# Appendix

## A. A Brief Review of the Economics Literature on Strategic Behavior

The threat of sabotage in which actors engage in non-productive effort to reduce the performance of another actor has been at the forefront of economic literature and noted in the earliest work on rank-order tournaments Lazear (1989), Dixit (1987). However, past empirical work on sabotage is scarce. Notable exceptions are the work on withholding helping behavior (Drago and Garvey 1998, Haas and Park 2010), which can be considered a weaker form of sabotage. The dearth of empirical insight into sabotage is (at least in part) attributable to the challenge of investigating sabotage empirically: actors try to hide their opportunistic behavior due to its immoral connotation so that "company data on sabotage is generally not available for research" (Dato and Nieken 2014, p. 65).

Self-promotion is somewhat better understood empirically (Li 2017, Keum and See 2017, Reitzig and Sorenson 2013, Edelman and Larkin 2015, Lerchenmueller et al. 2019), but it is unclear how it interacts with sabotage and when individuals would choose one form of strategic behavior over the other (i.e., whether the two are complements or substitutes). Thus, there are limited empirical insights into strategic behavior beyond theoretical work (Lazear 1989, Milgrom and Roberts 1988, Münster 2007), laboratory studies (Harbring and Irlenbusch 2011, Charness et al. 2014), or sports (Balafoutas et al. 2012, Deutscher et al. 2013, Garicano et al. 2005).[15]

Prior work from the economics literature has suggested that strategic behavior is contingent on the skill of contestants (e.g., Schotter and Weigelt 1992, Dixit 1987). However, existing empirical evidence is ambiguous. Some studies find that low skill is associated with engagement in strategic behavior (Carpenter et al. 2010, Charness et al. 2014) while others find evidence for higher skill (Harbring and Irlenbusch 2011). Yet others find no association between skill and strategic behavior at all (Dato and Nieken 2014). Hence, the effect (or lack of it) remains unclear.

## B. Theoretical Framework

### B.1. Model Setup in More Formalized Terms

We build on previous models of one-shot contest with sabotage (see, e.g., Harbring and Irlenbusch 2011, Konrad 2000), skill heterogeneous agents, and a single winner prize. The contest consists of two types of agents: a set $H$ of $h$ high-skill agents, and a set $L$ of $l$ low-skill agents. We refer to these as high and low types, respectively. Furthermore, the contest involves a set $N$ of $n$ outsiders. We assume that $h < l < n$, which should be satisfied in most realistic contest settings. Low and high types produce a contest submission (i.e. idea) of low quality $b_l$ and high quality $b_h$, respectively, which they enter into the contest. Outsiders do not enter the contest and thus remain neutral. The submissions of all agents are rated by each outsider and by each high and low type agent on a bounded quality scale in $\mathbb{R}_{\geq 0}$ with $[r_{min}, r_{max}]$. Without loss of generality, we normalize the rating scale to the unit interval so that $r_{min} = 0$ and $r_{max} = 1$. We assume that the quality bid of low type agents' submissions is $b_l$, whereas the quality bid of high type agents' submissions is $b_h$ with $0 < b_l < b_h < 1$.

While ratings can be sincere (i.e., $b_l$ or $b_h$, respectively), an agent can also sabotage any other agent, rating their submission at $r_{min} = 0$; or promote them by rating their submission at $r_{max} = 1$. Let $\Delta s_h = b_h$ be the damage done by sabotaging a high type and $\Delta s_l = b_l$ be the damage done by sabotaging a low type. Since $b_l < b_h$, it follows that the damage inflicted on a high type through sabotage is larger than on a low type (i.e., $\Delta s_l < \Delta s_h$). Further, let $\Delta sp_h = 1 - b_h$ denote the benefit gained from promoting a high type and $\Delta sp_l = 1 - b_l$ denote the benefit gained from promoting a low type. Since $b_l < b_h$ it follows that low types have more to gain from promotion than high types (i.e., $\Delta sp_l > \Delta sp_h$).

We assume that while rating sincerely is free of costs, sabotaging another agent costs $c_s$ and promoting any agent (including oneself) costs $c_p$. Costs associated with promotion and sabotage might arise from the costs of identifying suitable targets (Harbring et al. 2007, Münster 2007), the moral costs associated with lying (Abeler et al. 2019, Gneezy et al. 2018) or violating social norms (Elster 1989). As a consequence all $N$ outsiders rate contestants $h$ and $l$ sincerely. This implies that the size $n$ of the set of outsiders $N$ affects the effectiveness of any type of strategic behavior performed by the set of $H$ and $L$ agents: the more outsiders there are, the lower the effect that any individual strategic rating has. Effectively, $n$ can be considered a modifier in our model that governs how effective strategic behavior is. Note also that outsiders $n$ are not necessarily agents who never compete—they simply do not compete in the current contest. We will use this feature to identify strategic behavior in our empirical analysis where contestants do not enter every contest: they rate as neutral outsiders in some weeks and rate as competitors with stakes from idea generation—and incentives to rate strategically—in the contest in other weeks.

The rating given by agent $i$ to agent $j$ is denoted by $v_{ij}$. The value of agent $j$'s submission is the sum of all the ratings that this focal agent receives: $v_j = \sum_{i \in N \cup L \cup H} v_{ij}$. Throughout the remainder, let $v_h$ denote the value of some arbitrary high type (she), and $v_l$ denote the value of some arbitrary low type (he). As an example, in the case of sincere rating by everyone (no promotion and no sabotage), the value of a high type is $v_h = b_h(n + l + h)$ and that of a low type is $v_l = b_l(n + l + h)$. Finally, we also assume that the difference between contest entries $b_l$ and $b_h$ is large enough so that the relative order between high and low types cannot be changed through strategic behavior alone. That is, $v_l < v_h$ even if all low types engage in self-promotion and all high-types are sabotaged by everyone.

### B.2. Utility Function

The winner of the contest is determined based on the evaluations from $N$, $H$, and $L$. For simplicity, we model this as a Tullock contest (Tullock 1980) where the probability of winning is proportional to the aggregate rating $v_i$ plus an error term in relation to all other contest submissions. That is, each agent $i$'s probability of winning the contest is $p_i = \frac{v_i}{\sum_{j \in L \cup H} v_j}$. For simplicity, let $S = \sum_{j \in L \cup H} v_j$ denote the total value of all contestants and we write the Tullock contest success function as $p_i = \frac{v_i}{S}$. The winner of the contest receives a prize $M$, which is without loss of generality normalized to 1. Given the equilibrium values for productive effort and the resulting quality bid in the ordinary Tullock contest, agents maximize their expected winning probability minus their total costs of sabotage and cost of promotion. The utility function for agents' strategic contest behavior can then be written as:

$$E\Pi_i = Mp_i - c_s(\sum_j sab_{ij}) - c_p sprom_{ii}), \tag{3}$$

where $M$ is the winner prize, $p_i$ is the probability of agent $i$ winning the prize, $sab_{ij}$ ($prom_{ij}$) is 1 if agent $i$ sabotages (promotes) agent $j$ and 0 otherwise, and $c_s$ and $c_p$ are the cost of sabotage and promotion, respectively.

Formally, a strategy of agent $i$ is a list $(r_{ij})$ of rating behavior towards all agents $j \in L \cup H$, where $r_{ij} \in \{\text{sincere}, \text{sabotage}, \text{promote}\}$. Since sabotaging oneself and promoting any agent other than oneself are clearly strictly dominated and payoffs are symmetric with respect to actions towards other agents of the same type, we only consider equivalence classes of strategies of the form $s_i = (hsab_i, lsab_i, sprom_i)$, where $hsab_i$ is the number of high type agents $i$ sabotages, $lsab_i$ the number of low type agents $i$ sabotages, and $sprom_i$ takes the value 1 or 0, indicating whether or not $i$ self-promotes. We now investigate the strategic behavior for self-promotion and sabotage.

### B.3. Self-Promotion

Self-promotion increases agent $i$'s value $v_i$ and total contest output $S$ by the same amount and hence increases agent $i$'s chance of winning. The size of the this increase of course depends on agent $i$'s type, the number of other high and low types in the contest, and the number of outsiders.

Lemma 1. *In contests with a sufficiently large performance gap $g$ between low and high types, the expected gain from self-promotion is higher for a low type than a high type.*

*Proof.* The expected gain from self-promotion for a currently non-self-promoting high type agent $h$ is $\frac{v_h+\Delta sp_h}{S+\Delta sp_h} - \frac{v_h}{S} = \frac{\Delta sp_h(S-v_h)}{S(S+\Delta sp_h)}$, and $\frac{v_l+\Delta sp_l}{S+\Delta sp_l} - \frac{v_l}{S} = \frac{\Delta sp_l(S-v_l)}{S(S+\Delta sp_l)}$ for a currently non-self-promoting low type agent $l$. Considering the worst case value for a high skill agent gives us $v_h \geq b_h(n+1)$ (i.e., sincere rating by all $n$ outsiders and $h$ herself, all $l$ and $h-1$ sabotage and rate 0). Conversely, the best case value for a low type gives us $v_l \leq b_l(n+l+h)$ (i.e., sincere rating by everyone).

In a contest where the performance gap between high and low types is large with respect to the proportion of neutral outsiders $n$ (who vote truthfully) and the overall contest size (where $l$ and $h$ might potentially sabotage) we get $S > v_h > v_l$. Define $g$ as the performance gap between high and low types so that $v_h = gv_l$. Expressing $g$ as the inequality between the worst case value for a high type (receiving sabotage from all other high types and all low types) and the best case value for a low type (not being sabotaged at all)

$$g = \frac{b_h}{b_l}\frac{n+1}{n+l+h}. \tag{4}$$

This expression gives a lower bound on $g$ for a contest with a positive performance gap. A positive performance gap ensures that the high type values $v_h$ are higher than the low type values $v_l$ even if the high types are being sabotaged by all $H_{-i}$ and $L$ agents. This ensures that the $b_l < b_h$ relationship also holds for the values $v_l < v_h$. Note that this expresses the necessary size of the performance gap as a product of the difference in production skill and the proportion of neutral outsiders to overall contest size. Considering the smallest possible contest will have $n=3, l=2, h=1$ provides a lower bound for $g$ so that

$$g = \frac{b_h}{b_l}\frac{n+1}{n+l+h} \geq \frac{b_h}{b_l}\frac{2}{3} \tag{5}$$

Now consider a contest with a significant performance gap between high and low types so that $g \geq 1$ holds, then we can compare the expected gain from self-promotion between high and low types.

$$\frac{\Delta sp_l(S-v_l)}{S(S+\Delta sp_l)} - \frac{\Delta sp_h(S-v_h)}{S(S+\Delta sp_h)} = \frac{S((\Delta sp_l - \Delta sp_h)S + v_h - \Delta sp_l v_l) + \Delta sp_l(v_h - v_l)}{S(S+\Delta sp_h)(S+\Delta sp_l)}. \tag{6}$$

This expression is $>0$ as can be seen by substituting $gv_l$ for $v_h$ and $1-b_l$ for $\Delta sp_l$:

$$\begin{aligned} S((\Delta sp_l - \Delta sp_h)S + v_h - \Delta sp_l v_l) &\geq \\ S[(\Delta sp_l - \Delta sp_h)S + gv_l - (1-b_l)v_l] &= \\ S[(\Delta sp_l - \Delta sp_h)S + (g-1+b_l)v_l] &> 0 \end{aligned} \tag{7}$$

Thus, in contests where the performance gap between low and high types is large enough such that $g \geq 1$ is satisfied, the expected gain from self-promotion for a low type is larger than for a high type. □

A high type will of course self-promote if the gain from self-promotion is higher than its costs. More specifically:

LEMMA 2. *If, for a high type with value $v_h$ and a performance gap $g \geq 1$, $c_p < \frac{\Delta sp_h(S-v_h)}{S(S-\Delta sp_h)}$ in equilibrium, then all agents will engage in self-promotion.*

*Proof.* If the high type did not engage in self-promotion, her winning probability would fall by $\frac{v_h}{S} - \frac{v_h - \Delta sp_h}{S - \Delta sp_h} = \frac{\Delta sp_h(S-v_h)}{S(S-\Delta sp_h)}$. By assumption, this is larger than $c_p$, so her net gain would be negative. She will therefore adhere to self-promotion. By Lemma 1, this continues to hold for the low types if $g \geq 1$. Alternatively, low types will engage in self-promotion if $c_p < \frac{\Delta sp_l(S-v_l)}{S(S-\Delta sp_l)}$. □

### B.4. Sabotage

If an agent sabotages another agent, this decreases that agent's output $v_i$ and thus decreases $S$, which increases that agents probability of winning the contest. This decrease in $S$ also benefits all other agents and thus sabotage has—contrary to self-promotion—an important negative externality (Konrad 2000). We first show that if an agent sabotages another agent, the agent sabotages all agents of that type.

LEMMA 3. *The expected marginal gain in the probability of winning increases with the number of other agents (of the same type) being sabotaged.*

*Proof.* Consider a high type of value $v_h$ and sincere rating by everyone. Her probability of winning is $\frac{v_h}{S}$. If she sabotages one more high type, her probability of winning increases to $\frac{v_h}{S-b_h}$. The marginal gain in probability from this act of sabotage is $\frac{v_h}{S-b_h} - \frac{v_h}{S} = \frac{b_h v_h}{S(S-b_h)}$. We can then express the marginal gain of sabotage as a function of the number of other high types $x_h$ that the agent sabotages: $f(x_h) = \frac{x_h b_h v_h}{S(S - x_h b_h)}$. The first derivative of which is $f'(x_h) = \frac{b_h v_h}{(S - x_h b_h)^2}$ which is always positive for positive $x_h$ and given that $S > nb_h > hb_h > x_h b_h$ holds since by design $n > h$ and $h > x_h$ because the agent does not sabotage herself and at most sabotages $h-1$ high types. Thus, the marginal gain of sabotaging another high type increases with the number of high types already sabotaged. If the costs of sabotage are smaller than the (maximum) marginal gain from sabotaging the last other high type agent, a rational high type agent will therefore sabotage all other high type agents. An analogous argument shows that the marginal gain for a high type from sabotaging one more low type agent increases with the number of low type agents already sabotaged. The same arguments also hold for low types sabotaging other agents. □

We can use Lemma 3 to compute the bounds of sabotage costs $c_s$ when agents engage in strategic behavior. For example, consider a high type of value $v_h$ not currently sabotaging anyone and sincere voting by everyone else so that $v_h = b_h(n + l + h)$ and $v_l = b_l(n + l + h)$, and $S = hv_h + lv_l$. If she sabotages all other high types (remember she will not sabotage herself) her probability of winning is $\frac{v_h}{S - b_h(h-1)}$. If she were to sabotage only $h-2$ high types her probability of winning is $\frac{v_h}{S - b_h(h-2)}$. Consequently, her maximal marginal increase in the probability of winning from sabotaging other high types results from sabotaging all other $h-1$ instead of $h-2$ high type agents and is of size

$$\frac{v_h}{S - b_h(h-1)} - \frac{v_h}{S - b_h(h-2)} = \frac{v_h b_h}{(S - b_h(h-1))(S - b_h(h-2))} \tag{8}$$

If the costs of sabotage $c_s$ are smaller than the maximal marginal gain in Eq. 8, a rational high type will therefore sabotage all other high types. In a similar fashion we can compute the bounds for all other acts of sabotage (high types sabotaging low types, low types sabotaging high types etc.). We show all bounds in the appendix. Inspecting the bounds, the following relationships between the bounds are apparent.

LEMMA 4. *Both high and low types have more to gain from self-promotion than from sabotaging any other types.*

*Proof.* This follows directly from the externality associated with sabotage. □

### B.5. High Types Start Sabotaging High Types

LEMMA 5. *High types have more to gain from sabotaging other agents of a given type (all low types or all high types), than low types have to gain from sabotaging those agents.*

*Proof.* Consider a high type of value $v_h$ not currently sabotaging anyone and sincere voting by everyone else so that $v_h = b_h(n + l + h)$ and $v_l = b_l(n + l + h)$, and $S = hv_h + lv_l$. High types will sabotage other high types if the cost of sabotage $c_s$ is lower than the maximum marginal increase of sabotaging the last unit (i.e., going from sabotaging $h-2$ to $h-1$).

$$
\begin{aligned}
& \frac{v_h}{S-b_h(h-1)} - \frac{v_h}{S-b_h(h-2)} \qquad (9)\\
\Leftrightarrow\quad & \frac{v_h(S-b_h(h-2)) - v_h(S-b_h(h-1))}{(S-b_h(h-1))(S-b_h(h-2))}\\
\Leftrightarrow\quad & \frac{v_hS - v_hb_h(h-2) - v_hS + v_hb_h(h-1)}{(S-b_h(h-1))(S-b_h(h-2))}\\
\Leftrightarrow\quad & \frac{v_hb_h}{(S-b_h(h-1))(S-b_h(h-2))}
\end{aligned}
$$

□

### B.6. High Types Start Sabotaging Low Types

There are two possible transition points into this equilibrium: High types may start sabotaging low types while high types already sabotage other high types, but no sabotage otherwise. Or high types may start sabotaging low types, after low types have already started to sabotage high types. That is, the exact bound for high types to sabotage low types, will depend on whether it falls before or after the bound of low types sabotaging high types (in terms of $c_s$) as this determines the behavior of low types, which in turn affects the costs of sabotaging behavior for high types. We compute both bounds and then compare them to establish this sequence.

Lemma 6. *Both high and low types have more to gain from sabotaging high types than low types.*

**B.6.1. Option H1: High types sabotage all other high types and low types do not sabotage anyone** Consider a high type of value $v_h$ not currently sabotaging anyone and sincere voting by everyone else so that $v_h = b_h(n+l+h)$ and $v_l = b_l(n+l+h)$, and $S = hv_h + lv_l$. In this equilibrium, $h$ high types vote $-b_h$ on $h-1$ others.

$$
\begin{aligned}
& \frac{v_h}{S-hb_h(h-1)-b_ll} - \frac{v_h}{S-hb_h(h-1)-b_l(l-1)} \qquad (10)\\
\Leftrightarrow\quad & \frac{v_h(S-hb_h(h-1)-b_l(l-1)) - v_h(S-hb_h(h-1)-b_ll)}{(S-hb_h(h-1)-b_ll)(S-hb_h(h-1)-b_l(l-1))}\\
\Leftrightarrow\quad & \frac{v_hS - v_hhb_h(h-1) - v_hb_l(l-1) - v_hS + v_hhb_h(h-1) + v_hb_ll}{(S-hb_h(h-1)-b_ll)(S-hb_h(h-1)-b_l(l-1))}\\
\Leftrightarrow\quad & \frac{v_hb_l}{(S-hb_h(h-1)-b_ll)(S-hb_h(h-1)-b_l(l-1))}
\end{aligned}
$$

**B.6.2. Option H2: High and low types sabotage all high types** Consider a high type of value $v_h$ not currently sabotaging anyone and sincere voting by everyone else so that $v_h = b_h(n+l+h)$ and $v_l = b_l(n+l+h)$, and $S = hv_h + lv_l$. In this equilibrium, $h$ high types vote $-b_h$ on $h-1$ others and $l$ low types vote $-b_h$ on $h$ high types. This turns out to be the relevant bound (see proof below).

$$\frac{v_h}{S-hb_h(h-1)-lb_hh-b_ll}-\frac{v_h}{S-hb_h(h-1)-lb_hh-b_l(l-1)} \tag{11}$$

$$\Leftrightarrow \quad \frac{v_h(S-hb_h(h-1)-lb_hh-b_l(l-1))-v_h(S-hb_h(h-1)-lb_hh-b_ll)}{(S-hb_h(h-1)-lb_hh-b_ll)(S-hb_h(h-1)-lb_hh-b_ll)}$$

$$\Leftrightarrow \quad \frac{v_hS-v_hhb_h(h-1)-v_hlb_hh-v_hb_l(l-1)-v_hS+v_hhb_h(h-1)+v_hlb_hh+v_hb_ll}{(S-hb_h(h-1)-lb_hh-b_ll)(S-hb_h(h-1)-lb_hh-b_ll)}$$

$$\Leftrightarrow \quad \frac{v_hb_l}{(S-hb_h(h-1)-lb_hh-b_ll)(S-hb_h(h-1)-lb_hh-b_ll)}$$

$$\Leftrightarrow \quad \frac{v_hb_l}{(S-hb_h(h-l-1)-b_ll)(S-hb_h(h-l-1)-b_ll)}$$

□

### B.7. Low Types Start Sabotaging High Types

Similarly, there are two possible bounds when low types start to sabotage high types: one in the equilibrium where high types sabotage low types and one where they do not.

**B.7.1. Option L1: High types sabotage only other high types, but no low types** Consider a high type of value $v_h$ not currently sabotaging anyone and sincere voting by everyone else so that $v_h = b_h(n+l+h)$ and $v_l = b_l(n+l+h)$, and $S = hv_h + lv_l$. In this equilibrium, $h$ high types vote $-b_h$ on $h-1$ all others (high and low types) and low type start sabotaging $h$ high types with $-b_h$. This turns out to be the relevant bound (see proof below).

$$\frac{v_l}{S-hb_h(h-1)-b_hh}-\frac{v_l}{S-hb_h(h-1)-b_h(h-1)} \tag{12}$$

$$\Leftrightarrow \quad \frac{v_lb_h}{(S-hb_h(h-1)-b_hh)(S-hb_h(h-1)-b_h(h-1))}$$

$$\Leftrightarrow \quad \frac{v_lb_h}{(S-hb_hh)(S-b_hh^2+b_h)}$$

**B.7.2. Option L2: High types sabotage all other high types and low types** Consider a high type of value $v_h$ not currently sabotaging anyone and sincere voting by everyone else so that $v_h = b_h(n+l+h)$ and $v_l = b_l(n+l+h)$, and $S = hv_h + lv_l$. In this equilibrium, $h$ high types vote $-b_h$ on $h-1$ other high types and vote $-b_l$ on $l$ low types and low types start sabotaging $h$ high types with $-b_h$.

$$\frac{v_l}{S-hb_h(h-1)-hb_ll-b_hh}-\frac{v_l}{S-hb_h(h-1)-hb_ll-b_h(h-1)} \tag{13}$$

$$\Leftrightarrow \quad \frac{v_lb_h}{(S-hb_h(h-1)-hb_ll-b_hh)(S-hb_h(h-1)-hb_ll-b_h(h-1))}$$

$$\Leftrightarrow \quad \frac{v_lb_h}{(S-hb_hh-hb_ll)(S+b_h-b_hh^2-hb_ll)}$$

*Proof: Order of Bounds.* To establish the correct order of bounds, we show that maximum marginal gain for low types to sabotage high types is higher than the maximum marginal gain for high types to sabotage low types.

$$\frac{v_h b_l}{(S - hb_h(h-1) - b_l l)(S - hb_h(h-1) - b_l(l-1))} < \frac{v_l b_h}{(S - hb_h h)(S - b_h h^2 + b_h)} \tag{14}$$

This inequality holds given $v_l < v_h$ and $b_l < b_h$ which are true by construction. □

This shows that low types have more to gain from sabotaging high types than high types have to gain from sabotaging low types. Consequently, as the cost for sabotage decrease, low types will switch from not sabotaging anyone to sabotaging high types before high types will switch from sabotaging only other high types to sabotaging all other high types and low types. As a consequence, the relevant bound for low types switching to sabotaging high types is the one where only high types sabotage high types but no low types (option H2 above). Furthermore, the correct bound for high types to sabotage low types is one where everyone already sabotages high types (option L1 above). Given that we have now established a sequence of bounds, we establish that the set of possible Nash equilibria is NE = {no sab, h sab h, l sab h, h sab l, l sab l}

### B.8. Nash Equilibria

This contest has many Nash equilibria. Which equilibrium materializes depends on the size of the groups $n$, $l$, and $h$, the costs $c_s$ and $c_p$, and the performance gap $g$ between high and low types.

Proposition 1. *In contests with a positive performance gap $g \geq 1$ between high and low types, the following seven states are the set of possible Nash equilibria.*

(NE1) $s_i = (0,0,0)$ for $i \in H$, and $s_k = (0,0,0)$ for $k \in L$. No one self-promotes, and no one sabotages anyone.

(NE2) $s_i = (0,0,0)$ for $i \in H$, and $s_k = (0,0,1)$ for $k \in L$. Low types engage in self-promotion, and no one sabotages anyone.

(NE3) $s_i = (0,0,1)$ for $i \in H$, and $s_k = (0,0,1)$ for $k \in L$. All agents engage in self-promotion, and no one sabotages anyone.

(NE4) $s_i = (h-1,0,1)$ for $i \in H$, and $s_k = (0,0,1)$ for $k \in L$. All agents engage in self-promotion, high types sabotage each other but do not sabotage low types, and low types do not sabotage.

(NE5) $s_i = (h-1,0,1)$ for $i \in H$, and $s_k = (h,0,1)$ for $k \in L$. All agents engage in self-promotion, high types sabotage each other but do not sabotage low types, and low types sabotage only high types.

(NE6) $s_i = (h-1,l,1)$ for $i \in H$, and $s_k = (h,0,1)$ for $k \in L$. All agents engage in self-promotion, high types sabotage all other agents of both types, and low types sabotage only high types.

(NE7) $s_i = (h-1,l,1)$ for $i \in H$, and $s_k = (h,l-1,1)$ for $k \in L$. All agents engage in self-promotion, and all agents sabotage all other agents of both types.

*Proof.* This follows directly from Lemma 1 (low types will engage in self-promotion before high types do), Lemmas 4-6 which give an order of the bounds, and the additional proof that the benefit for low types sabotaging high types is larger than the benefit of high types sabotaging low types. □

### B.9. Revisiting the Performance Gap $g$

It is now interesting to revisit the condition of $g \geq 1$ for self-promotion. It is clear that the closer $b_h$ is to 1, the less high types have to gain from self-promotion. Consequently, they will only engage in self-promotion if the costs $c_s$ for doing so are increasingly smaller. In the equilibrium condition NE4, for example, in which high types sabotage each other and low types do not sabotage anyone, then $g \geq 1$ is violated if $(n+h+l)b_l > (n+l+1)b_h$ and the resulting gain from self-promotion would be lower for low types than high types. This would then lead to the interesting case where low types do not self-promote while high types do. Considering both self-promotion and sabotage, it is interesting to note the following opposing effect of the performance gap: The larger the performance gap, the more attractive *self-promotion* becomes for *low types* and the more attractive *sabotage* becomes for *high types*. As a result, in contests with a large performance gap, strategic behavior levels the playing field, thus making the contest more attractive to low types.

From the argument above, we can draw another observation: Self-promotion acts as an equalizing mechanism that reduces the performance difference between high and low-skill agents and thus reduces incentives for high-skill agents to sabotage (this can be seen by calculating and comparing Equation 8 in our model with and without self-promotion).

### B.10. Example and Additional Predictions

Let the prize be $M = \$5,000$ and let $h = 10$, $l = 30$, and $n = 100$. Further, let $b_h = 0.8$ and $b_l = 0.2$ (this corresponds to a common case of a finite rating scale from 0-stars to 5-stars where low type agents produce contest entries of 1-star quality and high type agents produce contest entries of 4-star quality). The resulting performance gap between low and high types is then $g = \frac{b_h}{b_l}\frac{n+1}{n+l+h} = \frac{0.8}{0.2}\frac{100+1}{100+30+10} > 1$ which means that high types always have higher values than low types, even when they are sabotaged. Figure A.I shows agent utility $(Mp)$ as a function of cost of sabotage $c_s$. The figure illustrates the transitions between equilibria and their relative sizes in terms of the range of $c_s$ values spanned. The figure suggests that there is one equilibrium in particular that has a large basin of attraction that spans a wide range of $c_s$ values that seem plausible in an online contest (\$0.06 to \$0.23): all agents self-promote and high types sabotage other high types while low types do not sabotage anyone (NE4).

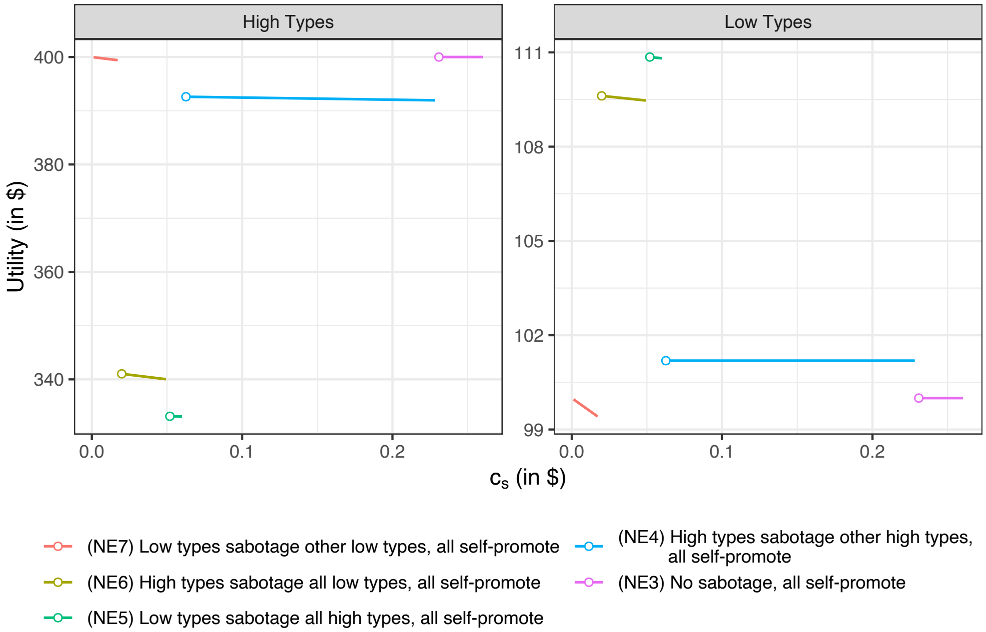
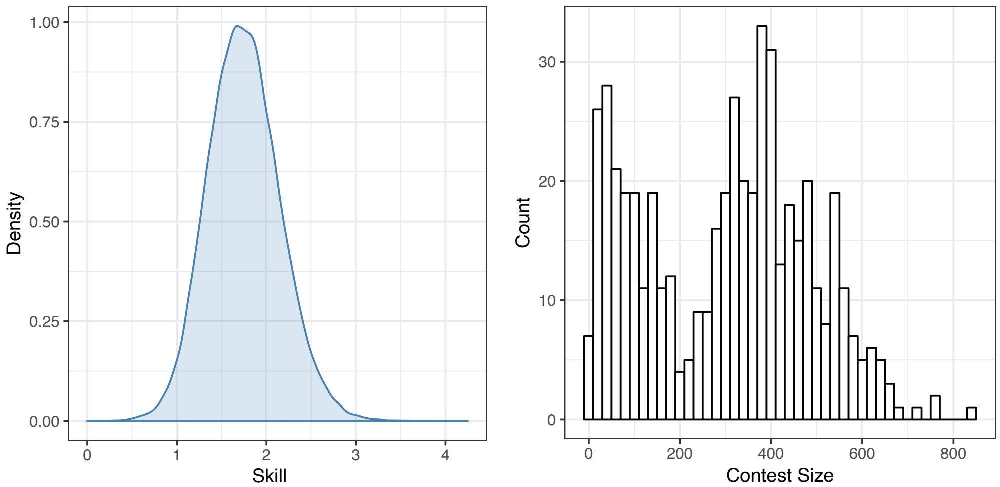



**Figure A.I Equilibria and Utilities.**

*Note:* According to Lemma 4, all equilibria with some form of sabotage also include self-promotion from all agents.

# C. Additional Empirical Analyses and Robustness Tests

## C.1. Distribution of Skill and Contest Size




(a) Distribution of Skill (b) Distribution of Contest Size

**Figure A.II Distributions of Skill and Contest Size.**

*Note:* (a) We compute skill as the average rating of all previous submissions by a designer. The skill distribution is right skewed (skewness: 0.21). (b) Distribution of contest sizes (i.e., number of contest entries; mean: 302).

## C.2. Career Length and Participation

On average, competitors have a career length from their first participation to their last participation that spans 42 contests. During their active tenure, they participate in some form in 9.1 contests, submit their own design in 1.5 contests and rate in 9.0. Submitting to a contest tends to be spaced out between periods of just rating. On average, competitors have 1.4 "flips" in which they go from rating as outsiders to rating as competitors, or from rating as competitors to rating as outsiders. These changes are the basis of our identification: we contrast rating behavior from weeks in which a participant rates (9 contests on average) and weeks in which participants submit their own design while also rating (on average 1.4 contests).

### C.3. Rating Distribution of Contest Winners

We show the percentile scores of contest winners in Figure A.III.

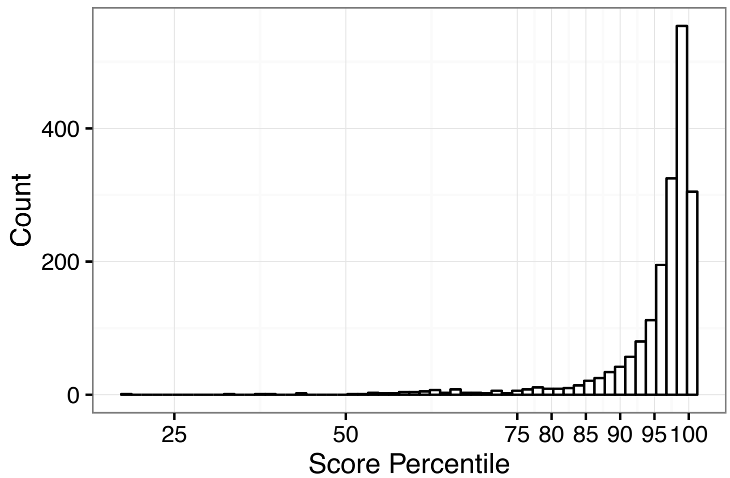



**Figure A.III Percentile score of contest winners.**

*Note:* Contest winners scored in the highest percentiles among all submissions in a given week. The median percentile score was 98th percentile.

### C.4. Heterogeneous Contestant Skill

*Subsample.* As we move from our overall analysis (Table 3) to our examination of heterogeneity by skill (Table 4), we have to restrict our analysis to a subsample of the data: observations for which skill measures for both the source (i.e., competitor casting the rating) and the target (i.e., competitor who submitted the contest entry being rated) are available. This means we have to exclude ratings by competitors who have never before submitted (and hence source skill is unknown) and ratings of submissions by first-time submitters (for whom target skill is unknown). In the interpretation of the main results, one may consequently worry about sampling of subsets of the data. Table A.I shows the main results across subsamples. We repeat the analysis of the full data (Model 1), and the subset for which both skill measures are missing (Model 4; i.e., the data used in Table 4).

Subsamples do not appear to drive our effects. The coefficients for sabotage remain stable across the subsamples ranging from $-0.007$ to $-0.011$. Never-before submitters are harsh raters and excluding these data increases the positivity bias (Model 2). Competitors greet first-timers with more positive ratings and the positivity bias decreases when we exclude those data (Model 3).

### C.5. Robustness of Cluster Level of Standard Errors

In the main text, we cluster standard errors to account for possible dependence of ratings of the same submissions and in the same contest. One may alternatively be concerned about capturing autocorrelation at the rater level. In table A.II we examine the robustness of our specification and cluster on the individual level instead. While we do see some changes in the standard errors, $p$-values are generally much smaller than 0.001 (i.e., $1e-7$) so that our substantial conclusions are not affected by the level of clustering of standard errors.

| Dependent Variable: | Linear Probability | | | |
|---|---|---|---|---|
| | **Sabotage:** 0-Star Rating | | | |
| | All data | Exclude ratings by never before submitters | Exclude ratings of submissions by first-timers | Exclude both |
| | (1) | (2) | (3) | (4) |
| Submitted to same contest: Yes | $-0.008^{***}$ | $-0.011^{***}$ | $-0.007^{***}$ | $-0.011^{***}$ |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Rate own submission: Yes | $-0.195^{***}$ | $-0.195^{***}$ | $-0.186^{***}$ | $-0.186^{***}$ |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| *Individual* | *Fixed* | *Fixed* | *Fixed* | *Fixed* |
| *Submission* | *Fixed* | *Fixed* | *Fixed* | *Fixed* |
| Adj. $R^2$ | 0.372 | 0.391 | 0.367 | 0.386 |
| Num. obs. | 38,102,880 | 27,188,751 | 26,317,775 | 18,787,584 |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

**Table A.I Analysis of subsamples.**

| Dependent Variable: | Linear Probability | |
|---|---|---|
| | **Sabotage** 0-Star Rating | **Self-Promotion** 5-Star Rating |
| | (1) | (2) |
| Submitted to same contest: Yes | $-0.057^{***}$ | $-0.000$ |
| | (0.008) | (0.001) |
| Rate own submission: Yes | $-0.179^{***}$ | $0.996^{***}$ |
| | (0.003) | (0.016) |
| Target skill | $-0.112^{***}$ | 0.008 |
| | (0.011) | (0.012) |
| Source skill | $-0.049^{**}$ | $-0.105^{***}$ |
| | (0.017) | (0.015) |
| Target skill × Source skill | 0.007 | $0.042^{***}$ |
| | (0.006) | (0.006) |
| Submitted to same contest: Yes | | |
| × Target skill | $0.015^{***}$ | |
| | (0.003) | |
| × Source skill | $0.010^{***}$ | |
| | (0.003) | |
| Rate own submission: Yes × Source skill | | $-0.080^{***}$ |
| | | (0.009) |
| *Individual* | *Fixed* | *Fixed* |
| *Contest* | *Fixed* | *Fixed* |
| Adj. $R^2$ | 0.360 | 0.208 |
| Num. obs. | 18,787,584 | |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

**Table A.II Robustness test of heterogeneous effects by skill level with errors clustered on the individual-level.**

*Note:* Skill is measured as the average quality of all past submissions. Contest-level fixed effects are used instead of submission-level as the ability of agents is time invariant at the submission level. Standard errors are in parentheses, clustered at the individual level.

### C.6. Skill Measure

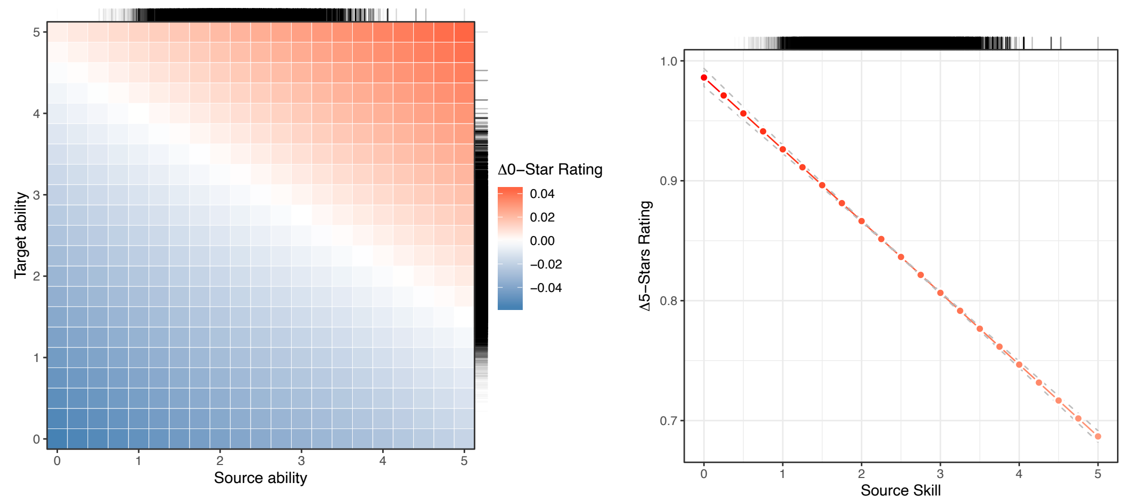
The analyses presented in the main body measure the idea generation skill of individuals as the average quality of all past submissions. We repeat that analysis using the quality of the best design they submitted in the past (Table A.III and Figure A.IV). We find very similar results.

| | Linear Probability | |
|---|---|---|
| Dependent Variable: | 0-Star Rating | 5-Star Rating |
| | (1) | (2) |
| Submitted to same contest: Yes | $-0.055^{***}$ | $-0.000$ |
| | (0.004) | (0.001) |
| Rate own submission: Yes | $-0.179^{***}$ | $0.986^{***}$ |
| | (0.004) | (0.007) |
| Target max skill | $-0.066^{***}$ | $-0.001$ |
| | (0.002) | (0.001) |
| Source max skill | $-0.019^{***}$ | $-0.084^{***}$ |
| | (0.004) | (0.003) |
| Target max skill × Source max skill | 0.002 | $0.027^{***}$ |
| | (0.001) | (0.001) |
| Submitted to same contest: Yes × Target max skill | $0.012^{***}$ | |
| | (0.001) | |
| Submitted to same contest: Yes × Source max skill | $0.008^{***}$ | |
| | (0.001) | |
| Rate own submission: Yes × Source ability | | $-0.060^{***}$ |
| | | (0.003) |
| *Individual* | *Fixed* | *Fixed* |
| *Contest* | *Fixed* | *Fixed* |
| Adj. $R^2$ | 0.360 | 0.211 |
| Num. obs. | 18,787,584 | |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

**Table A.III Estimates of heterogeneous effects using quality of the best prior design as skill measure.**

*Note:* Contest-level fixed effects are used instead of submission-level because the skill of agents is time invariant at the submission level. Standard errors are in parentheses, clustered at the contest level.




(a) SABOTAGE (b) SELF-PROMOTION

**Figure A.IV Strategic behavior by competitors of heterogeneous skill (max skill).**

*Note:* **Panel A (Sabotage).** Relative change in probability of rating 0-star when competing compared to not competing across skill levels. Outer margins show the distribution of data. There are over 15,000 (10,000) observations for source (target) skill greater than 4.5. **Panel B (Self-promotion).** Relative change in probability of rating 5-star when rating own submission compared to submissions by others of the same skill (error band is 95% confidence interval).

### C.7. Sources of Sabotage

In this section we explore tenure, experience, and being a past winner as different mechanism behind the decision to sabotage others (Table A.IV). Similar to the analysis in the main text, we introduce interaction terms between the *Submitted to same contest* dummy with different time-varying measures of individual attributes. All models control for source and target skill and thus need to be interpreted as drivers above and beyond the finding that sabotage is predominantly used among high-skill competitors targeting other high-skill competitors.

| | Linear Probability | | |
|---|---|---|---|
| Dependent Variable: | **Sabotage:** 0-Star Rating | | |
| | Tenure | Experience | Past Success |
| | (1) | (2) | (3) |
| Submitted to same contest: Yes | $-0.068^{***}$ | $-0.049^{***}$ | $-0.058^{***}$ |
| | (0.001) | (0.001) | (0.001) |
| Rate on own submission: Yes | $-0.192^{***}$ | $-0.196^{***}$ | $-0.194^{***}$ |
| | (0.001) | (0.001) | (0.001) |
| Tenure (in weeks) | $0.017^{***}$ | | |
| | (0.000) | | |
| Rating experience (log) | | $0.031^{***}$ | |
| | | (0.000) | |
| Submission experience (log) | | $-0.010^{***}$ | |
| | | (0.000) | |
| Source is past winner | | | $-0.026^{***}$ |
| | | | (0.000) |
| Target is past winner | | | $0.012^{***}$ |
| | | | (0.000) |
| Submitted to same contest: Yes | | | |
| × Tenure | 0.000 | | |
| | (0.000) | | |
| × Rating experience | | $-0.010^{***}$ | |
| | | (0.000) | |
| × Submission experience | | $0.011^{***}$ | |
| | | (0.000) | |
| × Source is past winner | | | $0.021^{***}$ |
| | | | (0.004) |
| × Target is past winner | | | $0.013^{***}$ |
| | | | (0.000) |
| *Skill controls* | *Yes* | *Yes* | *Yes* |
| *Individual* | *Fixed* | *Fixed* | *Fixed* |
| *Submission* | *Fixed* | *Fixed* | *Fixed* |
| Adj. $R^2$ | 0.391 | 0.391 | 0.391 |
| Num. obs. | | 27,188,751 | |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

**Table A.IV Regression of different mechanisms underlying strategic behavior.**

First, we explore tenure on the platform (measured in months between the date they signed up and the date a rating was submitted). We find no significant effect (Model 1). We also tested the interaction with a quadratic term (not shown) but find no effect either. Second, we find mixed results with regard to submission and rating experience (Model 2). While community members with higher rating experience appear to be less likely to sabotage, those with higher submission experience appear to be more likely. This is consistent with the interpretation that community members with a strong competitive motivation (making many submissions) are more likely to act strategically compared to community members who may be more socially motivated (evaluating many submissions by others). Third, we find past winners are significantly more likely to sabotage and are more likely to be the targets of sabotage.

## D. Leniency as Variation within Contests

Table A.V shows regression on the contest-level using standard deviation of ratings a user casts as dependent variable. We find that when individuals have submitted to the same contest, they make better use of the full rating spectrum and submit ratings with a higher standard deviation (Model 1: $\beta = 0.037; p < 0.001$). The effect is amplified by skill so that higher-skilled individuals submit ratings with an even higher standard deviation (positive interaction between skill and having submitted to the same contest in Model 2: $\beta = 0.017; p < 0.001$).

| | OLS | |
|---|---|---|
| Dependent Variable: | **Std.Dev. of Ratings** | |
| | (1) | (2) |
| Submitted to same contest: Yes | $0.037^{***}$ | 0.004 |
| | (0.002) | (0.009) |
| Source ability | | $-0.031^{***}$ |
| | | (0.007) |
| Submitted to same contest: Yes × Source ability | | $0.017^{***}$ |
| | | (0.005) |
| *Individual* | *Fixed* | *Fixed* |
| *Contest* | *Fixed* | *Fixed* |
| Adj. $R^2$ | 0.299 | 0.299 |
| Num. obs. | 439,882 | 439,882 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

**Table A.V User-contest level analysis of standard deviations of ratings cast.**

### D.1. Substitution of sabotage and self-promotion

| Dependent Variable: | Linear Probability<br>**Sabotage:** 0-Star Rating | |
|---|---|---|
| | (1) | (2) |
| Submitted to same contest: Yes | $-0.006^{***}$ | $-0.057^{***}$ |
| | (0.000) | (0.000) |
| Rate own submission: Yes | $-0.194^{***}$ | $-0.205^{***}$ |
| | (0.001) | (0.000) |
| Self-promoted in this contest | $-0.002^{**}$ | $0.031^{***}$ |
| | (0.001) | (0.001) |
| *Individual* | *Fixed* | *Fixed* |
| *Submission* | *Fixed* | *Fixed* |
| Adj. $R^2$ | 0.372 | 0.064 |
| Num. obs. | 38,102,880 | |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

**Table A.VI Regression models investigating whether individuals use sabotage and self-promotion together or substitute one for the other.**

### D.2. Robustness: Dyad Fixed Effects

As a robustness test, we investigate dyadic patterns underlying our findings of sabotage. Research on dyadic rivalry may be an alternative explanation (Kilduff et al. 2016, Grad et al. 2023). For example, individuals may “sabotage” others with whom they have competed in the past and against whom they have lost. One may hence wonder if there are (or emerge over time) dyads in which individuals systematically sabotage others. Some dyads may rate each other with 0-stars whether or not they are competitors in the current contest and rivalry could be an alternative explanation to the strategic motivation (i.e., the behavior is founded on the subjective feeling of rivalry instead of the strategic motivation to increase ones chance of winning the contest prize). If so, that may lead to an overestimation of our effects in the sense that we attribute it to being “strategic” rather than “rivalrous”. To investigate this alternative explanation, we estimate a variation of our main model with dyad-level fixed effects (Table A.VII). Since rivalry is generally assumed to require some time to form (Kilduff et al. 2016, Grad et al. 2023), we restrict our analysis to dyads with ten or more encounters (e.g., $A$ rating five of $B$’s submissions and $B$ rating five of $A$’s submissions). Note that the effect we report in the table is driven by many fewer observations: only those dyads who rate each other at least once in the same contest and once outside the same contest contribute to the parameter estimate. Controlling for unobserved dyadic rivalries through these dyad-level fixed effects (instead of our usual individual- and submission-level fixed effects) we find that likelihood to sabotage others when competing (the main effect) *increases* by more than half ($-0.070$ vs. $-0.043$). If the “sabotage” effect we report in our estimates were driven by unobserved rivalries, then we would expect the effect of rating ones competitors to *decrease*. Furthermore, we find significant positive effects for sabotage committed by high-skilled targeting other high-skilled just as in the analysis reported in the main text (interaction term coefficients $\beta = 0.015; p < 0.001$ for sources and $\beta = 0.002; p = 0.053$ for targets). This suggests that sabotage happens outside, and on top of, any pre-existing dyadic rivalries. That is, even within the same dyad, we still find a difference in the rating behavior between competing for the same prize vs. not competing. In summary, dyadic rivalry patterns are not sufficient to explain the strategically motivated sabotage.

| Dependent Variable: | Linear Probability<br>**Sabotage:** 0-Star Rating<br>10+ Encounters |
|---|---|
| | (1) |
| Submitted to same contest: Yes | $-0.043^{***}$ |
| | (0.004) |
| Target skill | $-0.029^{***}$ |
| | (0.008) |
| Source skill | $-0.025^{**}$ |
| | (0.009) |
| Target skill × Source skill | $-0.006^{\cdot}$ |
| | (0.004) |
| Submitted to same contest: Yes | |
| × Target skill | $0.002^{\dagger}$ |
| | (0.001) |
| × Source skill | $0.015^{***}$ |
| | (0.001) |
| *Source-Target Dyad* | *Fixed* |
| Adj. $R^2$ | 0.448 |
| Num. obs. | 36,94,941 |
| Num. dyads | 22,9404 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

**Table A.VII Robustness tests using dyad fixed effects.**

## D.3. Natural Experiment 1: Prize Increase

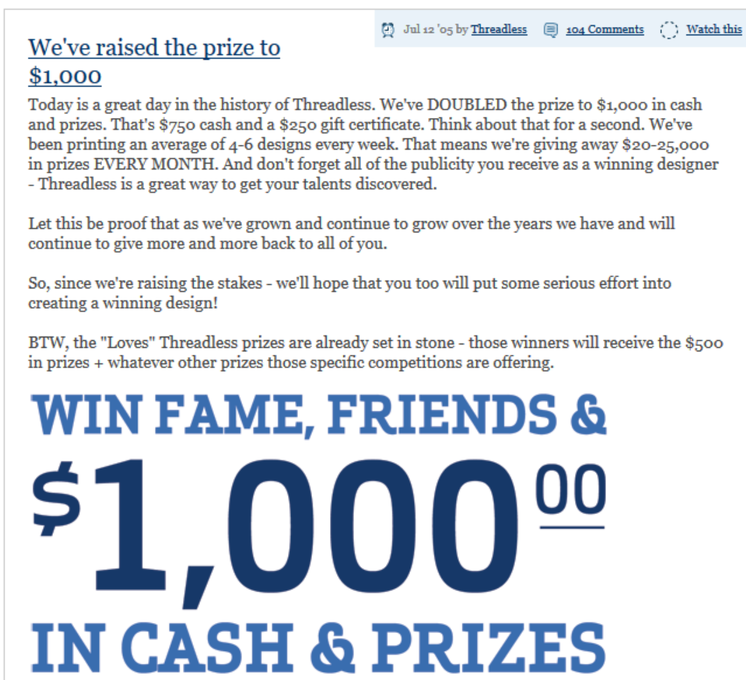



**Figure A.V Screenshot natural experiment sabotage.**

*Note:* Screenshot of the Threadless website showing the announcement of the prize increase posted on July 12, 2005.

### D.4. Natural Experiment 2: Scoring Rule Change

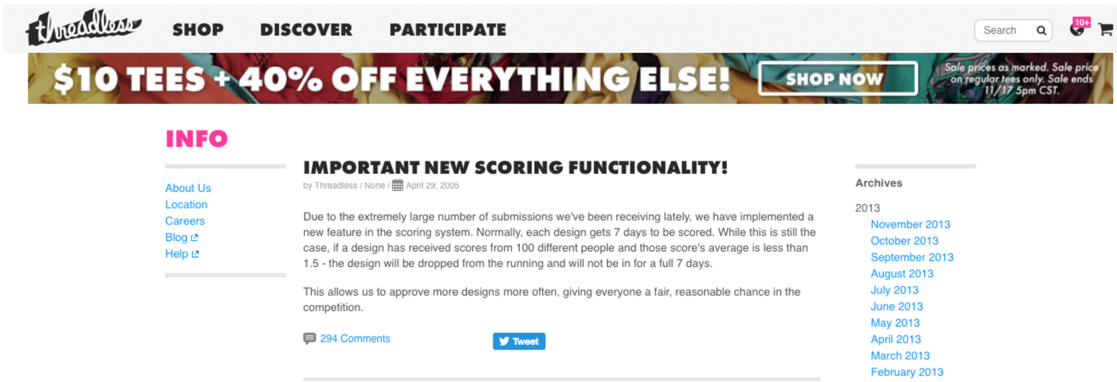



**Figure A.VI** **Screenshot natural experiment self-promotion**

*Note:* Screenshot of the Threadless website showing the announcement of the scoring rule change posted on April 29, 2005.

### D.5. Long-Term Participation

Table 6 in the main text shows coefficient estimates for the hazard rate of participation in the next round. The hazard rate can also be shown visually (Figure A.VII). The figure contrasts how high- versus low-skilled competitors (10th vs. 90th percentile, respectively) react to receiving high versus low levels of sabotage (also 10th vs. 90th percentile, respectively).

## E. Effect of Strategic Rating on Selection of Contest Winners

Does strategic rating behavior affect the rank-order of contest submissions, especially given the generally large contests we observe? To assess the impact of *self-promotion*, we exclude all ratings from designers who rated their own designs, recalculate average ratings for submissions, and then re-rank all submissions based on the new averages. We find that, keeping everything else constant, in seven out of 511 contests (1.4%), the winner of the contest changes; and in 28 of 511 contests (5.5%), there is a change in at least one of the top three ranks. To analyze the effect of *sabotage*, we exclude all ratings from competitors as they might be strategically motivated. Note that this also includes votes from designers on their own designs, i.e., it contains the previous reported effects. We find that in 12% of the contests, the winner of the contest changes; and that in 48% of the contests, there is a change in the top three ranks. The effect is especially pronounced in close contests where the contest winner would change in 25% of cases and 65% would see a change in the top three. This can be considered an upper boundary for the effect of strategic behavior via the rating mechanism in our setting.

This change in rankings, however, may be simply due to the fact that removing *any* ratings changes the resulting ranking. So to complement this first test, we compare the effect of removing the ratings of potentially strategic raters, to the effect of removing the ratings of an equal amount of randomly selected raters. Removing random raters, contest winner changes on average 10.7% of cases across our bootstrap simulation, while contest winners change 12% of the time when removing ratings by potentially strategic

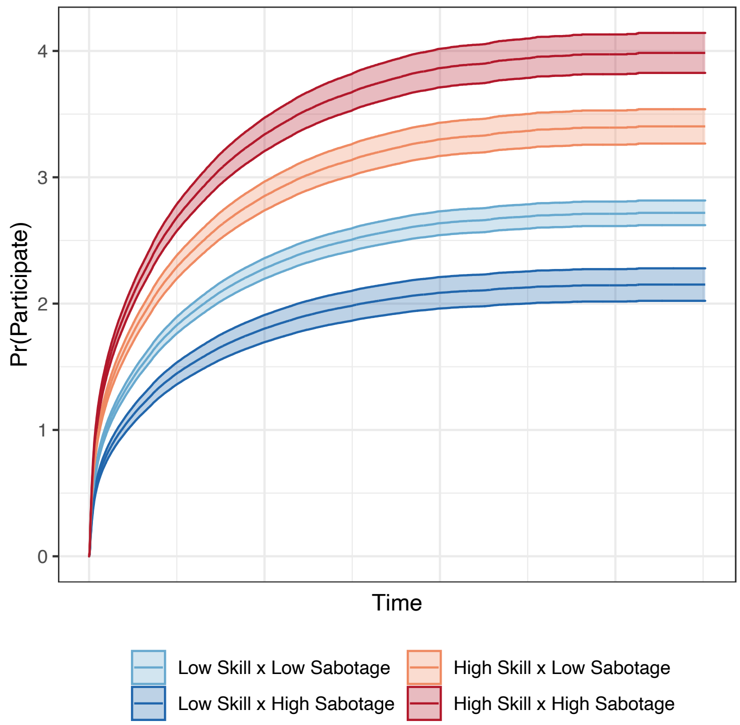



**Figure A.VII Likelihood of participating in a future contest.**

*Note:* Likelihood for high skilled (90th percentile) and low skilled (10th percentile) individuals based on experience in the current contest for individuals who received high (90th percentile) and low (10th percentile) levels of sabotage (cumulative hazard; based on Model 1 from Table 6; controls include rating received, number of ratings received, amount of competition, quality of competition)..

raters. In close contests, contest winner changes 13.8% in the null model vs. 25% winner change when removing potentially strategic raters. Removing strategic ratings affect outcomes more than just removing random raters which lends further support to our claim of the important role of strategic rating. Overall, these two analyses suggest that strategic behavior of idea generators during idea selection, does affect Threadless' ability to identify the most promising ideas. The effect is strongest in close competition where up to 25% of contest winners may change and around two-thirds of contests would have changes in the top-three.

To further quantify the effect of strategic rating, we compute a baseline of how much a contests rank-order would change if the ratings of some randomly chosen individuals were removed. For each contest, we compute the number of competitors who did cast potentially strategic ratings, and select the same amount of raters at random. To remove the same number of ratings, we need to remove about 20% more raters since competitors who submitted to the same contest rate more submissions that those who did not. We remove their ratings and recompute the ranking to see how much has changed. We repeat this 500 times for all contests in our data in order to get a robust null model of how much rankings would change after removing random raters.