Resource Competition in an Ideological Environment *

Justin Esarey
Assistant Professor of Political Science
Emory University
jesarey@emory.edu

March 25, 2012

Abstract

Previous experiments have indicated that tournament incentives can motivate workers to be productive in environments where other incentives are infeasible. This paper presents an experiment where tournament incentives are applied to an ideologically-charged environment, one in which the participants are identified and grouped ideologically and where the task has an ideological valence. Behavior in this experiment is compared to behavior in a non-ideological environment to assess the impact of ideology. I find that participants have a taste for the ideological valence of the task that makes them more or less productive, net of their financial incentives. I also find that, when competitors share an ideological commitment, they are more likely than those with a minimal group affiliation to collude against the principal by mutually lowering output. Conversely, competitors who have opposing ideological commitments compete harder than those have different minimal group affiliations. A substantive implication is that leveraging competitive incentives inside government bureaucratic agencies may be most effective when these agencies are ideologically heterogeneous.

Introduction

How are competitive incentives influenced by ideological motivations? A great deal of formal and experimental work indicates that tournaments, in which agents compete for rewards and are judged relative to one another, can motivate effort in situations where other managerial techniques cannot be employed. In particular, tournaments place lower information

*I thank the National Science Foundation for its generous support of my research (#0720055). I also thank William Berry, John Scholz, Bumba Mukherjee, TK Ahn, Tim Salmon, Eric Dickson, Jeff Lax, Tom Clark, and Randy Strahan for their helpful comments and feedback on earlier drafts of the paper.
and monitoring requirements on the supervisor: rather than comparing each agent’s effort or output levels to a benchmark, the supervisor need only be able to qualitatively rank agents relative to one another (even with some error) to determine who gets rewarded. This feature is important when output is difficult to monitor or quantitatively assess, as in many bureaucratic settings (Miller, 1992, Chapter 6; Brehm and Gates, 1999).

The effectiveness of tournament incentives, in combination with their comparatively low informational requirements, has led some scholars to suggest placing civil servants, agency subunits, or even entire agencies into explicit competition with one another for resources (Gailmard and Patty, 2007; Gersbach and Keil, 2005; Whitford, 2006; Ting, 2002). These resources might include awarding important projects, bigger budgets, supervisory authority, staffing increases, and other perks to the most productive or compliant units. This approach is especially sensible given the reality of (and strategic incentives to promulgate) extensive redundancy in bureaucratic systems (Landau, 1969; Bendor, 1985; Ting, 2003): by establishing competition for funding among redundant units, bureaucratic efficiency can be increased via reorganizing existing assets. Competitive budgeting also turns Niskanen’s (1971) classic argument to unexpected use: bureaucrats’ desire for larger agencies and more projects can be leveraged to increase the efficiency of the system and ensure its compliance with the goals of the elected government. In both these ways, the effects of competitive budgeting are counterintuitive: they turn supposed disadvantages of bureaucracy into useful tools.

But before such a project could be considered, I believe it important to ask the basic question of whether tournament incentives can be effective in an ideologically charged environment. For instance, it would be helpful to know (a) whether subjects’ ideological taste (or distaste) for a task influences their productivity in a tournament, and (b) whether ideological affinity with competitors makes an agent more or less willing to compete for financial rewards. I answer these questions using a simple laboratory experiment that abstracts away from the complexities of bureaucratic behavior to examine how ideology changes subjects’ response to resource competition. The experiment is designed to lay some micro-level groundwork for
understanding the interaction between ideology and resource competition before tackling the more involved question of how agents facing the complex incentives and sociological pressures of a bureaucracy might respond to competitive incentives.

I find that participants have a taste for the ideological valence of the task that makes them more or less productive, net of their financial incentives. I also find that, when competitors share an ideological commitment, they are more likely than those with a minimal group affiliation to collude against the principal by mutually lowering output. Conversely, competitors who have opposing ideological commitments compete harder than those have different minimal group affiliations. The substantive upshot is that key organizational choices might factor into the success of competitive incentives in an administrative bureau, and deserve further inquiry. In particular, there might be a tradeoff between picking agents that have an ideological preference for the task—making them want to contribute more effort—and ensuring enough ideological heterogeneity to ensure that competition is not undermined by common ideological group affiliation. In turn, the finding gives us a reason to think that a heterogeneous bureaucratic agency motivated by competitive incentives might be more efficient than a comparable organization without those incentives, while enjoying the expertise benefits of a merit civil service system.

**Resource Competition and Ideology**

Many standard forms of employment contracting are hard to implement in government bureaucracies because politicians do not influence the tenure and promotion of most civil servants.¹ These protections exist because (i) it is tempting for politicians to use civil service positions as patronage rewards, thereby undermining the competence, impartiality, and institutional memory of the administration (Johnson and Libecap, 1994, Chapter 2), and (ii)

---

¹For high-level positions (e.g., those in the Senior Executive Service), the more standard contracts are tenable because these individuals serve at the pleasure of the president and can be terminated at will. However, these positions form a tiny fraction of the total civil service: in 2004, the SES was allocated 7,868 positions out of a total of 2,649,300 civilian employees in the executive branch (United States Office of Personnel Management, 2005).
it is difficult to create an objective measure of productivity against which a civil servant can be measured (Miller, 1992, Chapter 6; Brehm and Gates, 1999). Hence, a civil servant’s extrinsic motivation for good performance is weak, as it is difficult to demote or fire them after a probationary period (Johnson and Libecap, 1989b, pp. 59-60; see also Johnson and Libecap, 1989a). Nor can politicians simply obey the ally principle (Crawford and Sobel, 1982; Bendor and Meirowitz, 2004) and hire their ideological comrades, who share their policy commitments and therefore have an intrinsic commitment for good performance: this is specifically forbidden by the merit civil service system for all but high-level appointees.

As a consequence, Ting (2002) proposes that separating jurisdiction over policy tasks and offering budgetary slack rewards to each agency for the successful completion of their own task would improve bureaucratic efficiency and compliance. Because redundancy is already a fact of bureaucracy (Landau, 1969; Bendor, 1985; Ting, 2003), the only costs are organizational; already existing units can be placed in competition. Over the year, the supervisor monitors agency performance. During annual budget formulation, the supervisor then reduces the budget of lower-performing individuals and/or units, awarding the surplus to more productive and compliant individuals and/or units. The incentives created are quite robust to error in monitoring because relative performances are compared instead of absolute performances; this is established in the formal model below. Proposals in a similar vein have been advanced by several others (Whitford, 2006; Gailmard and Patty, 2007).

But, even setting aside the many complexities of implementing competition inside the bureaucracy, there are several potential ways in which the ideologically charged nature of the work might change the effectiveness of competitive incentives. For example, experimental evidence indicates that identifying oneself as a member of a group—even a “minimal group” assigned in a lab by a trivial criterion, such as preference for one painting over another (Tajfel et al., 1971)—can influence a subject’s response to financial incentives in a strategic environment. Organizations, such as corporations and the military, are well-aware of the effect of identity on behavior and take measures to wed the identity of their members to
the organization in order to secure their loyalty (Akerlof and Kranton, 2005). But political environments tend to activate and highlight pre-existing ideological and/or partisan group identification, group memberships that are hard to reshape.

In general, subjects in laboratory experiments tend to value the welfare of fellow members of a group (including minimal groups) more than they value the welfare of non-members (Chen and Xin Li, 2009, p. 452). For example, subjects tend to make less-selfish choices in sequential Prisoner’s Dilemmas and dictator games when paired with minimal group members compared to non-members (Simpson, 2006; Chen and Xin Li, 2009). Group membership can also change how subjects respond to other characteristics of the environment. Pre-play discussion increases contributions to a public good, but only when one’s own group benefits from the good (Dawes, de Kragt and Orbell, 1988). Similarly, choosing in front of an audience of one’s own group members induces more aggressive play in coordination and Prisoner’s Dilemma games played against a non-group member, compared to private play (Charness, Rigotti and Rustichini, 2007). When playing the role of a third-part enforcer, people are less likely to punish other subjects who enrich themselves at another’s expense if the victim is not a group member (Bernhard, Fehr and Fischbacher, 2006; Gozette, Huffman and Meier, 2006; Chen and Xin Li, 2009). On the other hand, subjects in a voting experiment tend to respond less powerfully to group affiliation in their voting choices when their financial self-interest conflicts with this identity (Bassi, Morton and Williams, 2011).

Given this literature, there are at least two possibilities worth investigating. First, work that benefits one’s own group (e.g., fellow ideologues) may be valued more highly than work that benefits an opposing group. If distaste for effort outweighs the financial incentives—and these are reasonably weak in a bureaucratic setting—then the incentives will not be effective. A similar effect would exist if the bureaucrat’s work had an undesirable policy valence (worked toward policies that the bureaucrat did not support).

Second, and more subtly, bureaucrats might not compete as hard when pitted against others who share their party or ideology. As researchers have recognized for years (?) , the
efficiency of competitive incentives is threatened when competitors agree to mutually lower their output; if successful, this allows everyone to maintain the same probability of receiving resources but at a lower effort cost. Experiments have confirmed the empirical possibility of collusion in these environments (??). In an ideologically charged production environment, collusion may be enhanced because (a) competitors are concerned with their fellow partisans’ or ideologues’ welfare and may internalize it as a part of their own, and (b) may trust fellow partisans or ideologues more than outsiders. Why do these two considerations matter? Working harder raises one’s own expected payoff at the expense of co-workers’ expected payoffs, an undesirable outcome if one internalizes the co-worker’s welfare as one’s own. A competitor might therefore be tempted to cut productivity to benefit his/her fellows. Similarly, the increased trust between co-workers may enable them to collude against the supervisor: they can both scale back their effort levels, thereby cutting back on the costs of effort, without changing their probability of a budget cut (recall that this probability depends on relative, not absolute, effort levels because of the supervisor’s limited information).

Thus, there are two questions of interest. The first, and most straightforward, is how the ideological valence of production activities changes behavior. The second question involves how the organizational structure of the production environment can condition the effect of ideology on behavior. That is, it will be valuable to know not only the direct effect of ideological taste for a task on work effort, but also how the configuration of ideological preferences inside the organization conditions work effort. Consider three cases: (a) an ideologically homogeneous group, (b) a group that is ideologically heterogeneous as a whole but has comparatively homogeneous work units that are different from one another, and (c) an ideologically heterogeneous group with heterogeneous work units. Because of the direct effect of ideological tastes and the indirect effect of social preferences created by ideological homogeneity or heterogeneity, all three of these cases may be very different.
Illustrating the Possibilities with a Formal Model

It may be helpful to illustrate the ways in which ideological preferences can change behavior in a resource competition using a formal model. The model I propose is closely related to that of Schotter and Weigelt (1992), who developed a framework for analyzing competitive behavior in asymmetric tournaments that I use here (see also Bull, Schotter and Weigelt, 1987; Lazear and Rosen, 1981). Their work, however, does not consider the role of ideological taste and social preferences as a part of their model. My model incorporates these preferences into the utility functions of bureaucrats, allowing me to describe how competitors’ behavior can change as ideology and social preference become a more important part of the production environment. The initial cut at the model uses a one-shot game; I then examine how it changes if the game is infinitely repeated over time.

Utility Functions

Assume that there are two agents, $i$ and $j$, working under a supervisor. Each agent cares about the benefits s/he receives from the job (personal, career, and ideological) and the costs of exerting effort to safeguard those benefits. In a competitive environment, s/he wants to avoid losing those benefits by exerting as much or more effor than his/her co-workers. Finally, the agent may care directly about the well-being of his/her co-workers because of social preferences.

This preference structure can be represented in terms of either gains or losses. In terms of gains, we might write:

$$u_i = B\pi_i + A + (F + \alpha_i)(e_i + e_j) - e_i^2 + \delta_i[B\pi_j + A + (F + \alpha_j)(e_i + e_j) - e_j^2]$$  \hspace{1cm} (1)

That is, an agent starts with a base level of resources $A$, competes for some benefit $B$ with a probability of receiving the benefit $\pi_i$, has a cost of exerting effort $e_i$. An agent may derive a personal direct benefit $F$ in proportion to total effort exerted toward the task.
(e.g., effort toward environmental protection may benefit a bureaucrat’s health), and an ideological benefit according to an individual parameter $\alpha_i$. The $\alpha_i$ parameter captures the benefit or harm that an agent receives for exerting effort that benefits some ideological group and/or has a policy valence. An agent $i$ may value agent $j$’s utility as his own, with a weight $\delta_i \in (-1, 1)$ assuming that $i$ values his own utility more than that of others.

Alternatively, we can write the utility in the form of losses:

$$u_i = B - C\pi_i + (F + \alpha_i) (e_i + e_j) - e_i^2 + \delta_i \left( B - C\pi_j + (F + \alpha_j) (e_i + e_j) - e_j^2 \right)$$  \hspace{1cm} (2)

Here, the only difference is that an agent starts with a bundle of resources $B$ and faces the potential for a resource cut $C$ that is levied with probability $\pi_i$. decreases in the size of the some penalty $C$. There may be a psychological (non-financial) cost of losing the competition regardless of the consequences, which gets wrapped up in the $C$ term as an additional factor. We may speak of $C = C_o + C_p$, where $C_o$ is the objective (financial or career) cost of losing the competition and $C_p$ is the strictly psychological cost of losing.$^2$

For the remainder of the paper and with little loss of generality, I use the “loss” form of the utility function, equation (2), because, with finite resources to distribute among multiple units, a cut is presumably easier to implement than an expansion. There may also be a motivational reason to frame the allocations as cuts: while the predictions for behavior would be the same either way if agents are risk-neutral, risk-averse agents will tend to see budget cuts as a greater influence on their utility and therefore be more motivated to avoid them.

Each unit wants to avoid receiving a cut in their baseline level of resources, $C_o$, and may also dislike losing regardless of the consequences, $C_p$. The probability of receiving such a cut,$^2$Bureaucrats are risk-neutral in this model, but making them risk averse would only reinforce the conclusions of the model. Budget cuts $C$, the only probabilistically-determined payoffs in the utility function (and thus the only ones relevant to risk aversion), would be more heavily valued relative to the reference point $B$, and thus bureaucrats would respond more strongly to threatened cuts. The only effect is to make the cuts $C$ more effective; consider replacing $C$ in my model with $C' = \zeta \ast C$ where $\zeta$ represents overvaluation due to risk aversion. All comparative statics involving $C$ would simply involve $C'$, a simple transformation of the original budget cut $C$.}
\(\pi_i\), will be a function of the effort exerted by their own unit \(e_i\) and by other, similar units that are competing for resources \(e_j\). Of course, even without these incentives, agents will take pride in and derive ideological benefits from their work, at the rate \(\alpha_i\), and may even directly benefit from it at the rate \(F\). Social preferences may cause civil servants to want to work more or less hard: they may identify with their colleagues' welfare and therefore be reticent to compete against them (\(\delta_i > 0\)), or they may dislike their colleagues and be indifferent toward their utility or even desire to harm them (\(\delta_i \leq 0\)).

**Supervision**

The supervisor, a non-strategic actor in this scenario, compares the output of the two agents \((e_i\) and \(e_j\)) and punishes the one whose output s/he perceives to be lower by cutting that agent’s resources. In organizations with difficult-to-observe work outputs and limited monitoring ability, it is likely that the supervisor will be unable to perfectly observe \(e_i\) and \(e_j\). The supervisor will instead compare \(e_i + \varepsilon_i\) and \(e_j + \varepsilon_j\), where \(\varepsilon_i\) is an error-in-observation term for agent \(i\)'s production, and make the following decision:

- If \(e_i - e_j > \varepsilon_j - \varepsilon_i\), cut \(j\)'s resources.
- If \(e_i - e_j < \varepsilon_j - \varepsilon_i\), cut \(i\)'s resources.
- If \(e_i - e_j = \varepsilon_j - \varepsilon_i\), cut either agent’s resources with 1/2 probability.

Assume that \(\varepsilon_j - \varepsilon_i\) is distributed according to the uniform distribution \(U(-e_j, e_i)\). This assumption is made to show that the incentives are robust to a supervisor with extremely noisy information: the uniform distribution implies (for example) that one-third of the time the supervisor cuts the budget of an agent exerting twice as much effort as his/her competitor!

Under this assumption, the probability of receiving a cut is:

---

3The supervisor could be made a strategic actor by, e.g., providing him/her with a simple incentive to pick the higher-producing agent subject to the monitoring constraints described below. But this extension would predict the same supervisor behavior described below, and so I maintain the supervisor as a non-strategic actor for simplicity.
\[
\pi_i = \frac{e_j}{e_i + e_j}
\]

This probability is intuitively equivalent to a raffle draw in which effort levels constitute tickets. Hence, exerting greater effort always increases the probability of avoiding the budget cut, but convergence to certainly avoiding the cut is slower than if the supervisor’s ability to distinguish effort levels were finer-grained.

**Equilibrium Comparative Statics**

As demonstrated in the appendix, the utility functions of each actor are concave with respect to his/her own effort choice. Consequently, each actors’ best response to the other’s effort level must be unique. Any action profile \(e^* = [e_i^*, e_j^*]\) that simultaneously solves the relevant first order conditions must therefore be a Nash equilibrium. Solving for an explicit \(e^*\) would be very challenging analytically. Fortunately, we can determine the comparative statics of how \(e_i^*\) and \(e_j^*\) change as some parameter \(\gamma\) changes using the implicit function theorem without explicitly solving for \(e_i^*\) and \(e_j^*\).

The point of the model is to illustrate two ways in which effort might change as a result of ideological influence: the relationship between effort \(e^*\) and ideological taste for the task \(\alpha\), and the relationship between \(e^*\) and utility interdependence \(\delta\) changes. Both \(\alpha\) and \(\delta\) might be functions of ideology: ideological preferences can make a task more (dis)tasteful, and can provide a basis for solidarity (or division) with co-workers. I assume that \(\delta \in (-1, 1)\)—that is, I presume that utility interdependence is less important than a person’s own utility.

**Ideological Affinity for the Task**

In a standard economic production environment, workers presumably have little ideological attachment to the task at hand or the products thereof; the political or social preferences of an assembly line worker are probably not particularly enhanced by the production of
cars. If moving from a generic economic production environment into a politically charged environment corresponds to an increase in the subjects’ $\alpha$ value, one can expect to see a change in effort levels corresponding to $\frac{\partial e^*}{\partial \alpha}$.

The details are worked out in the appendix, but according to this illustrative model a person’s own effort $e^*_i$ always increases in his/her ideological affinity for the task $\alpha_i$. His or her competitor’s effort $e^*_j$ increases in $\alpha_i$ as well, up to a cut point where $(1 + \frac{\delta_i(1-\delta_i)}{(1-\delta_j)} e_j \approx e_i$, and then decreases.\(^4\) That is, at first greater effort by $i$ spurs greater competition from $j$, but eventually $j$’s increasing costs cause him/her to lower effort in response to $i$’s increases in effort.

The appendix also illustrates what we expect if both affinities for the task change simultaneously and symmetrically; this change corresponds to the difference we would expect moving two workers from a neutral environment into a political bureaucracy where they have a political preference for or against the task. If both people increase their affinity for the task simultaneously by the same amount, then we expect $e^*_i$ and $e^*_j$ to rise. If $\alpha_i$ rises and $\alpha_j$ falls by the same amount—that is, if one person $i$ likes the task more and another person $j$ likes it less—we expect $e^*_i$ to rise and $e^*_j$ to fall.

This boils down to two stylized relationships:

**Stylized Relationship 1:** Increasing ideological affinity for the task ($\alpha$) increases effort under a competitive (tournament) incentive regime when ideological preferences are homogenous among competitors.

**Stylized Relationship 2:** When ideological preferences among competitors are heterogeneous, increasing the ideological importance of the task (greater positive affinity for one agent and greater negative affinity for the other) should increase the effort of the agent who likes the task and decrease the effort of the agent who dislikes it under a competitive (tournament) incentive regime.

\(^4\)The equality is not exact; see the appendix for details.
Social Preferences and Utility Interdependence

The body of experimental work about social preferences discussed earlier in the paper provides us with a reason to believe that people will act more cooperatively (that is, less competitively) with people they consider a part of their own group. As many of these experiments are about minimal groups—that is, groups that are artificially created and not organized around shared characteristics of substantive importance—we might expect such an effect to exist in any kind of organizational environment. But in an ideologically charged environment, we might expect this effect to be stronger still: competition within and among units takes place against those who are likely to identify strongly with one another.

As a result, I suspect that ideologically charged environments might correspond to high \( \delta \) values: agents with shared ideologies could identify strongly and positively with each others’ utility (\( \delta > 0 \)) and those with opposing ideologies may identify negatively (\( \delta \leq 0 \)). Essentially, people may feel worse about out-competing an ideological comrade than they would out-competing a neutral party (or an ideological opponent). Thus, in this illustrative model, a more ideologically charged production task will change effort levels by \( \frac{\partial e^*}{\partial \delta} \). Two stylized relationships come out of the model.

First, an agent’s own effort \( e^*_i \) is decreasing in \( \delta_i \) as long as \( C \) is sufficiently large.\(^5\) That is, as \( i \) cares more about \( j \)’s utility, \( i \) lowers his/her effort level to decrease the likelihood that \( j \) receives the penalty. Second, the other agent’s effort \( e^*_j \) is (for sufficiently large \( C \)) increasing in \( \delta_i \) up to the point where \( e^*_j = e^*_i \), at which point \( e^*_j \) is (for sufficiently large \( C \)) decreasing in \( \delta_i \). Colloquially, when \( e^*_i > e^*_j \), \( i \)’s working less hard creates an opportunity for \( j \) to work harder and decrease his/her probability of receiving the penalty. This continues until \( e^*_i = e^*_j \), at which point further declines in \( e_i \) necessitate less competition (and justify slackening effort) from \( e_j \). However, when \( e^*_i > e^*_j \), \( j \)’s increased effort is always outweighed by \( i \)’s decreased effort.

---

\(^5\)This is true as long as \( C \left( \frac{e_j}{e_i + e_j} \right) > F + \alpha_j \), which should hold for all but the lowest penalty probabilities given a reasonably large penalty \( C \). The same condition applies for the “sufficiently large \( C \)” statements that follow this one.
Stylized Relationship 3: Agents that have common group affinity ($\delta > 0$) will produce lower total levels of effort than those that have different group affinity ($\delta \leq 0$).

What happens if both $\delta_i$ and $\delta_j$ rise (or fall) by the same amount, akin from moving from a neutral environment into one with positive or negative social preferences? In the appendix, I show that under most conditions, a rise in $\delta_i$ and $\delta_j$ should cause a fall in $e^*_i$ and $e^*_j$: positive social preferences cause a mutual lowering of output because both agents want to lower the effort cost (and raise the probability of victory) for their fellow competitors. If we believe that social preferences are more important when ideology is the basis for group organization rather than some minimalistic criterion—that is, that ideology is a multiplier on $\delta$—then I can infer that increasing positive utility interdependence is associated with falling effort. Conversely, increasing negative utility interdependence will be associated with rising effort. Combining these two statements, in this model moving from neutral groups to ideologically-defined groups will make the heterogeneous pairs of agents exert more effort and the homogenous pairs exert less.

Stylized Relationship 4: If ideology activates group affinity, the gap in effort between competing agents that have common group affinity ($\delta > 0$) and those who have different group affinity ($\delta \leq 0$) will be larger in when groups are defined on the basis of ideology rather than on minimalistic criteria.

Repetition of the Stage Game

The above model assumes a one-shot competition, but the equilibrium and comparative statics still hold in any finite repetition of the stage game. Backward induction ensures the symmetry: any attempt to agree on a different strategy would break down in the final period of the game, propagating backward through to the beginning. Still, experimental

\[\text{As shown in the appendix, this prediction holds for all but the most extremely imbalanced settings, with dramatically larger } \delta_i > \delta_j \text{ and/or } \alpha_i > \alpha_j.\]
evidence has shown that some subjects seem to treat finitely repeated games as though they are infinitely repeated, particularly during the early stages of play: cooperation often breaks down as the end of the game approaches (Selten and Stoecker, 1986; Normann and Wallace, 2006).

The consequences of repeated competition are like those of repeating the Prisoner’s Dilemma. Like the PD, tournaments allow subjects to cooperate to their mutual benefit. Thus, just as in the PD, indefinitely repeating the tournament creates the potential for agents to collude with one another to lower their effort levels—in other words, to collude against the supervisor and subvert the incentives to produce. Thus, in repeated settings (or settings that actors treat as though they are repeated), we expect an across-the-board increase in propensity for collusion (mutually lowered effort); the relative predictions mentioned above are unaffected. This has been observed in previous experiments (??).

**Experiment: How does Ideology influence Resource Competition?**

I conduct an experiment that compares the behavior of subjects in the competitive environment described by the model under several settings. First, to determine whether social preferences are a factor, I compare the effort level of subjects who are members of the same group (with $\delta > 0$) to those who are members of different groups (with $\delta \leq 0$). Second, to determine whether ideological preferences for a task ($\alpha$) matter, I compare the effort level of subjects in a non-ideological production setting to those in an ideological setting. To determine whether social preferences are stronger in an ideological setting, I determine whether the gap between matched and unmatched group pairings is larger in the ideological setting than in the non-ideological setting. Finally, to see whether lowered output might be the result of mutual cooperation (as allowed by the Folk Theorem in an indefinitely repeated environment), I look for an upward trend in productivity toward the end of the experiment.
indicating a gradual breakdown in trust as the shadow of the future diminishes.

Experimental Design

In the experiment, 6 subjects are first broken into two groups, depending on the treatment. In the ideological treatment, subjects are asked about their general political ideology before any of the other aspects of the experiment are described; the three most conservative subjects are then placed into “the conservative group,” while the three most liberal subjects are placed into “the liberal group.” This treatment is designed to create a meaningful grouping that activates the subjects’ social preferences. In the non-ideological treatment, the survey question is not asked; instead, group assignment occurs randomly into either “Group A” or “Group B.” This minimal grouping treatment is designed to allow us to determine the power of ideologically meaningful groups relative to basic, essentially meaningless groupings.

Once assigned to a group, subjects are sorted into pairs with a counterpart. Two pairs contain members of the same group (like pairings), with one conservative/conservative or A/A pairing and one liberal/liberal or B/B pairing, while the third pair contains a mixed conservative/liberal or A/B pairing (unlike pairings). Combined with the ideological/non-ideological group treatment variation, the overall design of the experiment is thus a 2x2 block: like or unlike pairing, and ideological or non-ideological groups.

Each subject is given 10 tokens and 250 ECUs, or Economic Currency Units ($1.00 = 200 ECUs). Subjects may then choose whether to contribute tokens to one of two groups, or keep the tokens for themselves. Each token contributed to a group creates 2 ECUs for every member of that group, but costs the subject money: \( x \) tokens given to a group costs a subject \( x^2 \) ECUs.

The tournament aspect of the experiment is similar to the one described in Bull, Schotter, and Weigelt (1987, see also Schotter and Weigelt, 1992). Subjects are informed that a

\footnote{The instructions for this experiment are included as a reviewer appendix, but will be available as a web appendix after publication.}
penalty\textsuperscript{8} of 150 ECUs will be assessed to themselves or their counterpart, with the penalty assessed according to the number of tokens that they and their counterpart give to one of the two groups, which I call the target group.\textsuperscript{9} The probability of a subject receiving the penalty is given by the cumulative distribution function of the uniform distribution: if subject \(X\) gives \(x\) tokens to the target group while subject \(Y\) gives \(y\) tokens, then subject \(X\)’s probability of receiving the penalty is:

\[
\text{Pr}(X \text{ gets penalty}) = \frac{y}{x + y}
\]

The penalty selection process is explained to the subjects as having the computer simulate placing all contributed tokens into a bag, then randomly drawing one token from the bag. If the token drawn is one’s own, the counterpart receives the penalty; if the token drawn is the counterpart’s, than oneself receives the penalty. If \(x + y = 0\), as explained to the subjects, the penalty is determined by a simulated coin flip. Note that, by design, subjects cannot lose money in the experiment: the lowest possible payoff is 0 ECUs.

The principal-agent game embedded in this experiment is between the experimenter (or the computer program), who represents the collective interest of the target group, and the subjects. Subjects have the choice between contributing to the target group (at a cost to themselves) and avoiding the penalty, shirking by contributing to neither group (and avoiding the cost of contribution), or subverting the principal’s wishes by contributing to the non-target group. These actions correspond to the actions of working, shirking, and sabotage described in Brehm and Gates (Brehm and Gates, 1999): contributing to the target group is productive effort toward the principal’s goal, withholding contributions is diverting effort toward personal self-interest, and contributing to the non-target group is productive effort

\textsuperscript{8}As also noted above, this experiment uses penalties rather than rewards because in applied situations governments may find it easier to cut budgets than add to them and because, under risk neutrality, the predictions are the same for either model.

\textsuperscript{9}Subjects were not given the label of “target group”, but merely told that “a penalty of 150 ECUs will be assessed to you or your counterpart. This penalty will be assessed on the basis of the number of tokens that you and your counterpart gave to the <name of target> group.”
toward a collective goal that is not desired by the principal.\textsuperscript{10}

The experiment lasts 20 periods. For the first set of 10 periods, the target group remains the same. The second set of 10 periods switches the target group to the other group. A subject’s pairing with the counterpart remains the same throughout all these periods. Hence, there are potential ordering effects in the experiment depending on which of the two groups was the target group first. To counter potential ordering effects, I systematically vary the ordering of the target group, with half of the sessions using the conservative group/Group A as the first target group and the other half using the liberal group/Group B as the first target group.\textsuperscript{11}

**Data and Statistical Analysis**

The experiments were conducted in the xsfs laboratory for experimental social science at Florida State University, using 102 undergraduate students as subjects (54 in the ideological treatment, 48 in the non-ideological treatment). The laboratory provides a large number of private, networked computing carrels that allow subjects to make their decisions anonymously. Each session of the experiment was run using 12 subjects in front of private computer terminals, with the 12 subjects randomly assigned to two subsessions of 6 subjects simultaneously taking place in the same room.\textsuperscript{12} The experiments were programmed using zTree (Fischbacher, 2007), a software toolbox for economic experiments. All subjects received a $10 show-up fee in addition to what they earned in the experiment; subjects earned a total of $27.13 on average.

\textsuperscript{10}In the formal theory described above, sabotage effort could be characterized as a negative effort value $e_i$. I, however, take a slightly different approach: sabotage is not directly destructive of the supervisor’s desired policy effort totals, but instead works toward an opposed policy which the supervisor does not desire. Hence, the supervisor counts this sabotage effort as equivalent to $e_i = 0$.

\textsuperscript{11}At the conclusion of the experiment, a risk-assessment task and demographic survey were administered; these treatments do not enter into the present analysis and took place strictly after the treatments analyzed here. Subjects did not know about these future elements of the experiment when participating in the tournament.

\textsuperscript{12}One additional session of 6 subjects was run in the ideological treatment as a pilot session, and is included in these results.
The vast majority (92%) of subject contributions in the data set are made to the target group. In later rounds (period 15 and later), this percentage rises to 95.7%. This outcome is not entirely surprising, as there is little financial incentive to contribute to the non-target group. Given these facts, analysis will be focused on contributions to the target group.

Experimental methodology can eliminate many of the threats to inference that typically exist in a panel data set like this one, but a few remain to be dealt with. The censoring of contributions between the boundaries of 0 and 10 tokens can be explicitly modeled using tobit regression, which separately models the process of (i) being censored and (ii) the observed value of the dependent variable conditional on not being censored. The problem of unit heterogeneity (variation specific to the nature of individual subject) can be handled using a random effects approach, especially appropriate for this setting as zero correlation between the regressors and the unit effect is guaranteed by random assignment. In addition, there may be time trends in the data (as anticipated by the collusion argument above) and an effect of being in the first sequence of competitions (where a target group is selected with no expectation of a later switch in the target group for a second sequence) that should be removed from the data to make the observation errors completely exchangeable.

To ensure that none of my results are confounded by these problems, I use a random-effects tobit model that corrects for censoring and individual heterogeneity. Dummy variables capture the average token contribution levels for each of the four possible treatment combinations, and the coefficients on these dummies are manipulated as required to conduct the necessary tests for statistical significance. In addition, I include two controls to handle any trend and ordering effects present in the data. A logged period variable models the trend present in the data. A dummy variable for the first 10 rounds of play handles any constant ordering effect between the first and second sequences of competition.

13In particular, random assignment of the key regressors ensures that omitted variable bias is not an issue.
Stylized Relationships 1 and 2: Ideological Task Preference

According to the stylized relationships, those whose preferences are congruent with the task should contribute more than non-members, and the gap should be wider in the ideological treatment than in the non-ideological treatment. Table 1 shows the results of a random effects tobit model of contributions; note that the coefficient for each treatment condition indicates the predicted number of tokens contributed on average by subjects in that treatment. At first glance, we see these stylized relationships in the data: target group members contribute more than non-target group members, and the difference is larger in the ideological treatment. However, further tests are needed to determine whether these differences are statistically significant.

Figure 1 shows the estimated difference in contribution levels attributable to ideology from the results of Table 1 for all subjects. The first column in the figure shows that target group members contribute about 1.82 more tokens than non-target members in the ideological treatment. The second column shows that target group members contribute about .67 more tokens (on average) than non-target group members. As the figure shows, both of these differences are statistically significant. There may be some small group-affiliation
benefit from contributing to the target task in the minimal group treatment, but it is more likely that target group member contributions are higher in this treatment just because target group members have their contributions partially subsidized by benefitting from the total pool of group contributions. Thus, the really interesting difference to examine is whether the ideological treatment creates wider contribution differences than the minimal group treatment, since both have identical extrinsic incentives. Column three in the figure depicts this comparison:

\[
\left( \hat{\beta}_{\text{target}} - \hat{\beta}_{\text{non-target}} \right)_{\text{ideological}} - \left( \hat{\beta}_{\text{target}} - \hat{\beta}_{\text{non-target}} \right)_{\text{minimal}}
\]

Column three shows that the gap between contributions of target and non-target group members is 1.15 tokens wider in the ideological treatment than in the non-ideological treatment, a difference that is highly statistically significant \((p < 0.001, \text{two-tailed})\).\(^{14}\)

To summarize:

- Those who had a reason to ideologically prefer the task tend to contribute more toward it, compared to those who might have an ideological distaste for the task, and
- preference for the task has a stronger relationship with contributions when the groups and tasks are ideologically defined, compared to the minimal group case.

**Stylized Relationships 3 and 4: Social Preferences**

The formal model illustrates that competitors who share a group affiliation produce less than those who do not, and that this gap gets larger in the ideological treatment compared to the minimal group treatment. The results of the appropriately specified Tobit model are given in Table 2. Again, the stylized relationships are in the data: there is a positive gap in contributions between matched and unmatched pairings in both the minimal and ideological

\(^{14}\)I also calculated the same statistics separately for subjects who were matched with people in the same group and for those who were matched with people in a different group. The tobit regressions supporting these additional results are unreported, but substantially similar to the results of Table 1.
Figure 1: Increase in Contributions Attributable to Ideological Preference for the Task ($\alpha$), with 95% Confidence Intervals*

*The first two columns depict the difference in contributions between target and non-target group members. Positive values indicate that ideological groups elicited larger contributions from target members compared to non-target members. The "ID - Minimal" column depicts the difference between these the first column’s gap and the second column’s gap. Estimates taken from the model in Table 1.
Table 2: RE Tobit of Contributions by Pairing Structure

<table>
<thead>
<tr>
<th>variable</th>
<th>coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched Pairings, Minimal Treatment</td>
<td>5.43</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Unmatched Pairings, Minimal Treatment</td>
<td>5.86</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Matched Pairings, Ideological Treatment</td>
<td>4.68</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Unmatched Pairings, Ideological Treatment</td>
<td>7.09</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Controls</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>first sequence</td>
<td>-1.92</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ln periods 1-10</td>
<td>1.31</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ln periods 11-20</td>
<td>0.430</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Results derived from a random-effects Tobit model estimated in Stata 11.0; two-tailed p-values are reported. N=102, T=20, NT=2040. 181 observations were left-censored, 277 were right-censored.

treatment conditions, and this gap is larger for the ideological groups than for the minimal groups. Are these relationships statistically significant?

The predicted difference in contributions between matched and unmatched pairings from the Tobit model, both for minimal groups and for ideological groups, is shown in Figure 2. The quantities shown in the first two columns of the figure are equal to:

\[
\left( \hat{\beta}_{\text{matched pair}} - \hat{\beta}_{\text{unmatched pair}} \right)_{\text{treatment}}
\]

In the ideological treatment, those competing against people who shared their ideology tended to compete less hard compared to those competing against people with a different ideology, to the tune of 2.41 fewer tokens. In the minimal group treatment, the same gap exists but it is not as large (matched pairings contributed only .427 tokens less on average) and is statistically indistinguishable from zero. Thus, utility interdependence is as described in the stylized model, but minimal group affiliation only weakly activates this effect of social preferences on behavior.

As one might expect from Figure 2, ideology strongly activates group antipathy (or

\[15\]The quantity and its standard error are calculated using lincom in Stata 11.0.
Figure 2: Decrease in Contribution Levels Attributable to Change in Utility Interdependence (\( \delta \)), with 95% Confidence Intervals*

*The first two columns depict the contributions gap between matched (both competitors from same group) and unmatched (one competitor from each group) pairings. Negative values indicate that the matched pair contributed less per person than the unmatched pair. The "ID - Minimal" column depicts the difference between these the first column’s gap and the second column’s gap. Estimates taken from the model in Table 2.
solidarity). To see this, I calculated:

\[
\left( \hat{\beta}_{\text{matched}} - \hat{\beta}_{\text{unmatched}} \right)_{\text{ideological}} - \left( \hat{\beta}_{\text{matched}} - \hat{\beta}_{\text{unmatched}} \right)_{\text{minimal}}
\]

This quantity is shown in the third column of Figure 2. Each of the terms in parentheses measures how much harder unmatched groups competed against each other compared to matched groups; the difference between the two is the effect of ideological group formation on competitiveness (compared to minimal groups). The calculated quantity is $-1.99$, indicating that the ideological treatment caused the contributions gap between matched and unmatched pairings of competitors to widen by 2 tokens when compared to the non-ideological treatment. This quantity is on the border of statistical significance ($p = 0.055$, two-tailed).

To summarize, forming groups on the basis of ideology seems to make social preferences much more consequential:

- heterogenous group pairings tend to contribute more than same group pairings in the ideological treatment, and
- the contribution gap between heterogeneous and same group pairings is larger in the ideological treatment than in the minimal group treatment.

**Collusion and the Breakdown of trust**

The evidence of Tables 1 and 2—particularly the statistically significant trend effects in both the first and second sequences of play—suggests that common group affiliation engenders trust between subjects, allowing them to collude against the experimenter and mutually lower their effort levels. Both regressions reveal that contributions tended to rise over the course of play, as would be consistent with a cooperative agreement to lower output for mutual benefit that gradually broke down as the shadow of the future shortened toward the end of the game.

The primary rival explanation for this finding is simple learning: subjects gradually
competed harder as they became familiar with their opponent and the environment. This explains the need to separately examine trends in the first and second sequences of play; by the second sequence of play, the effect of learning should be substantially smaller as subjects are already familiar with the environment.

Although the trend in the second set of periods is smaller than that from the first, suggesting that some of the increase in the first ten periods is explained by subjects learning how to compete, there is still a detectable upward trend in periods 11-20 that, for like pairs, corresponds to about 1.4 more tokens contributed in period 20 compared to period 11. Thus, even accounting for the existence of a learning effect, there seems to be an upward trend in effort that would be consistent with a collusion agreement that broke down as the end of the game approached.

Discussion

This paper presents an experimental finding that might have broader implications for future research into using resource competition inside of bureaucracies. The experiment indicates that adding a relatively narrow, innocuous ideological charge to the production task in a tournament competition game might have substantively relevant impacts on behavior in that game. Specifically, imposing an ideological valence on the task (and, by implication, activating the ideological preferences of the subjects in the experiment) can either weaken or strengthen the incentives of the tournament. Ideological congruence with the task is associated with increased contribution toward that task, and incongruence weakens contributions. However, there are also social preferences that are activated by ideology: subjects compete harder against their ideological opponents compared with their ideological comrades.

The analogy from this experiment to the live environment of a bureaucracy is far from direct. However, several avenues of potentially fruitful research are suggested by this research. First, it is worth determining whether individual bureaucrats, agency subunits, and whole
agencies that do not share a common group affiliation or ideology (including, but not limited to, political ideology in a conservative-liberal sense) will compete harder for resources than more ideologically homogenous units. For example, units with similar responsibilities but different loyalties or policy commitments might compete harder for funding awarded according to the success of their respective programs when compared to units that share common organizational loyalties. When the Defense Department runs programs with broadly similar goals (missile defense, nuclear deterrence, etc.) that compete for funding, it’s worth determining whether competition for resources organized between branches of service (Navy, Air Force, and Army) is fiercer than competition for resources inside any one of these services. In this case, the different cultures, warfighting doctrines and sociological structures of the three branches (and not conservative-liberal differences) might provide the ideological tension necessary. This sort of competition is harder to imagine inside of the Environmental Protection Agency, where employees have a common ideological commitment to the environment and all share a group affiliation with the department. It would be worth determining whether resource competition inside of the EPA was less effective at increasing the efficiency or policy compliance (with the executive branch) of the agency.

My findings might also explain a conclusion from some prior work: in apparent contradiction of the ally principle (Bendor and Meirowitz, 2004), ideological disagreement with the executive has been observed not to worsen the principal-agent problem when it comes to low- and mid-level work that is ideologically charged. Multiple case studies of the bureaucracy have shown that bureaucrats are, by and large, loyal to the goals of the president regardless of ideological disagreement (Rosen, 1989, Chapter 9; Wilson, 1991, pp. 274-275; Brehm and Gates, 1999; Golden, 2000). My work suggests two potential explanations: (1) the effect of internal competition among ideologically heterogeneous bureaucrats for assignments and promotions overwhelms the effect of ideological taste, and (2) the lowered output by some who have a distaste for a particular principal tends to be compensated for by increased output by those who agree with that principal, and if evenly balanced then changes in party
control over the executive branch will have little effect on bureaucratic performance. If confirmed in future work, this implication of the experiment could significantly influence our understanding of bureaucratic behavior.

Last and not least, I believe that my findings add a quantum of support for the merit-based civil service system. An agency that hires on the basis of merit and without regard to ideological conviction might, on the basis of this experiment’s empirical findings, be easier to manage via internal competition: the resulting workforce will be more heterogeneous and therefore more competitive. If confirmed in field data and future experiments, that is good news for democratic governance in the context of a professionalized administration.

References


**Mathematical Appendix: Comparative Statics on Ideological Affinity and Concern for Others’ Utility**

In this appendix, I will use the implicit function theorem to derive predictions about how a subject’s effort will change as (1) his/her affinity for the task, (2) the other’s affinity for the task, (3) his/her own concern for the other’s utility, and (4) the other’s concern for his/her own utility changes. As a part of the process, I will also demonstrate that utility functions for each actor, $u_i$ and $u_j$, are concave in the actors’ own effort choices.
As noted in the body of the paper, the uniqueness of the best response solutions implies that any action profile \( e^* = [e_i^*, e_j^*] \) that simultaneously solves the first order conditions must be a Nash equilibrium:

\[
\frac{\partial u_i}{\partial e_i} = 0 \quad \frac{\partial u_j}{\partial e_j} = 0
\]

Generally speaking, I will stack the first order conditions of an optimization problem into a vector called \( u' \). According to the implicit function theorem, we can write an implicit function \( \phi = [\phi_i(\gamma), \phi_j(\gamma)] \) such that \( \phi_i(\gamma) = e_i^* \), \( \phi_j(\gamma) = e_j^* \), and conclude that:

\[
\frac{\partial e^*}{\partial \gamma} = \frac{\partial \phi}{\partial \gamma} = -[D_{e'u'}]^{-1} D_{\gamma} u'
\]

The following sections apply this methodology to obtain comparative statics on the parameters of interest.

**Ideological Affinity**

By the implicit function theorem,

\[
\frac{\partial e^*}{\partial \alpha_i} = -[D_{e'u'}]^{-1} D_{\alpha_i} u'
\]

Considering the first term:

\[
D_{e'u'} = \begin{bmatrix}
\frac{\partial^2 u_i}{\partial e_i^2} & \frac{\partial^2 u_i}{\partial e_i \partial e_j} \\
\frac{\partial^2 u_j}{\partial e_i \partial e_j} & \frac{\partial^2 u_j}{\partial e_j^2}
\end{bmatrix} = \begin{bmatrix}
a & b \\
c & d
\end{bmatrix}
\]

We need to compute and sign each of these conditions, then take the inverse to determine

---

\[16\] I include conditions for the case where both subjects are members of the target group, but the results are no different for other situations: the only difference between these cases is a trailing \(-2\) term on terms
the sign of the components of the inverse. The conditions\textsuperscript{17} are:

\[
\begin{align*}
a &= \frac{\partial^2 u_i}{\partial e_i^2} = -C(1 - \delta_i) \frac{2e_j}{(e_i + e_j)^3} - 2 \\
b &= \frac{\partial^2 u_i}{\partial e_i \partial e_j} = C(1 - \delta_i) \left( \frac{e_i - e_j}{(e_i + e_j)^3} \right) \\
c &= \frac{\partial^2 u_j}{\partial e_i \partial e_j} = C(1 - \delta_j) \left( \frac{e_j - e_i}{(e_i + e_j)^3} \right) \\
d &= \frac{\partial^2 u_j}{\partial e_j^2} = -C(1 - \delta_j) \frac{2e_i}{(e_i + e_j)^3} - 2
\end{align*}
\]

The \( a \) and \( d \) conditions are negative as long as \( \delta < 1 \)—that is, as long as one’s own welfare is valued more than the competitor’s. This demonstrates that \( u_i \) is concave in \( e_i \) and likewise for \( u_j \). The sign of the cross-derivatives depends on the relative size of \( e_j \) and \( e_i \): one will be positive or negative depending on which is larger (and, if equal, they will both be zero).

We need the inverse of the matrix whose terms we have just calculated. To compute an inverse of a 2x2 matrix, we can use a simple rule:

\[
\begin{bmatrix}
a & b \\
c & d
\end{bmatrix}^{-1} = 
\begin{bmatrix}
d & -b \\
-c & a
\end{bmatrix}
\]

Now, consider the second term:

\[
D_{\alpha_i} u' = \begin{bmatrix}
\frac{\partial^2 u_i}{\partial \alpha_i \partial e_i} \\
\frac{\partial^2 u_j}{\partial \alpha_i \partial e_j}
\end{bmatrix} = \begin{bmatrix}
1 \\
\delta_j
\end{bmatrix}
\]

Because of the simple structure of the second term, the final derivative reduces to a simple

\textsuperscript{17}NOTE TO REVIEWERS: I have numerically verified that each of the following derivations are correct. The verification is in the included file proof\_numerical\_verification\_2-19-2011.r. This file will not be included with the final draft, but I included it so that you may verify that these conditions are correct, if you wish.
the epsilon is necessary because the \( \alpha \) is a cut point

\[
\frac{\partial e^*}{\partial \alpha_i} = \begin{bmatrix}
\frac{\partial e^*_i}{\partial \alpha_i} \\
\frac{\partial e^*_j}{\partial \alpha_i}
\end{bmatrix} = -\begin{bmatrix}
d\delta_{ji}b \\
\frac{d\delta_{ji}}{ad-bc}
\end{bmatrix}
\]

We know that \( ad > 0 \) and \( bc < 0 \), and hence the denominator is positive. We can compute the numerator:

\[
d - \delta_j b = -\frac{C}{(e_i + e_j)^3} \left[ (2(1 - \delta_j) - \delta_j(1 - \delta_i)) e_i + \delta_j(1 - \delta_i) e_j \right] - 2
\]

This term is strictly negative as long as \( 2(1 - \delta_j) - \delta_j(1 - \delta_i) > 0 \), or \( \frac{2}{3 - \delta_j} \geq \delta_i \). If we limit \( \delta_j \) to a reasonable range, say \( \delta_j \in [-1, 1] \), this restriction corresponds to \( \delta_i \leq \frac{1}{2} \). As long as this condition holds, then \( \frac{\partial e^*_i}{\partial \alpha_i} > 0 \).

\( \frac{\partial e^*_i}{\partial \alpha_i} \) depends on the relative size of \( e_i \) and \( e_j \):

\[
\delta_j a - c = \frac{C(1 - \delta_j)}{(e_i + e_j)^3} \left[ e_i - \left( 1 + \frac{\delta_j(1 - \delta_i)}{1 - \delta_j} \right) e_j \right] - 2\delta_j
\]

Thus, \( \frac{\partial e^*_i}{\partial \alpha_i} > 0 \) if \( e_i - \left( 1 + \frac{\delta_j(1 - \delta_i)}{1 - \delta_j} \right) e_j < \epsilon \approx 0 \), and \( \frac{\partial e^*_i}{\partial \alpha_i} < 0 \) if \( e_i - \left( 1 + \frac{\delta_j(1 - \delta_i)}{1 - \delta_j} \right) e_j > \epsilon \approx 0 \); the epsilon is necessary because the \(-2\delta_j\) term moves the threshold for \( e_j \) slightly larger (if \( \delta_j > 0 \)) or smaller (if \( \delta_j < 0 \)) than it otherwise would be.

If \( \frac{C(1 - \delta_i)}{(e_i + e_j)} \left[ e_i - \left( 1 + \frac{\delta_j(1 - \delta_i)}{1 - \delta_j} \right) e_j \right] = 2\delta_j \), \( \frac{\partial e^*_i}{\partial \alpha_i} = 0 \). From this fact, we can infer that there is a cut point \( \alpha_i \) above which \( \frac{\partial e^*_i}{\partial \alpha_i} < 0 \) and below which \( \frac{\partial e^*_i}{\partial \alpha_i} > 0 \): when \( e_i = e_j \), \( \frac{\partial e^*_i}{\partial \alpha_i} \) is still positive and thus an increase in \( \alpha_i \) will cause \( e_i \) to grow larger than \( e_j \), moving into the regime where \( \frac{\partial e^*_i}{\partial \alpha_i} < 0 \).

A symmetric proof, omitted for reasons of space, can be given for \( \frac{\partial e^*_j}{\partial \alpha_j} \); the proof consists primarily of switching subscripts around in the proof above.

What happens if both \( \alpha_i \) and \( \alpha_j \) change at once (by the same amount)? We can approximate the total derivative using \( \frac{\partial e^*_i}{\partial \alpha_i} + \frac{\partial e^*_j}{\partial \alpha_j} \), which for \( e_i \) is:

\[
\frac{C(1 - \delta_i)}{(e_i + e_j)^3} \left[ \left( \frac{2(1 - \delta_j)}{1 - \delta_i} - \delta_j \right) e_i + \delta_j e_j \right] + 2 + \frac{C(1 - \delta_i)}{(e_i + e_j)^3} \left[ \left( 1 + \frac{\delta_i(1 - \delta_j)}{1 - \delta_i} \right) e_i - e_j \right] + 2\delta_i
\]
As long as $\delta_i$ and $\delta_j$ are both greater than 1, this term is strictly positive: if both $\alpha_i$ and $\alpha_j$ rise, then $e_i^*$ should rise. If $\alpha_i$ rises and $\alpha_j$ falls, then $e_i^*$ will rise as long as:

$$\frac{C(1 - \delta_i)}{(e_i + e_j)^3} \left[ \frac{2(1 - \delta_j)}{(1 - \delta_i)} - \delta_j \right] e_i + \delta_j e_j + 2 > \frac{C(1 - \delta_i)}{(e_i + e_j)^3} \left[ (1 + \frac{\delta_i(1 - \delta_j)}{(1 - \delta_i)}) e_i - e_j \right] + 2\delta_i$$

or, slightly rearranged,

$$\frac{C(1 - \delta_i)}{(e_i + e_j)^3} \left[ \frac{2(1 - \delta_j)}{(1 - \delta_i)} - \delta_j \right] e_i + \delta_j e_j - \left[ (1 + \frac{\delta_i(1 - \delta_j)}{(1 - \delta_i)}) e_i - e_j \right] > 2\delta_i - 2$$

This is always true for $\delta_i < 1$.

**Concern for Others’ Utility**

The process is the same as above, except for the second term in the IFT-derived expression:

$$\frac{\partial e^*}{\partial \delta_i} = -[D_{e^*}u']^{-1} D_{\delta_i}u'$$

The 2x2 inverse matrix of derivatives is the same as above. The 2x1 matrix of derivatives with respect to $\delta_i$ is simple:

$$D_{\delta_i}u = \begin{bmatrix} \frac{\partial^2 u_i}{\partial e_i \partial \delta_i} \\ \frac{\partial^2 u_j}{\partial e_j \partial \delta_i} \end{bmatrix} = \begin{bmatrix} -C \frac{e_j}{(e_i + e_j)^2} + F + \alpha_j \\ 0 \end{bmatrix}$$

With these terms in hand, deriving the predictions is straightforward:

$$\frac{\partial e^*}{\partial \delta_i} = \begin{bmatrix} \frac{\partial e_i^*}{\partial \delta_i} \\ \frac{\partial e_j^*}{\partial \delta_i} \end{bmatrix} = -\begin{bmatrix} \left( \frac{d}{ad - bc} \right) \left( -C \frac{e_j}{(e_i + e_j)^2} + F + \alpha_j \right) \\ \left( \frac{c}{ad - bc} \right) \left( -C \frac{e_i}{(e_i + e_j)^2} + F + \alpha_j \right) \end{bmatrix}$$

Presume that $C \frac{e_j}{(e_i + e_j)^2} > F + \alpha_j$ (that is, $i$’s probability of receiving the penalty is sufficiently high). We know that $d < 0$ as long as $\delta_j < 1$. Then $\frac{\partial e_i^*}{\partial \delta_i} < 0$. We also know that the sign of $c$
depends on the relative values of $e_j$ and $e_i$: if $e_j > e_i$, then $\frac{\partial e_j^*}{\partial \delta_i} < 0$; if $e_j < e_i$, then $\frac{\partial e_j^*}{\partial \delta_i} > 0$. Because $|d| > |c|$ when $e_j < e_i$, the effect of an increase in $\delta_i$ on combined effort $e_i^* + e_j^*$ is always negative. Furthermore, we know that there must be a cut point $\delta_i$ that divides these two regimes: when $e_j = e_i$, then $\frac{\partial e_j^*}{\partial \delta_i} = 0$ but $\frac{\partial e_i^*}{\partial \delta_i}$ is still negative; hence, further increases in $\delta_i$ will move into the regime where $e_j > e_i$ and $\frac{\partial e_j^*}{\partial \delta_i} < 0$.

A symmetric proof, omitted for reasons of space, can be given for $\frac{\partial e_i^*}{\partial \delta_j}$; the proof consists primarily of switching subscripts around in the proof above.

What happens when both $\delta$ values rise at the same time (by the same amount)? We can predict this using $\frac{\partial e_i^*}{\partial \delta_i} + \frac{\partial e_j^*}{\partial \delta_j}$, which in this case will be:

$$
\left( \frac{d}{ad - bc} \right) \left( C \frac{e_j}{(e_i + e_j)^2} + F + \alpha_j \right) - \left( \frac{b}{ad - bc} \right) \left( C \frac{e_i}{(e_i + e_j)^2} + F + \alpha_i \right)
$$

The first term is strictly less than zero because $d < 0$:

$$
d = -C \left(1 - \delta_j\right) \frac{2e_i}{(e_i + e_j)^3} - 2
$$

The second term’s sign depends on the value of $b$:

$$
b = C \left(1 - \delta_i\right) \left( \frac{e_i - e_j}{(e_i + e_j)^3} \right)
$$

In order for the total derivative to be negative, we require:

$$
\left( C \left( 1 - \delta_j \right) \frac{2e_i}{(e_i + e_j)^3} + 2 \right) \left( C \frac{e_j}{(e_i + e_j)^2} + F + \alpha_j \right) > \left( C \left( 1 - \delta_i \right) \left( \frac{e_i - e_j}{(e_i + e_j)^3} \right) \right) \left( C \frac{e_i}{(e_i + e_j)^2} + F + \alpha_i \right)
$$

The condition above will always hold when the following, weaker condition holds:
\[
2e_i (1 - \delta_j) \left( C \frac{e_j}{(e_i + e_j)^2} + F + \alpha_j \right) > (1 - \delta_i) (e_i - e_j) \left( C \frac{e_i}{(e_i + e_j)^2} + F + \alpha_i \right)
\]

Slightly simplifying terms:

\[
C (1 - \delta_j) \frac{2e_i e_j}{(e_i + e_j)^2} + 2e_i (1 - \delta_j) (F + \alpha_j) > C (1 - \delta_i) \frac{e_i (e_i - e_j)}{(e_i + e_j)^2} + (e_i - e_j) (1 - \delta_i) (F + \alpha_i)
\]

We can see that this condition will hold unless \( e_i \) is much larger than \( e_j \), presuming that \( \delta_i \approx \delta_j \) and \( \alpha_j \approx \alpha_i \). The first term on the left hand side will be larger than the first term on the right as long as
\[
2e_i e_j > e_i (e_i - e_j), \quad \text{or} \quad 3e_j > e_i.
\]
The second term on the left hand side will be bigger than the second term on the right so long as
\[
2e_i > (e_i - e_j), \quad \text{or} \quad e_i > -e_j \quad \text{(true by definition)}.
\]
Thus, for all but the most extremely imbalanced settings—with dramatically larger \( \delta_i > \delta_j \) and/or \( \alpha_i > \alpha_j \)—if both actors’ utility interdependence \( \delta \) rises, I expect output \( e_i \) to fall (and likewise for \( e_j \)).