributor welfare in the case of strategic independence (see Section 1) generalises to other degrees of complementarity. As a starting point, consider the benchmark case without an adversary, and the payoff-dominant Nash effort e_B that is invariant to the degree of complementarity for all other cases than perfect complementarity. Let a single effort now be sabotaged to the maximal extent, and assume that the contributors do not adapt their efforts to this fact, but keep on exerting effort e_B. Then, clearly, the larger the degree of complementarity, the lower the welfare of the contributors, as the fact that one effort is reduced by S will have more impact the larger the degree of complementarity.[15] This suggests that sabotage diminishes contributor welfare by more the larger the degree of complementarity.

Yet, as Proposition 3 shows, given that contributors strategically adapt their efforts to the presence of an adversary, if contributors coordinate on the payoff-dominant equilibrium effort, they can improve their welfare compared to the case without an adversary, and do so in particular for a sufficiently *large* degree of complementarity. As shown in Proposition 3, the degree of complementarity does not have to be considerable to achieve a welfare increase, as strategic complementarity is a sufficient condition. In line with Proposition 2, when all contributors

[14] As closed-form solutions for the payoff-dominant equilibrium effort can only be derived for the linear production function, for strategic independence, and for perfect complementarity, Figure 2 depicts this effort for all degrees of complementarity in a specific numerical example. It should be noted that the payoff-dominant equilibrium effort need not monotonically increase in the degree of complementarity. In the example of Figure 2, while this is not visible, the payoff-dominant equilibrium effort slightly decreases, approaching perfect complements.

[15] Formally, given that the adopted CES production function is a generalised mean, this result follows from the generalised mean inequality (Bullen, 2003, p. 202).

respond to the presence of an adversary by exerting S extra effort, in the case of strategic substitutes contributors further best respond to each other by lowering their effort levels. It is therefore intuitive that welfare increases occur if the degree of complementarity is not too close to the linear production function.

PROPOSITION 3. *For a sufficiently large degree of complementarity ($\sigma > (1-\alpha)(n-1)^{-2}$), an interval of sabotage budgets $(0, \bar{S}]$ exists such that, for each budget in this interval, contributors who play the payoff-dominant subgame perfect equilibrium are strictly better off in the presence of an adversary.*

Proposition 4 further calculates for key degrees of complementarity the sabotage budget that maximises contributor welfare ($S_{\max}$), assuming that the contributors coordinate on the payoff-dominant equilibrium (for other degrees of complementarity, $S_{\max}$ cannot be calculated in closed form). This confirms that in the limit case of a linear production function, while contributors may increase their effort levels with an adversary, their welfare cannot increase. With this production function, as the sum of contributors' net efforts determines team production, and as they jointly take S more gross effort, production is not changed by the presence of the adversary; however, contributors' costs are larger. Approaching perfect complements, a maximal welfare increase can be achieved with a vanishingly small but positive sabotage budget (where welfare approaches that with the socially optimal effort in the absence of an adversary). Intuitively, this is because approaching perfect complements, a small increase in effort by each contributor to compensate for small sabotage triggers contributors to exert higher efforts in response, which make it a best response to increase effort levels further, and so on.

Finally, Proposition 4(*ii*) shows that, with strategic independence (in line with Lemma 1), $S_{\max}$ is strictly positive if there are at least three contributors. As the proof of Proposition 4 shows, $S_{\max}$ in this case can be calculated in closed form. The fact that $S_{\max}$ is zero for a linear production function and approaches zero for perfect complements, but (for a sufficiently large number of contributors) is strictly positive for strategic independence, suggests that $S_{\max}$ is a hill-shaped function of the degree of complementarity.

PROPOSITION 4. *Denote by $S_{\max}$ the sabotage budget that maximises welfare for contributors who play the payoff-dominant subgame perfect equilibrium. Then,*

(*i*) *with a linear production function ($\sigma = 0$), $S_{\max} = 0$;*
(*ii*) *with strategic independence ($\sigma = 1 - \alpha$), $S_{\max} = 0$ for $n = 2$ and $S_{\max} > 0$ for $n \geq 3$;*
(*iii*) *approaching perfect complementarity ($\sigma \to +\infty$), $S_{\max} \to 0$. With contributors who play the payoff-dominant subgame perfect equilibrium, for vanishingly small but positive S, contributor welfare then approaches the welfare achieved in the social optimum in the absence of an adversary.*

5. Robustness

As our model is stylised, the question arises whether the results are robust when deviating from the simplifying assumptions taken. In this section, we summarise the conclusions from a robustness analysis in Online Appendix B, and refer to sections and propositions in this appendix.

First, we have assumed constant marginal costs for the contributors. With increasing marginal costs, intuitively contributors are less inclined to compensate for sabotage. Yet, as Proposition

B1 shows, approaching perfect complements it continues to be the case that a small increase in the sabotage budget away from zero suffices to make it possible that contributors increase their welfare. This shows that, with increasing marginal costs, our results are maintained for sufficiently large degrees of complementarity.

Second, we have assumed a fixed sabotage budget for the adversary. In a more general model set out in Section B.2, the adversary ex post also strategically sets the total sabotage level, and faces higher costs for higher total sabotage levels.[16] As long as the adversary's marginal costs of total sabotage increase to a sufficient extent and/or his marginal benefits of reducing team production decrease to a sufficient extent, it continues to be the case that, for large enough degrees of complementarity, contributors who coordinate on the payoff-dominant subgame perfect equilibrium are better off in the presence of an adversary (Proposition B2). Intuitively, the adversary who directs all sabotage at the lowest-effort contributor sets a higher total sabotage level the higher the effort of this contributor. A contributor who considers exerting less effort than the other contributors anticipates this, and the extent to which she considers her marginal product of effort to be larger when she sets an effort level just below the effort level of others is mitigated. Under the conditions specified in Proposition B2, this mitigating effect is small enough for our results to be maintained. Proposition B2 further shows that the impact of the mitigating effect is reduced in the realistic case where the adversary's marginal cost of undoing a particular contributor's effort by a certain amount is larger the higher this effort is; this is because the adversary is then less inclined to increase sabotage when the lowest-effort contributor increases her effort level.

Third, we have assumed that players are able to coordinate on the payoff-dominant equilibrium. Section B.3 extends our results to equilibria that are not payoff dominant for both the games without (Proposition B3) and with (Proposition B4) an adversary. As Proposition B4 shows, the effort in the payoff-dominant subgame perfect equilibrium is the highest of a continuum of equilibrium efforts. As contributors who play the payoff-dominant subgame perfect equilibrium are strictly better off with an adversary for a sufficiently large degree of complementarity, the same result continues to apply for a range of equilibrium efforts just below the effort in the payoff-dominant equilibrium (Corollary B1). At the same time, for all considered degrees of complementarity, contributors can coordinate on a subgame perfect equilibrium where they are worse off in the presence of an adversary (Proposition B5). Intuitively, the kink created in the individual contributor's payoff by the fact that the adversary targets the smallest effort can also lock contributors into low effort.

The question then arises to what extent our results are maintained when considering other criteria for equilibrium selection than payoff dominance, such as risk dominance (Harsanyi and Selten, 1988), where the reasoning is that contributors who are uncertain about others' efforts minimise their risk. In the context of continuous games, an equivalent to a risk-dominant equilibrium is found when deriving the equilibrium with the maximal potential (Monderer and Shapley, 1996), which we do in Proposition B6 for both the game with and without an adversary. As this proposition shows, in the game without an adversary, in the maximal potential equilibrium, players exert the effort $e_B = [\alpha/(cn^2)]^{1/(1-\alpha)}$ characterised in Proposition 1, including for the case of perfect complementarity. However, in the game with an adversary, in the maximal potential equilibrium players exert an effort lower than the effort $\bar{e}$ characterised in Proposition 2. By

[16] If the adversary can ex ante commit to an overall sabotage level, he can avoid any common-enemy effect by committing to a low sabotage level. A similar result applies in De Jaegher and Hoyer (2019), where a government commits to a low level of repression to prevent a backfiring effect where high repression increases cooperation among dissidents ('velvet-glove' strategy).

Proposition B7, a small increase in S above zero does not increase the welfare of contributors who play the maximal potential equilibrium. Still, numerical examples in Section B.3 show that, for intermediate degrees of complementarity, sabotage budgets exist such that contributors playing the maximal potential equilibrium are better off with an adversary. Intuitively, for the highest degrees of complementarity, the payoff reduction when other contributors defect is large, and with it the risk to contributors of cooperating; in the maximal potential equilibrium, this leads contributors to reduce their effort levels, explaining why the positive welfare effects of the presence of an adversary are maintained only for intermediate degrees of complementarity.

Fourth, our model assumes that the adversary has perfect information about the contributors' efforts. When the adversary does not have such information, our game effectively becomes a simultaneous-moves game, which is analysed in Section B.4. As shown in Proposition B8, in a symmetric equilibrium, the adversary still maximally sabotages a single contributor, even though this contributor is now randomly determined. Contributors set an effort level that is lower than that in the payoff-dominant equilibrium of the sequential game in Section 2, where this lower effort turns out to be equal to the effort in the maximal potential equilibrium of the sequential game. The results about the possibility of contributor welfare increases in the presence of an adversary in Section B.3 therefore continue to apply.

Fifth, we have only looked at homogeneous contributors. A variant of the game with heterogeneous contributors is analysed in Section B.5. When contributors are heterogeneous in obtaining unequal shares of the team's production, in the game without an adversary they set different effort levels (Proposition B9). While the discontinuity at perfect complementarity continues to exist, increasing the degree of complementarity without reaching perfect complementarity means that team production decreases, as the fact that the lowest-benefit contributor sets a lower effort level has more impact then. With an adversary, for sufficiently small heterogeneity, all contributors exert the same effort in the payoff-dominant equilibrium, determined by the maximisation problem of the lowest-benefit contributor (Proposition B10). This fact limits the height of the symmetric effort that contributors can achieve in the presence of an adversary. However, for sufficiently large degrees of complementarity, it continues to be the case that the presence of an adversary can make contributors better off.

6. Discussion

Previous literature has theoretically grounded the common-enemy hypothesis by arguing that the presence of a common enemy makes each player's effort critical, and incentivises each player to do her part. One may refer to this in short as the criticality argument. In particular, it is argued that the efforts of players facing a common enemy are perfect complements (Hirshleifer, 1983; Harrison and Hirshleifer, 1989; Vicary, 1990). As the level of cooperation then depends on the lowest effort among the players, they can lock each other into high effort. However, this argument does not generalise to less extreme production functions, even if these closely approach perfect complementarity (see Section 3). Moreover, if players are heterogeneous in the benefits they obtain from collective action, as one approaches perfect complementarity by increasing the degree of complementarity, equilibrium effort decreases rather than increases (see Online Appendix B, Section B.5). Finally, even if the players' efforts are perfect complements, it is plausible that players still coordinate on a low effort, as the impact of uncertainty about other players' efforts is maximal in this case. Experimental evidence suggests that with perfect complements, players are only able to coordinate on high effort in small groups, and when effort costs are low (Devetag and Ortmann, 2007). We conclude that the criticality argument may only explain the common-enemy hypothesis in very specific circumstances. Yet, our paper comes to the rescue of the criticality argument by endogenising the common enemy, which we model as a strategic adversary who ex post sabotages players' efforts in contributing to team production. In such a model, the common-enemy effect is predicted to apply in a much wider set of circumstances.

If common enemies counter free-riding in an effective way, one would expect revealed preference for assigning common enemies to groups. Such revealed preference may be found in the managerial practice of encouraging within-firm conflict (Fjermestadt, 2019) or competition (Marino and Zábojnik, 2004). A first example of this practice is found in the corporate use of *devil's advocacy*, which has been suggested as a technique for dealing with complex managerial decisions in ambiguous environments (Schweiger *et al.*, 1986; Katzenstein, 1996). For instance, let a management allow a team among its workers to lay out a plan for a new product, after which management decides whether or not to approve the plan. Because of the complexity of this decision, it is plausible that management is unaware of all possible strengths and weaknesses that the plan may have (in the sense of having no conception of them; Schipper, 2014), and takes those revealed by the team at face value (Heifetz *et al*., 2021). Given the team's interest in having the plan approved (in terms of prestige within the firm, career opportunities, etc.), it may then focus on revealing strengths. This bias may be countered by the assignment to the team of a devil's advocate, who 'assumes the role of an adverse and often carping critic', and is tasked ' . . . to determine all that is wrong with the plan and to expound reasons why the plan should *not* be adopted' (Mason, 1969, p. 407). Yet, the effectiveness of devil's advocacy has been questioned in that it is 'destructive rather than constructive' (Mason, 1969, p. 407): as the devil's advocate does not present alternatives, his criticism can only induce rejection of the plan. Management can avoid this problem and still counter the team's bias by asking an individual or another team to make a counter-plan (Schwenk, 1984). From this perspective, it is puzzling that devil's advocates continue to be assigned to within-firm teams (for examples at Anheuer-Busch, IBM and 3M, see Ivancevich *et al.*, 2008, p. 304, and Jones, 2012, p. 379).

Applying the model in this paper, we argue that devil's advocacy has the added advantage that it counters free-riding. Let the management assess each attribute of the plan (design, marketing, engineering, . . .) separately by weighing the attribute's strengths against its weaknesses, as they are revealed by the team and by the devil's advocate. Also, let the management consider the different attributes of the plan as complementary, so that perceived poor performance on a single attribute suffices to considerably decrease the probability that management approves the plan.[17] As the devil's advocate is tasked with ex post minimising the probability of approval, he does not just pick random weaknesses, but focuses all his efforts in finding weaknesses on the plan's poorest attribute. A team member in charge of a specific attribute of the plan who considers defecting from high team effort, in turn anticipates that this attribute will be singled out by the devil's advocate, in which case this team member's effort becomes to a larger extent critical in assuring approval. For this reason, the presence of a devil's advocate can lock team members into high effort.[18] Even though ex post the devil's advocate reduces the probability of approval,

[17] Such a decision process may take place implicitly or explicitly. In the latter case, management may use a scoring function to weigh strengths against weaknesses for each attribute, following a SWOT scoring methodology (Flavel and Williams, 1996). Also, the generalised mean in our model has been proposed in operations research as a basis for decision support systems in the context of multi-attribute decision-making (Wang *et al.*, 2015).

[18] Similarly, management may task a test engineer with attempts to break the prototype of a new product (so-called destructive testing; Bell, 1991). The test engineer has an incentive to put maximal pressure on the prototype's weakest spot.

ex ante he may increase it, because each team member invests in producing a larger number of strengths for the attribute for which she is in charge.[19] Moreover, more standard techniques for solving the free-rider problem may not be operable in this context. Individual punishment of underperforming team members is not an option for the management if, as is often the case in team production, it is not possible to trace back output attributes to individual team members' efforts (Blair and Stout, 1999). Imposing ex ante a quality standard for the entire plan (cf. Holmström, 1982) or similarly for its individual attributes (cf. McAfee and McMillan, 1991), where the plan is automatically rejected if this standard is not achieved, may not be viable in innovative environments such as product development, where it is impossible to anticipate every future contingency (Belloc, 2012). Devil's advocacy then functions as an internal quality control system, in spite of the fact that a quality standard cannot be imposed.

As a second case, consider management's practice of introducing *within-firm team competition*, and this in spite of the fact that this may lead teams to mutually sabotage each other. In their popular management book, Peters and Waterman (1982) hail team competition as one of the key features that distinguish successful companies from unsuccessful ones. As an example, the authors offer Procter & Gamble's practice of selling several competing brands, e.g., brands of toothpastes. In an interview gathered by the authors, a Procter & Gamble employee states: 'I remember I was a quality control manager when Crest [a Procter & Gamble toothpaste brand] was certified by the American Dental Association a few years back. The next week I ran into a brand manager from one of our other toothpastes. He said, only half kidding, 'Can't you put some bugs in that stuff?'' (Peters and Waterman, 1982, p. 217). While the authors refer to sabotage to illustrate the effectiveness of introducing within-firm team competition, others have warned against such competition, for fear that the detrimental effects of sabotage outweigh the beneficial effects of competition. Indeed, Van Knippenberg (2003, p. 383) reported anecdotal evidence that when a container terminal in the port of Rotterdam introduced team competition among dock workers, one team was caught entering the port after working hours to sabotage the other team's set-up. To avoid such detrimental sabotage, it has been argued that team competition should only be allowed between geographically distant teams (Verbeke *et al.*, 2016). Yet, our model suggests that the possibility of mutual sabotage among teams promotes collective action within teams, offering a possible explanation for why firms continue to encourage team competition.[20]

We end by noting some directions for future research. First, in our model the adversary is a single player. Yet, in terms of the product development example, pointing out the disadvantages

If management rejects the prototype conditional on the test engineer's success at breaking it, each member of the team developing the prototype realises that shirking on the prototype attribute for which she is responsible, disproportionately increases the probability of rejection.

[19] Three conditions are essential for this rationale for assigning a devil's advocate. First, weaknesses revealed by the devil's advocate should reduce the probability of approval. This is plausible when management does not have a pre-given conception of all the weaknesses that the plan may have, and cannot deduce any information from non-revelation of weaknesses, so that the unravelling argument (Milgrom, 1981) does not apply. Second, the devil's advocate must have incentives to minimise the probability of approval. Following Nemeth *et al.* (2001), this is guaranteed when the devil's advocate is an outsider who is genuinely opposed to approval. Examples in Gersen and Vermeule (2012) of delegation to enemies can be understood in this light. Third, the devil's advocate should not be able to commit ex ante; otherwise, a devil's advocate opposed to the plan who anticipates that pointing out weaknesses is counterproductive commits to not pointing out weaknesses (see also footnote 16). Inability to commit is plausible when the devil's advocate does not know ex ante about the plan; ex post, the best he can do is to target the weakest link.

[20] Not all forms of sabotage are beneficial. Ex ante sabotage where the adversary raises contributors' costs (Salop and Scheffman, 1986) unambiguously diminishes contributor welfare. Non-targeted sabotage reducing team production (e.g., bugs in toothpaste) does not have the same incentive effects as targeted sabotage. Ex post sabotage only has beneficial effects when defensive efforts to counter it directly increase team production (e.g., when dock workers safeguard their gear against tampering, there are no benefits beyond prevention of tampering).

of specific aspects of a proposed product may be a specialised task, and for this reason, the task of a devil's advocate may be taken up by a team (often called the 'red team'). Alternatively, management may let two competing teams comment on each other's products. If a specialist in one team can only comment on the attribute for which a corresponding specialist in the other team is responsible then multiple equilibria may exist where only a single specialist sabotages (Diekmann, 1985), and efficiently sabotaging the weakest link in the other team requires coordination within the sabotaging team, which deserves separate attention.

Second, the common-enemy effect in its broadest interpretation is not limited to threats by strategic adversaries, but may also apply to threats by Nature, such as natural disasters (Hirshleifer, 1983; Calo-Blanco *et al.*, 2017). In a variant of our model where contributors face a threat by Nature, each contributor may independently face a randomly distributed reduction of the effort that she invests in the team. This differs from the setting in our model, where even if the strategic adversary is uninformed, the effort of a single contributor is maximally reduced, whereas all other contributors' efforts remain intact. For this reason, the theoretical analysis of a threat by Nature in the form of an independent, random reduction in each contributor's effort also deserves separate attention.

Finally, while our analysis focuses on a formalisation of the common-enemy effect where the presence of an adversary makes contributors to team production more cooperative because the adversary changes their incentives, this does not exclude the possibility that the presence of an adversary changes contributors' social preferences or psychology. In order to identify behavioural common-enemy effects, present experimental literature designs experiments where incentive effects caused by the common enemy are eliminated (e.g., Theelen and Böhm, 2021). Future experimental work may establish the relative importance of incentive effects compared to the effect of changes in social preferences or psychology.

Appendix A: Proofs

A.1. *Proof of Lemma 1*

In the given case, contributor welfare equals

$$\frac{1}{n}{e_B}^{\alpha} + \frac{n-1}{n}(e_B + S)^{\alpha} - cn\left(e_B + S\right). \tag{A1}$$

The first derivative of (A1) with respect to S equals

$$\frac{\alpha\left(n-1\right)}{n}(e_B + S)^{\alpha-1} - cn \tag{A2}$$

and the second derivative is negative. Substituting $e_B = [\alpha/(cn^2)]^{1/(1-\alpha)}$ into (A2) and evaluating (A2) at $S = 0$, it becomes clear that (A2) is zero for $n = 2$, and is positive for $n \geq 3$. It follows that zero S maximises (A1) for $n = 2$, and positive S maximises (A1) for $n \geq 3$.

A.2. *Proof of Proposition 1*

Consider first the social optimum. By Minkowski's inequality (Hardy *et al.*, 1934), the production function is quasiconcave. In solving the social welfare maximiser's problem by first minimising costs for each possible production level note that, for any two efforts e_i and e_j, it is the case on the isoquant where level Q is produced that $(\partial Q/\partial e_i)/(\partial Q/\partial e_j) = -1$ when $e_i = e_j$. At the same time, given the constant marginal cost c, on any isocost hyperplane where the sum of the contributors' costs equals C, it is the case that $(\partial C/\partial e_i)/(\partial C/\partial e_j) = -1$ when $e_i = e_j$. As the isoquants are convex, it follows that the effort vector where all efforts are equal is cost minimising, in which case given a symmetric effort level e, we have $e = Q$ because of the form of the production function. Knowing the cost-minimising effort vector for each Q, the social welfare maximiser can now be seen as setting the optimal Q, by solving $\max_Q Q^\alpha - cnQ$. The second-order condition of this maximisation problem is valid as $0 < \alpha < 1$. The maximisation problem always has an interior solution, as the first derivative is infinite around $Q = 0$. Solving the maximisation problem yields the effort level e_B^* given in the proposition.

Consider next the actual game, and consider first finite σ. The first-order condition of the individual contributor i takes the form

$$\frac{\alpha}{n^2}\left[\sum_{\ell=1}^{n}\left(\frac{1}{n}\right){e_\ell}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1} e_i^{-\sigma} - c \;=\; 0, \tag{A3}$$

and the second-order condition is

$$\frac{\alpha(1-\sigma)}{n^3}\left[\frac{\alpha}{1-\sigma}-1\right]\left[\sum_{\ell=1}^{n}\left(\frac{1}{n}\right){e_\ell}^{1-\sigma}\right]^{\alpha/(1-\sigma)-2} e_i^{-2\sigma}\frac{\alpha\sigma}{n^2}\left[\sum_{\ell=1}^{n}\left(\frac{1}{n}\right){e_\ell}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1} e_i^{-\sigma-1}$$

$$\leq 0,$$

which can be written as

$$-\frac{\alpha}{n^2}\left[\sum_{\ell=1}^{n}\left(\frac{1}{n}\right){e_\ell}^{1-\sigma}\right]^{\alpha/(1-\sigma)-2} e_i^{-\sigma-1}\left\{\frac{1-\alpha}{n}e_i^{1-\sigma} + \sigma\left[\sum_{\ell\neq i}\left(\frac{1}{n}\right){e_\ell}^{1-\sigma}\right]\right\} \leq 0,$$

so that the second-order condition is valid. Looking at the first-order condition in (A3), given that only the coefficient $e_i^{-\sigma}$ in the first term on the left-hand side (LHS) differs across contributors, it follows that, for $\sigma > 0$, there can be at most one interior Nash equilibrium, which has symmetric efforts; in such a candidate equilibrium, by (A3) each contributor i sets $e_i = [\alpha/(cn^2)]^{1/(1-\alpha)} = e_B$. For $\sigma = 0$, such a symmetric effort profile is just one of the solutions to the first-order condition.

By the second-order condition, given that all other players exert the candidate equilibrium effort $[\alpha/(cn^2)]^{1/(1-\alpha)}$, for any smaller effort set by the individual contributor, including zero effort, the marginal product of effort is larger than the constant marginal cost. Given that there are no fixed costs to effort, exerting effort $[\alpha/(cn^2)]^{1/(1-\alpha)}$ is necessarily a best response to the individual contributor.

Consider finally $\sigma = +\infty$, and let all other contributors exert effort e_B^*, so that contributor i maximises $(1/n)\min(e_B^*, \ldots, e_i, \ldots, e_B^*)^\alpha - ce_i$. The derivative of i's payoff function with respect to e_i equals

$$\frac{\alpha}{n}{e_i}^{1-\alpha} - c \quad \text{for } e_i \leq e_B^* \quad \text{and} \quad -c \quad \text{for } e_i > e_B^*.$$

Given that $0 < \alpha < 1$, the second-order condition is valid. It follows that it is a best response for i to exert effort e_B^* as well.

A.3. *Lemma A1 and Lemma A2*

We formulate and prove a more extensive version of Lemma 2, indicated here as Lemma A1.

LEMMA A1. *Order efforts as $e_1 \leq e_2 \leq \cdots \leq e_n$. Then,*

(*i*) *with $e_1 \geq S$, it is a best response for the adversary to set $s_1 = S$. If, additionally, $e_1 < e_2$, this is the unique best response;*

(*ii*) *with $e_1 < S$, consider the largest k such that $\sum_{\ell=1}^{k} e_i \leq S$. Then it is a best response for the adversary to put $s_\ell = e_\ell$ for $\ell = 1, 2, \ldots, k$. If $\sum_{\ell=1}^{k} e_\ell = S$ then, additionally, the adversary puts $s_\ell = 0$ for every $\ell > k$; if $\sum_{\ell=1}^{k} e_\ell < S$ then she puts $s_{k+1} = S - \sum_{\ell=1}^{k} e_\ell$ and $s_\ell = 0$ for every $\ell > k + 1$.*[21]

PROOF. (i) By Minkowski's inequality (Hardy *et al.*, 1934), the production function is quasiconcave, and for a fixed sabotage vector, the upper contour sets of the form $Q \geq \bar{Q} = [\sum_{\ell=1}^{n} (1/n)(e_\ell - s_\ell)^{1-\sigma}]^{1/(1-\sigma)}$ in $(e_1, e_2, \ldots)$ space are convex. It follows that, for a fixed effort profile, the lower contour sets of the form $Q \leq \bar{Q} = [\sum_{\ell=1}^{n} (1/n)(e_\ell - s_\ell)^{1-\sigma}]^{1/(1-\sigma)}$ in $(s_1, s_2, \ldots)$ space are concave. In $(s_1, s_2, \ldots)$ space, the adversary can be seen as looking for the isoquant with the lowest Q on the hyperplane $\sum_{\ell=1}^{n} s_\ell = S$. Given that the specified isoquants are concave, and given that reducing smaller efforts has more impact than reducing larger ones, with $e_1 \geq S$, it is a best response for the adversary to direct the entire S at this effort.

(*ii*) This follows by induction from (*i*). Once the adversary has reduced net effort 1 to zero, by the same reasoning as under (*i*), the next effort to be sabotaged is effort 2, and so on. □

Before proving the results for the model with an adversary, we prove Lemma A2, establishing the conditions for existence of an interior equilibrium in the case of a fixed sabotage vector.

LEMMA A2. (CONDITIONS FOR EXISTENCE OF AN INTERIOR EQUILIBRIUM). *Consider the fixed sabotage vector with $s_i = S$, and consider any given effort profile of all other players than i. Then, for any finite S, there exist sufficiently small but positive c, and for any finite c, there exist sufficiently small but positive S, such that the best response of contributor i is to set $e_i \geq S$. Moreover, under these conditions, in any symmetric interior equilibrium, ε_i is invariant to S.*

PROOF. Consider contributor i's payoff function as a function of ε_i, where her total costs can then be reinterpreted as consisting of fixed cost cs_i and a variable cost $c\varepsilon_i$. We first establish that, with $\sigma > 0$, the marginal benefit of contributor i is infinite at $\varepsilon_i = 0$. The marginal benefit equals

$$\frac{\alpha}{n^2}\left[\sum_{\ell=1}^{n}\left(\frac{1}{n}\right)\varepsilon_\ell^{1-\sigma}\right]^{\alpha/(1-\sigma)-1} \varepsilon_i^{-\sigma}, \tag{A4}$$

from which the result immediately follows for $\sigma = (1 - \alpha)$. For other values of σ, rewrite (A4) as

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)\varepsilon_i^{-(1-\alpha)(1-\sigma)/[\sigma-(1-\alpha)]} + \left(\frac{1}{n}\right)\varepsilon_i^{-\sigma(1-\sigma)/[\sigma-(1-\alpha)]}\sum_{\ell\neq i}\varepsilon_\ell{}^{1-\sigma}\right]^{[\sigma-(1-\alpha)]/(1-\sigma)}. \tag{A5}$$

For $0 < \sigma < (1 - \alpha)$ and $\sigma > 1$, it is the case that $-(1 - \alpha)(1 - \sigma)/[\sigma - (1 - \alpha)] > 0$, $-\sigma(1 - \sigma)/[\sigma - (1 - \alpha)] \geq 0$ and $[\sigma - (1 - \alpha)]/(1 - \sigma) < 0$. It follows that, for $0 \leq \sigma < (1 - \alpha)$ in general, and for $\sigma > 1$ conditional on $\varepsilon_j > 0$ for $j \neq i$, at $\varepsilon_i = 0$ (A5) equals

[21] With $\sigma \geq 1$, this is just one of the best responses of the adversary. Any sabotage strategy that reduces the lowest effort to zero is a best response, as this automatically means that production is zero.

(α/n^2) times 0 to a negative power, and therefore equals plus infinity (where we note that, for $\sigma > 1$, the result is guaranteed by the fact that $\alpha < 1$). For $(1-\alpha) < \sigma < 1$, it is the case that $-(1-\alpha)(1-\sigma)/[\sigma-(1-\alpha)] < 0$, $-\sigma(1-\sigma)/[\sigma-(1-\alpha)] < 0$ and $[\sigma-(1-\alpha)]/(1-\sigma) > 0$, so that at $\varepsilon_i = 0$, $\varepsilon_i^{-(1-\alpha)(1-\sigma)/[\sigma-(1-\alpha)]}$ and $\varepsilon_i^{-\sigma(1-\sigma)/[\sigma-(1-\alpha)]}$ converge to plus infinity, meaning that the expression between square bracket equals plus infinity; as this is taken to a positive power, the result follows.

Let it now be the case that $s_i = S$. For any finite S, as a decrease in c both shifts down the contributor's costs and tilts them clockwise (i.e., decreases both fixed and variable costs), one can always find a sufficiently small but positive c such that benefit as a function of ε_i has parts that exceed the costs. Moreover, given that marginal benefit of contributor i is infinite at $\varepsilon_i = 0$, for any finite c, one can always find sufficiently small but positive S such that benefit as a function of ε_i has parts that exceed the costs.

Consider next $\sigma = 0$. In this case the marginal benefit (A4) equals $(\alpha/n^2)[\sum_{\ell=1}^{n}(1/n)\varepsilon_\ell]^{\alpha-1}$. Let all contributors $j \neq i$ exert net effort $\varepsilon_j = [\alpha/(cn^2)]^{1/(1-\alpha)}$, such that an equilibrium where contributor i exerts the same net effort becomes possible. Then at $\varepsilon_i = 0$, the marginal benefit equals $(\alpha/n^2)\{[(n-1)/n][\alpha/(cn^2)]^{1/(1-\alpha)}\}^{\alpha-1} = [n/(n-1)]^{1-\alpha}c$, which is larger than c. Let it again be the case that $s_i = S$. Then it follows that, for any c, one can find sufficiently small S such that contributor i prefers to set $\varepsilon_i = [\alpha/(cn^2)]^{1/(1-\alpha)}$ rather than $\varepsilon_i = 0$. Moreover, for any finite S, one can find sufficiently small c such that the same continues to be true, as follows from the fact that c reduces both contributor i's fixed cost cS and her variable cost $c\varepsilon_i$.

The fact that net effort is invariant to S follows from the fact that, with an increase in S, (A4) can only continue to equal the constant marginal cost if ε_i does not change. □

A.4. *Proof of Lemma 3*

Let all contributors other than i exert effort e, with $e \geq S$. The necessary conditions for the existence of a symmetric equilibrium where all contributors exert effort e are that, for the individual contributor i,

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(e_i-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)e^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}(e_i-S)^{-\sigma}-c \geq 0 \quad \text{for } S \leq e_i < e \quad \text{(A6)}$$

and

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)e_i^{1-\sigma}+\left(\frac{1}{n}\right)(e-S)^{1-\sigma}+\left(\frac{n-2}{n}\right)e^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}e_i^{-\sigma}-c \leq 0 \quad \text{for } e_i \geq e. \quad \text{(A7)}$$

These two conditions are compatible for e_i approaching e, as the first term on the LHS of (A6) is then larger than the first term on the LHS of (A7). Since, by the proof of Lemma A2, the first term on the LHS of (A6) approaches plus infinity for e_i approaching S, under the conditions of Lemma A2, the individual contributor prefers exerting effort e rather than 0.

Consider first finite σ. Then the first-order conditions are given by (A6) and (A7). The second-order conditions for (A6) and (A7) are valid and take the form

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(e_i - S)^{1-\sigma} + \left(\frac{n-1}{n}\right)e^{1-\sigma}\right]^{\alpha/(1-\sigma)-2}(e_i - S)^{-\sigma-1}$$
$$\times\left\{-\frac{1-\alpha}{n}(e_i - S)^{1-\sigma} - \sigma\left(\frac{n-1}{n}\right)e^{1-\sigma}\right\}$$
$$< 0, \tag{A8}$$

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)e_i^{1-\sigma} + \left(\frac{1}{n}\right)(e - S)^{1-\sigma} + \left(\frac{n-2}{n}\right)e^{1-\sigma}\right]^{\alpha/(1-\sigma)-2}{e_i}^{-\sigma-1}$$
$$\times\left\{-\frac{1-\alpha}{n}{e_i}^{1-\sigma} - \sigma\left[\left(\frac{1}{n}\right)(e - S)^{1-\sigma} + \left(\frac{n-2}{n}\right)e^{1-\sigma}\right]\right\}$$
$$< 0. \tag{A9}$$

For $\sigma \geq 1$, (A6)–(A9) suffice for the existence of the symmetric equilibrium. A distinguishing feature of technologies with $\sigma \geq 1$ is that each contributor's effort is essential (Irmen and Maussner, 2017), in that no public good can be produced if net effort is zero for at least one contributor. Given this fact, for $e_i < S$, the first derivate for contributor i equals $-c$, and under the conditions of Lemma A2, contributor i will prefer to exert $e > S$ rather than any effort lower than S. Yet, for $\sigma < 1$, production is still non-zero when $e_i < S$, and given that all other contributors set $e > S$, we cannot exclude a local maximum for contributor i where

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(e - (S - e_i))^{1-\sigma} + \left(\frac{n-2}{n}\right)e^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}(e - (S - e_i))^{-\sigma} - c$$
$$= 0 \quad \text{for } 0 \leq e_i < S. \tag{A10}$$

The second-order condition for (A10) is valid, and takes the form

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(e - (S - e_i))^{1-\sigma} + \left(\frac{n-2}{n}\right)e^{1-\sigma}\right]^{\alpha/(1-\sigma)-2}(e - (S - e_i))^{-\sigma-1}$$
$$\times\left\{-\frac{1-\alpha}{n}(e - (S - e_i))^{1-\sigma} - \sigma\left(\frac{n-2}{n}\right)e^{1-\sigma}\right\}$$
$$< 0.$$

To see the possibility of such a local maximum, note that the LHS of (A6) as e_i approaches S is positive; also, the LHS of (A7) as e_i approaches e and the LHS of (A10) as e_i approaches S are the same, and therefore both negative. In such a local maximum, denoting the production level achieved as $f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$, contributor i obtains payoff $(1/n)f(0, e - (S - e_i), e, \ldots, e)^\alpha - ce_i$; whereas in the local maximum proposed in the proposition, she obtains $(1/n)f(e - S, e, e, \ldots, e)^\alpha - ce$. For sufficiently small S and/or c, the latter is larger, and for $\sigma < 1$, we may therefore equally well focus on first-order conditions (A6) and (A7).

By (A6) and (A7), a subgame perfect equilibrium is therefore obtained for any e such that

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(e - S)^{1-\sigma} + \left(\frac{n-1}{n}\right)e^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}(e - S)^{-\sigma} - c \geq 0 \tag{A11}$$

and

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(e-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)e^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}e^{-\sigma}-c\leq 0. \tag{A12}$$

We next look for the highest effort $\bar{e}$ that can be obtained in equilibrium. The derivative of the LHS of (A11) with respect to e can be written as

$$\begin{aligned}&\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(e-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)e^{1-\sigma}\right]^{\alpha/(1-\sigma)-2}(e-S)^{-\sigma-1}\\&\times\left\{-(1-\alpha)\left(\frac{1}{n}\right)(e-S)^{1-\sigma}-(1-\alpha)\left(\frac{n-1}{n}\right)e^{-\sigma}(e-S)-S\sigma\left(\frac{n-1}{n}\right)e^{-\sigma}\right\}\\&<0.\end{aligned} \tag{A13}$$

Moreover, the LHS of (A11) can be written as

$$\begin{aligned}\frac{\alpha}{n^2}\Bigg[&\left(\frac{1}{n}\right)(e-S)^{-(1-\sigma)(1-\alpha)/[\alpha-(1-\sigma)]}\\&+\left(\frac{n-1}{n}\right)e^{1-\sigma}(e-S)^{-\sigma(1-\sigma)/[\alpha-(1-\sigma)]}\Bigg]^{[\alpha-(1-\sigma)]/(1-\sigma)}-c.\end{aligned} \tag{A14}$$

As $e\to\infty$, $e^{1-\sigma}(e-S)^{-\sigma(1-\sigma)/[\alpha-(1-\sigma)]}\to e^{-(1-\sigma)(1-\alpha)/[\alpha-(1-\sigma)]}$. For $\sigma<(1-\alpha)$ and $\sigma>1$, $-(1-\sigma)(1-\alpha)/[\alpha-(1-\sigma)]>0$ and $[\alpha-(1-\sigma)]/(1-\sigma)<0$, and (A14) converges to $-c$ as $e\to\infty$. For $(1-\alpha)<\sigma<1$, $-(1-\sigma)(1-\alpha)/[\alpha-(1-\sigma)]<0$ and $[\alpha-(1-\sigma)]/(1-\sigma)>0$, (A14) also converges to $-c$ as $e\to\infty$. If follows that $\bar{e}$ is finite. Given this fact, as the LHS of (A11) is larger than the LHS of (A12), and by (A13), it follows that $e=\bar{e}$ is found when the LHS of (A11) is zero, in which case the LHS of (A12) is negative, so that $\bar{e}$ is implicitly given by

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(\bar{e}-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)\bar{e}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}(\bar{e}-S)^{-\sigma}-c=0. \tag{A15}$$

Given (A8), the effect of S holding $\bar{e}$ fixed on the LHS of (A15) is positive; starting from $S=0$ and $\bar{e}=e_B$, increasing S therefore means that $\bar{e}>e_B$. Furthermore, the effort $\bar{e}$ defined by (A15) maximises contributor welfare among the subgame perfect equilibria, as the (non-equilibrium) effort that maximises contributor welfare is larger than $\bar{e}$ by the $1/n$ problem, so that contributors are best off in the subgame perfect equilibrium with the largest effort. Indeed, contributor welfare is maximised for an effort e^* such that

$$\frac{\alpha}{n}\left[\left(\frac{1}{n}\right)(e^*-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)e^{*1-\sigma}\right]^{\alpha/(1-\sigma)-1}(e^*-S)^{-\sigma}-c=0,$$

as follows from the fact that the first-order conditions for contributor welfare maximisation have a similar form as those for individual contributor payoff maximisation, except for the coefficient $(1/n)$.

A.5. *Proof of Proposition 2*

For $\sigma = 0$, (A11) and (A12) can only both be valid if

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(e-S)+\left(\frac{n-1}{n}\right)e\right]^{\alpha-1} - c = 0.$$

In a symmetric subgame perfect equilibrium, this is only possible for $e = e_B + S/n$. As this is the unique symmetric equilibrium, it is also the case that $\bar{e} = e_B + S/n$. This completes the proof of (*i*).

It is easily checked that, for $\sigma = (1-\alpha)$, (A15) is valid when $\bar{e} = e_B + S$ if and only if $\bar{e} - S = e_B$. For $\sigma \neq (1-\alpha)$, looking at the same value for $\bar{e}$ on the LHS of (A11), we obtain

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right){e_B}^{1-\sigma}+\left(\frac{n-1}{n}\right)(e_B+S)^{1-\sigma}\right]^{\alpha/(1-\sigma)-1} {e_B}^{-\sigma} - c. \tag{A16}$$

At the same time, we know from Proposition 1 that

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right){e_B}^{1-\sigma}+\left(\frac{n-1}{n}\right){e_B}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1} {e_B}^{-\sigma} - c = 0.$$

The effect of S on the LHS of (A16) is negative when $\sigma < (1-\alpha)$, so that $\bar{e}$ needs to be decreased in this case to make (A15) valid, or $\bar{e} < e_B + S$. When $\sigma > (1-\alpha)$, the effect of S on the LHS of (A16) is positive, so that $\bar{e}$ needs to be increased, or $\bar{e} > e_B + S$. This completes the proof of (*ii*).

Finally, for $\sigma = +\infty$, let other players exert effort e. Then the marginal payoff for contributor i is

$$\frac{\alpha}{n}(e_i - S)^{\alpha-1} - c \quad \text{for } e_i < e$$

and

$$-c \quad \text{for } e_i > e.$$

The individual contributor i therefore never exerts an effort higher than e. Also, $(\alpha/n)(e_i - S)^{\alpha-1}$ equals plus infinity for $e_i = S$, and diminishes. As long as $(\alpha/n)(e_i - S)^{\alpha-1} - c \geq 0$ for $e_i < e$, the individual contributor wants to set her effort level as close as possible to e. It follows that in the symmetric equilibrium with the maximal effort, $(\alpha/n)(\bar{e} - S)^{\alpha-1} - c = 0$, or $\bar{e} = [\alpha/(cn)]^{1/(1-\alpha)} + S = e_B^*$.

To complete the proof of (*iii*), we note that there is no discontinuity in $\bar{e}$ as $\sigma \to +\infty$. To see this, note that in (A14), for $e = bare$, as σ becomes large, $-\sigma(1-\sigma)/[\alpha-(1-\sigma)]$ approaches σ, so that $\bar{e}^{1-\sigma}[(\bar{e}-S)^{-\sigma(1-\sigma)/[\alpha-(1-\sigma)]}$ approaches $[(\bar{e}-S)/\bar{e}]^{\sigma}$, meaning that the second term within the square brackets approaches zero. The first term approaches $(1/n)(\bar{e}-S)^{(1-\alpha)}$, and the power $[\alpha-(1-\sigma)]/(1-\sigma)$ approaches minus one, so that the entire expression approaches $(\alpha/n)(\bar{e}-S)^{\alpha-1} - c$. From this, it follows that $\bar{e} \to [\alpha/(cn)]^{1/(1-\alpha)} + S$.

A.6. *Proof of Proposition 3*

The derivative of (A15) with respect to S equals

$$\frac{\alpha}{n^2}\left[\left(\frac{1}{n}\right)(\bar{e}-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)\bar{e}^{1-\sigma}\right]^{\alpha/(1-\sigma)-2}(\bar{e}-S)^{-\sigma-1}$$
$$\times\left\{(1-\alpha)\left(\frac{1}{n}\right)(\bar{e}-S)^{1-\sigma}+\sigma\left(\frac{n-1}{n}\right)\bar{e}^{1-\sigma}\right\}$$
$$>0. \tag{A17}$$

Using (A13) and (A17), and applying implicit differentiation to (A15), it follows that

$$\frac{d\bar{e}}{dS}=\frac{(1-\alpha)(\bar{e}-S)^{1-\sigma}+\sigma(n-1)\bar{e}^{1-\sigma}}{(1-\alpha)(\bar{e}-S)^{1-\sigma}+(1-\alpha)(n-1)\bar{e}^{1-\sigma}+S[\sigma-(1-\alpha)](n-1)\bar{e}^{-\sigma}}.$$

The marginal effect of an increase in S on the sum of the contributors' payoffs equals

$$\left\{\alpha\left[\left(\frac{1}{n}\right)(\bar{e}-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)\bar{e}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}\right.$$
$$\left.\times\left[\left(\frac{1}{n}\right)(\bar{e}-S)^{-\sigma}+\left(\frac{n-1}{n}\right)\bar{e}^{-\sigma}\right]-cn\right\}\frac{d\bar{e}}{dS}$$
$$-\frac{\alpha}{n}\left[\left(\frac{1}{n}\right)(\bar{e}-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)\bar{e}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}(\bar{e}-S)^{-\sigma}. \tag{A18}$$

Applying (A15) and solving for cn, we obtain

$$cn=\frac{\alpha}{n}\left[\left(\frac{1}{n}\right)(\bar{e}-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)\bar{e}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}(\bar{e}-S)^{-\sigma}.$$

Substituting into (A18), it follows that (A18) equals

$$\left\{\frac{\alpha}{n}\left[\left(\frac{1}{n}\right)(\bar{e}-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)\bar{e}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}(n-1)\bar{e}^{-\sigma}\right\}\frac{d\bar{e}}{dS}$$
$$-\frac{\alpha}{n}\left[\left(\frac{1}{n}\right)(\bar{e}-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)\bar{e}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}(\bar{e}-S)^{-\sigma}$$
$$=\frac{\alpha}{n}\left[\left(\frac{1}{n}\right)(\bar{e}-S)^{1-\sigma}+\left(\frac{n-1}{n}\right)\bar{e}^{1-\sigma}\right]^{\alpha/(1-\sigma)-1}\left\{(n-1)\bar{e}^{-\sigma}\frac{d\bar{e}}{dS}-(\bar{e}-S)^{-\sigma}\right\}.$$

Around $S=0$, the latter expression becomes

$$\frac{\alpha}{n}\bar{e}^{\alpha-1}\left[(n-1)\frac{(1-\alpha)+\sigma(n-1)}{n(1-\alpha)}-1\right],$$

which, for finite n, is larger than zero if $\sigma>(1-\alpha)(n-1)^{-2}$.

A.7. *Proof of Proposition 4*

(*i*) As shown in Proposition 2(*i*), with $\sigma=0$, it is the case that $\bar{e}=e_B+S/n$, meaning that $Q=[(1/n)(e_B+S/n-S)+((n-1)/n)(e_B+S/n)]=e_B$. Production is therefore exactly the same as in the absence of an adversary. As the presence of an adversary raises effort levels and therefore costs, but does not change production, $S_{\max} = 0$ maximises contributor welfare.

(*ii*) For $\sigma = (1-\alpha)$, by Proposition 2(*ii*), it is the case that $\bar{e} = e_B + S$. We first find the maximum sabotage budget S^* such that the individual contributor does not prefer to exert zero effort rather than $\bar{e}$ when all other contributors exert effort $\bar{e}$. This is the case if

$$\frac{n-1}{n^2}\left[\left(\frac{\alpha}{cn^2}\right)^{1/(1-\alpha)} + S\right]^{\alpha} + \frac{1}{n^2}\left(\frac{\alpha}{cn^2}\right)^{\alpha/(1-\alpha)} - c\left[\left(\frac{\alpha}{cn^2}\right)^{1/(1-\alpha)} + S\right]$$
$$\geq \frac{n-2}{n^2}\left[\left(\frac{\alpha}{cn^2}\right)^{1/(1-\alpha)} + S\right]^{\alpha} + \frac{1}{n^2}\left(\frac{\alpha}{cn^2}\right)^{\alpha/(1-\alpha)}$$
$$\iff \quad S \leq S^* = \left(\frac{\alpha}{cn^2}\right)^{1/(1-\alpha)}[1-\alpha^{1/(1-\alpha)}].$$

Denoting by S^{**} the sabotage budget that maximises (A1), by the proof of Lemma 1, for $n = 2$, $S^{**} = 0$ and (A1) decreases everywhere in S. For $n \geq 3$, $S^{**} = (\alpha/(cn^2))^{1/(1-\alpha)}[(n-1)^{1/(1-\alpha)} - 1]$ and (A1) increases in S for $S < S^{**}$.

Denote now by $S_{\max}$ the sabotage budget that maximises contributor welfare for contributors who coordinate on effort $\bar{e}$. Given that S^{**} is only larger than zero for $n \geq 3$, it follows that, for $n = 2$, it is the case that $S_{\max} = 0$. For $n \geq 3$, one can distinguish between two cases: either $S^{**} \leq S^*$, in which case $S_{\max} = (\alpha/(cn^2))^{1/(1-\alpha)}[(n-1)^{1/(1-\alpha)} - 1] > 0$—this is the case when $0 < \alpha \leq (1/(n-1))$—or $S^{**} > S^*$, in which case $S_{\max} = (\alpha/(cn^2))^{1/(1-\alpha)}[1-\alpha^{1/(1-\alpha)}] > 0$—this is the case when $(1/(n-1)) < \alpha < 1$.

(*iii*) This follows directly from the fact that approaching perfect complements (see Proposition 2(*iii*)), the maximal equilibrium effort approaches the socially optimal effort without an adversary, plus S. For this reason, a vanishingly small but positive S maximises the contributors' payoffs.

Utrecht University, The Netherlands

Additional Supporting Information may be found in the online version of this article:

Appendix B. Robustness analysis

References

Adams, C.P. (2006). 'Optimal team incentives with CES production', *Economics Letters*, vol. 92(1), pp. 143–8.
Alchian, A.A. and Demsetz, H. (1972). 'Production, information costs, and economic organization', *American Economic Review*, vol. 62(5), pp. 777–95.
Arrow, K.J., Chenery, H.B., Minhas, B.S. and Solow, R.M. (1961). 'Capital-labor substitution and economic efficiency', *Review of Economics and Statistics*, vol. 43(3), pp. 225–50.
Baer, M., Leenders, R.T.A.J., Oldham, G.R. and Vadera, A.K. (2010). 'Win or lose the battle for creativity: The power and perils of intergroup competition', *Academy of Management Journal*, vol. 53(4), pp. 827–45.
Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E. and Mitts, T. (2016). 'Can war foster cooperation?', *Journal of Economic Perspectives*, vol. 30(3), pp. 249–74.
Bell, T.E. (1991). 'When bust is best (destructive testing)', *IEEE Spectrum*, vol. 28(3), pp. 56–9.
Belloc, F. (2012). 'Corporate governance and innovation: A survey', *Journal of Economic Surveys*, vol. 26(5), pp. 835–64.
Bergstrom, T.C., Blume, L. and Varian, H.R. (1986). 'On the private provision of public goods', *Journal of Public Economics*, vol. 29(1), pp. 25–49.
Bierbrauer, F.J. and Hellwig, M.F. (2016). 'Robustly coalition-proof incentive mechanisms for public good provision are voting mechanisms and vice versa', *Review of Economic Studies*, vol. 83(4), pp. 1440–64.
Blair, M.M. and Stout, L.A. (1999). 'A team production theory of corporate law', *Virginia Law Review*, vol. 85(2), pp. 247–328.

Bullen, P.S. (2003). *Handbook of Means and Their Inequalities*, Dordrecht: Kluwer Academic Publishers.

Calo-Blanco, A., Kovářík, J., Mengel, F. and Romero, J.G. (2017). 'Natural disasters and indicators of social cohesion', *PloS One*, vol. 12(6), e0176885.

Charness, G., Masclet, D. and Villeval, M.C. (2014). 'The dark side of competition for status', *Management Science*, vol. 60(1), pp. 38–55.

Chowdhury, S.M. and Gürtler, O. (2015). 'Sabotage in contests – a survey', *Public Choice*, vol. 164(1), pp. 135–55.

Cornes, R.C. (1993). 'Dyke maintenance and other stories: Some neglected types of public goods', *Quarterly Journal of Economics*, vol. 108(1), pp. 259–71.

Dawes, R.M., Orbell, J.M., Simmons, R.T. and Van de Kragt, A.J.C. (1986). 'Organizing groups for collective action', *American Political Science Review*, vol. 80(4), pp. 1171–85.

De Jaegher, K. (2021). 'Common-enemy effects: Multidisciplinary antecedents and economic perspectives', *Journal of Economic Surveys*, vol. 35(1), pp. 3–33.

De Jaegher, K. and Hoyer, B. (2016a), 'Collective action and the common-enemy effect', *Defence and Peace Economics*, vol. 27(5), pp. 644–64.

De Jaegher, K. and Hoyer, B. (2016b), 'By-product mutualism and the ambiguous effects of harsher environments: A game-theoretic model', *Journal of Theoretical Biology*, vol. 393, pp. 82–97.

De Jaegher, K. and Hoyer, B. (2019). 'Preemptive repression: Deterrence, backfiring, iron fists and velvet gloves', *Journal of Conflict Resolution*, vol. 63(2), pp. 502–27.

Devetag, G. and Ortmann, A. (2007). '"When and why?" A critical survey of coordination failure in the laboratory', *Experimental Economics*, vol. 10(3), pp. 171–8.

Diekmann, A. (1985). 'Volunteer's dilemma', *Journal of Conflict Resolution*, vol. 29(4), pp. 605–10.

Doğan, S., Keskin, K. and Sağlam, C. (2019). 'Sabotage in team contests', *Public Choice*, vol. 180(3), pp. 383–405.

Fjermestad, J. (2019). 'Structured conflict approaches used in strategic decision making: From Mason's initial study to virtual teams', *Foundations and Trends in Information Systems*, vol. 3(3–4), pp. 234–400.

Flavel, R. and Williams, J. (1996). *Strategic Management: A Practical Approach,* Sydney: Prentice-Hall.

Gersen, J.E. and Vermeule, A. (2012). 'Delegating to enemies', *Columbia Law Review*, vol. 112(8), pp. 2193–238.

Hardy, G., Littlewood, J.E. and Pólya, G. (1934). *Inequalities*, Cambridge: Cambridge University Press.

Harrison, G.W. and Hirshleifer, J. (1989). 'An experimental evaluation of weakest link/best shot models of public goods', *Journal of Political Economy*, vol. 97(1), pp. 201–25.

Harsanyi, J.C. and Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*, Cambridge MA: MIT Press.

Heckathorn, D.W. (1993). 'Collective action and group heterogeneity: Voluntary provision versus selective incentives', *American Sociological Review*, vol. 58(3), pp. 329–50.

Heckathorn, D.W. (1996). 'The dynamics and dilemmas of collective action', *American Sociological Review*, vol. 61(2), pp. 250–77.

Heifetz, A., Meier, M. and Schipper, B.C. (2021). 'Prudent rationalizability in generalized extensive-form games with unawareness', *BE Journal of Theoretical Economics*, vol. 21(2), pp. 525–56.

Hirshleifer, J. (1983). 'From weakest-link to best-shot: The voluntary provision of public goods', *Public Choice*, vol. 41(3), pp. 371–86.

Holmström, B. (1982). 'Moral hazard in teams', *Bell Journal of Economics*, vol. 13(2), pp. 324–40.

Hoyer, B. and De Jaegher, K. (2012). 'Network disruption and the common-enemy effect', TKI Discussion Paper Series 12-06, Utrecht University.

Hoyer, B. and Haller, H. (2019). 'The common enemy effect under strategic network formation and disruption', *Journal of Economic Behavior and Organization*, vol. 162, pp. 146–63.

Hugh-Jones, D. and Zultan, R. (2012). 'Reputation and cooperation in defense', *Journal of Conflict Resolution*, vol. 57(2), pp. 327–55.

Irmen, A. and Maussner, A. (2017). 'A note on the characterization of the neoclassical production function', *Macroeconomic Dynamics*, vol. 21(7), pp. 1827–35.

Ivancevich, J.M., Konopaske, R. and Matteson, M.T. (2008). *Organizational Behavior and Management*, 8th edn., New York: McGraw-Hill.

Jones, G.R. (2012). *Organizational Theory, Design, and Change*, 7th edn., Upper Saddle River, NJ: Prentice Hall.

Katzenstein, G. (1996). 'The debate on structured debate: Toward a unified theory', *Organizational Behavior and Human Decision Processes*, vol. 66(3), pp. 316–32.

Kempa, K. and Rusch, H. (2019). 'Dissent, sabotage, and leader behaviour in contests: Evidence from European football', *Managerial and Decision Economics*, vol. 40(5), pp. 500–14.

Kovenock, D. and Roberson, B. (2012). 'Coalitional Colonel Blotto games with application to the economics of alliances', *Journal of Public Economic Theory*, vol. 14(4), pp. 653–76.

Marino, A.M. and Zábojnik, J. (2004). 'Internal competition for corporate resources and incentives in teams', *RAND Journal of Economics*, vol. 35(4), pp. 710–27.

Marks, M. and Croson, R. (1998). 'Alternative rebate rules in the provision of a threshold public good: An experimental investigation', *Journal of Public Economics*, vol. 67(2), pp. 195–220.

Mason, R.O. (1969). 'A dialectical approach to strategic planning', *Management Science*, vol. 15(8), pp. B403–14.

McAfee, R.P. and McMillan, J. (1991). 'Optimal contracts for teams', *International Economic Review*, vol. 32(3), pp. 561–77.

Mesterton-Gibbons, M. and Dugatkin, L.A. (1992). 'Cooperation among unrelated individuals: Evolutionary factors', *Quarterly Review of Biology*, vol. 67(3), pp. 267–81.
Milgrom, P.R. (1981). 'Good news and bad news: Representation theorems and applications', *Bell Journal of Economics*, vol. 12(2), pp. 380–91.
Monderer, D. and Shapley, L.S. (1996). 'Potential games', *Games and Economic Behavior*, vol. 14(1), pp. 124–43.
Münster, J. and Staal, K. (2011). 'War with outsiders makes peace inside', *Conflict Management and Peace Science*, vol. 28(2), pp. 91–110.
Myatt, D.P. and Wallace, C. (2009). 'Evolution, teamwork and collective action: Production targets in the private provision of public goods', Economic Journal, vol. 119(534), pp. 61–90.
Nemeth, C., Brown, K. and Rogers, J. (2001). 'Devil's advocate versus authentic dissent: Stimulating quantity and quality', *European Journal of Social Psychology*, vol. 31(6), pp. 707–20.
Nöldeke, G and Peña, J. (2020). 'Group size and collective action in a binary contribution game', *Journal of Mathematical Economics*, vol. 88, pp. 42–51.
Olson, M. (1965). *The Logic of Collective Action – Public Goods and the Theory of Groups*, Cambridge MA: Harvard University Press.
Peters, T.J. and Waterman, R.H., Jr. (1982). *In Search of Excellence – Lessons from America's Best-Run Companies*, New York: Harper & Row.
Rahmani, M., Roels, G. and Karmarkar, U.S. (2018). 'Team leadership and performance: Combining the roles of direction and contribution', *Management Science*, vol. 64(11), pp. 5234–49.
Ray, D, Baland, J.-M. and Dagnelie, O. (2007). 'Inequality and inefficiency in joint projects', Economic Journal, vol. 117(522), pp. 922–35.
Salop, S.C. and Scheffman, D.T. (1983). 'Raising rivals' costs', *American Economic Review*, vol. 73(2), pp. 267–71.
Schipper, B.C. (2014). 'Unawareness – a gentle introduction to both the literature and the special issue', *Mathematical Social Sciences*, vol. 70, pp. 1–9.
Schweiger, D.M., Sandberg, W.R. and Ragan, J.W. (1986). 'Group approaches for improving strategic decision making: A comparative analysis of dialectical inquiry, devil's advocacy, and consensus', *Academy of Management Journal*, vol. 29(1), pp. 51–71.
Schwenk, C.R. (1984). 'Devil's advocacy in managerial decision-making', *Journal of Management Studies*, vol. 21(2), pp. 153–68.
Selten, R. (1975). 'Reexamination of the perfectness concept for equilibrium points in extensive games', *International Journal of Game Theory*, vol. 4(1), pp. 25–55.
Sutton, J., Butcher, C.R. and Svensson, I. (2014). 'Explaining political jiu-jitsu: Institution-building and the outcomes of regime violence against unarmed protests', *Journal of Peace Research*, vol. 51(5), pp. 559–73.
Theelen, M.P. and Böhm, R. (2021). 'The conflict-cooperation effect persists under intragroup payoff asymmetry', *Group Processes and Intergroup Relations*, vol. 24(5), pp. 815–35.
Tullock, G. (1980). 'Efficient rent seeking', in (J.M. Buchanan, R.D. Tollison and G. Tullock, eds.), *Toward a Theory of the Rent-Seeking Society*, pp. 3–16, Boston MA: Springer.
Van Knippenberg, D. (2003). 'Intergroup relations in organizations', in (M.A. West, D. Tjosvold and K.G. Smith, eds.), *International Handbook of Organizational Teamwork and Cooperative Working*, pp. 381–99, Chichester: Wiley.
Verbeke, W., Bagozzi, R.P. and Belschak, F.D. (2016). 'The role of status and leadership in sales contests: A natural field experiment', *Journal of Business Research*, vol. 69(1), pp. 4112–20.
Vicary, S. (1990). 'Transfers and the weakest-link – an extension of Hirshleifer's analysis', *Journal of Public Economics*, vol. 43(3), pp. 375–94.
Wang, X., Chan, H.K. and Li, D. (2015). 'A case study of an integrated fuzzy methodology for green product development', *European Journal of Operational Research*, vol. 241(1), pp. 212–23.