Specific and General Information Sharing
Among Competing Academic Researchers

Carolin Haeussler[a]

Lin Jiang[b]

Jerry Thursby[c]

Marie Thursby[c,d,e]

a. University of Passau, Faculty of Economics and Business, Innstr. 27, D-94032 Passau, Germany
b. University of Missouri, Cornell Hall, Columbia MO, 65203
c. Georgia Institute of Technology, College of Business, 800 W. Peachtree St. NW, Atlanta GA 30308
d. National Bureau of Economic Research, Cambridge MA, 02138
e. Corresponding author: Marie.thursby@scheller.gatech.edu; tel: 404 894 5249; fax: 404 385 4894

Abstract

We examine information sharing among academics during the research process and show it is context dependent because of differences in trade-offs. When researchers respond to specific requests for information or materials, potential future reciprocity is weighed against the current loss of competitiveness, while general sharing intermediate results in an open forum is driven by the need for feedback versus potential misappropriation. We formally model these trade-offs and empirically test for differences using a survey of German and UK bio-scientists. Increased competition has a negative impact on sharing in both contexts. But career stage has an effect only on specific sharing with untenured faculty less likely to share. Further, scientists in larger teams are more likely to share specifically, but less likely to share generally. The importance of patents for one's reputation reduces sharing in both contexts, but the effect is greater for general information sharing.

Keywords: Information sharing, open science, scientific competition, knowledge diffusion, misappropriation.

1   Introduction

Information sharing is critical to scientific progress, so much so that the Mertonian norm of unconditional sharing of knowledge is considered one of the defining features of academic life (Merton, 1973). In principle, this norm is enforced by a priority-based scientific reward system in which the first person to discover a result gets whatever "prize" is associated with discovery (Dasgupta and David, 1987; Stephan, 1996). There is a tension, however, between communal sharing and the competitive incentives for researchers during the research process itself (Hagstrom, 1965; 1974; Dasgupta and David, 1994; Murray and O'Mahony, 2007). This tension along with incentives created by the commercial potential of academic research has made restrictive sharing a common problem and heightened interest in information sharing among academic researchers (Blumenthal *et al*., 1996; Cohen and Walsh, 2008; Murray, 2010).

We examine, theoretically and empirically, what drives competing academic researchers to share information during the research process. We consider two important contexts: *specific sharing*, where researchers share information about their work privately in response to a request and *general sharing*, where a researcher shares new, unpublished results publicly. Common examples of the former include sharing data, materials (such as cell line or mice), and information about techniques or details not released publicly. General sharing includes presentations, webpostings, and working papers.

While the communal desire for information sharing is universal, particularly among public funding agencies (e.g., NSF, 2011; NIH, 2013), factors that encourage sharing in one context may not in another. In fact, we show that sharing can be highly context dependent because of differences in individual trade-offs and incentives. Our formal modeling allows us to characterize these trade-offs and set up an empirical test for differences in specific and general sharing.

In specific sharing, a researcher who shares in response to a request from another bears the cost of preparing materials or documentation, but also, by providing access to the information, the researcher increases the probability that someone else will solve her research problem first. There is a potential benefit, but only in the future, and only if the requesting scientist has information of value and reciprocates. Hence, sharing depends on the likelihood of future reciprocity and the level of competition. In general sharing, a researcher who shares new results can gain immediate feedback, particularly if the audience includes researchers who have solved complementary parts of the problem. However, as with specific sharing, she increases the probablity someone else will win the prize for solving the problem. General sharing also has the benefit of providing credit, but only if the work is acknowledged. Unknown researchers in the audience may use the information without acknowledgement unless they risk independent verification or have a strong belief in the norms of open science. Important factors of sharing in this context are the value of feedback, beliefs about acknowledgement, and the level of competition.

The empirical analysis uses a survey of bio-scientists in Germany and the United Kingdom which

provides individual level data on willingness to share in specific and general contexts.. The survey also includes information on the respondents' perceptions of the level of competition in their field and demographic information such as career stage, the extent to which their research is basic, and the size of their research team. Our econometric analysis suggests that, except for three regressors (most notably the level of competition), what drives researchers to share information differs markedly. For example, we find members of larger teams are more likely to engage in specific sharing but less likely to share generally. Intuitively, economies of scale in larger teams reduce the cost of sharing in the context of specific sharing. With regard to general sharing, researchers in larger labs have a larger built-in network of colleagues for immediate feedback so that general sharing is less valuable, *ceteris paribus*. Moreover, in specific sharing, we find tenured faculty are more likely to share than untenured; in general sharing, tenure has no effect. This is consistent with the importance of future gains from reciprocity in specific sharing since untenured faculty may not be in a position to share in future periods.

To our knowledge, this is the first study to formally model academic information sharing in these two contexts. Despite evidence that academics withhold information, there has been little theory to guide our understanding (Blumenthal *et al*., 1996; Campbell *et al*., 2002). With few exceptions, the theory has focused on firms (Von Hippel, 1987; Anton and Yao, 2002; 2004; Lerner and Tirole, 2002, Gächter *et al.,* 2010; Baker and Mezzetti, 2005; Hellmann and Perotti, 2011; Gans *et al*., 2011; Gill, 2008; Stein, 2008). The studies of academics focus on different issues, such as the trade-off between secrecy and publication (Mukherjee and Stern, 2009; Gans and Murray, 2010) or the impact of academic misconduct on research and publication decisions (Hoover, 2006; Lacetera and Zirulia, 2011). In contrast, we focus on sharing during the research process.

With respect to empirics, prior work has either aggregated information sharing across contexts or focused solely on specific sharing. For example, Blumenthal *et al.'s* (2006) study of genetics and other life sciences examines perceived competition and sharing, but it combines information across the contexts we consider into a single variable. Hong and Walsh (2009) and Walsh *et al.* (2007) examine specific sharing and find that cost, involvement in business, students' ability to publish, and scientific competition decrease sharing and Haeussler (2011) finds expected reciprocity and perceived adherence to the scientific norm of communalism to be important. While this literature compares the behavior of academic and industrial scientists, we are unaware of empirical evidence on sharing of the same individual in specific and general contexts.

Combined, our results have interesting implications for policy. One factor which operates the same way in both models is competition or the value of the prize. This suggests negative unintended consequences for competitions with prizes, as well as a cautionary message for funding agencies to the extent that their funding levels affect the nature of competition. The general sharing model has the

interesting implication that journal policies regarding misappropriation have a spillover to sharing during the research process itself. Misappropriation is considered to be a major problem in science (Bailey *et al.*, 2001; Enders and Hoover, 2004; Birnholtz, 2006; Couzin-Frankel and Grom, 2009). We join Lacetera and Zirulia (2011) in our message for more journal attention to these issues.

In the next section, we construct two simple models of specific and general sharing which highlight salient factors in the two context. Section 3 relates the propositions from the theory to empirical measures from the survey and presents the econometric analysis, as well as a discussion of empirical differences in the models. Section 4 provides clarifying remarks and implications for funding agencies and journals.

2 Games of Information Sharing

In developing the theory, it is important to keep in mind that the two types of sharing in question are not necessarily alternatives. A researcher may share some information privately (e.g., materials, techniques), while sharing other information (e.g., an entire working paper) publicly. Also the appropriate assumptions on timing and identity of the players who receive the information by specific and general sharing are quite different. Accordingly, we construct two models.

To capture the rivalry created by the scientific reward system, we model the situations as games in which researchers are trying to solve a common problem. A complete solution earns a prize, W, such as publication, a Fields Medal or Nobel Prize with academic value and/or it can have commercial value.

In both games, we assume that each researcher has solved a portion of the problem and/or developed materials of use in solving it. If a researcher shares her solution or materials, she makes it easier for the recipient(s) to earn the prize. The information could be materials (cell line, reagents), data or methods (software, lab technique), or intermediate research results.

2.1 Specific Sharing

Consider a game played in two periods. In the first, researcher 1 receives a request from researcher 2, and she decides whether or not to share. With probability α, the game continues to the second period and researcher 2 has information of value to researcher 1, in which case researcher 2 must decide whether or not to share. Each researcher's expected payoff is the expected value of the prize plus the asset value of any information or materials gained in exchange minus the cost of the exchange.

The game is "winner take all" so that each researcher gets W with probability less than one.[1] Unilateral sharing by a researcher lowers her/his probability of winning the prize and raises the rival's

---

[1] In this model W is fixed, but we have also examined a model with the value of the prize as a positive function of the material shared and the results of interest are qualitatively similar.

probability. In the absence of sharing, $x(r_1)$ and $1\text{-}x(r_1)$ represent the probability that researcher 1 and 2, respectively, wins the prize where $r_1 \geq 0$ is the amount or quality of researcher 1's information requested by researcher 2. Unilateral sharing by researcher 1 reduces her probability of winning by $\delta_1(e_2, r_1)$ and increases 2's probability by δ_1.[2] We assume that δ_1 is an increasing function of r_1 and 2's ability to exploit the information, $e_2 \geq 0$. The shared information has an asset value $V(e_2, r_1) \geq 0$ regardless of whether researcher 2 wins the prize. $V(e_2, r_1)$ reflects the extent to which the shared information reduces researcher 2's project cost and its value for use in other projects under the assumption that researcher 1 cannot credibly prevent such use. There is also a cost c_1 for researcher 1 to share the information (e.g., ship a lab mouse). We assume the costs are sufficiently low that all expected payoffs are positive, and that the discount factor is one.

  To determine equilibrium behavior, we first need to consider researcher 2's behavior should the game continue into the second period. For simplicity, we suppose the two researchers are again competing on a research problem with a prize W. In this period, it is researcher 2's information that is of value. We represent researcher 1 and 2's probabilities of winning by $1\text{-}\,y(r_2)$ and $y(r_2)$, respectively, where r_2 is the amount or quality of researcher 2's information requested by researcher 1. If researcher 2's first period request was denied, he will never share since he will raise researcher 1's probability of winning by δ_2 and bear c_2 without any gain. If, however, researcher 1 honored his request in the first period, his expected payoff is $(y - \delta_2)W + V(e_2, r_1) - c_2$ if he shares and $yW + V(e_2, r_1) - R_S(e_2, r_1, \eta)$ if he does not. We include the term R_S to reflect the common cultural norm that a person receiving a favor is expected to reciprocate. Thus, R_S is researcher 2's reputational loss if he does not share. It is natural to think of this loss as increasing in all arguments and critically dependent on η, a norm of reciprocity for favors within the community (i.e., $R_S(e_2, r_1, 0) = 0$). Thus researcher 2 will share if and only if

$$R_S > \overline{R_S} \equiv \delta_2(e_1, r_2)W + c_2. \tag{1}$$

  The first researcher makes her decision taking researcher 2's behavior and her expectation that the game will continue with researcher 2 having information of value, which we denote as α. It is easy to see that if $R_S < \overline{R_S}$, the first researcher will refuse to share because she knows researcher 2 will not share in the second period and the return to her generosity in the first period, α [(1-y)W] + (1- α)[(x- δ₁)W-c₁], is smaller than the return to not sharing, α [(1-y)W]+ (1- α)xW. On the other hand, if $R_S > \overline{R_S}$, she

---

[2] Absent competition, the researcher who shares the information would not necessarily reduce his/her own probability of winning. An example would be the case when the recipient uses the same techniques or information for a different problem. But here we focus on competing researchers. Also in a model of sharing among researchers with varying levels of competition, the competition parameter acts much like the *W* in our model.

knows that researcher 2 will share should the game continue, and her expected payoff is $\alpha[(1 - y + \delta_2)W + V(e_1, r_2)] + (1 - \alpha)[(x - \delta_1)W - c_1]$. This expression is greater than her expected payoff from not sharing, $\alpha[(1-y)W] + (1 - \alpha)xW$, for

$$\alpha > \frac{\delta_1(e_2, r_1)W + c_1}{\delta_1(e_2, r_1)W + \delta_2(e_1, r_2)W + V(e_1, r_2) + c_1} \equiv \bar{\alpha}. \tag{2}$$

Sharing by both researchers is a pure strategy subgame perfect equilibrium if and only if conditions (1) and (2) hold.

If either α or η equals zero, neither researcher shares in equilibrium. If $\alpha = 0$, researcher 1 either foresees no need for information from the researcher who requested the information or, alternatively, may not envision future opportunities for him to share. An example of the latter situation, could be if either researcher faces a tenure decision or is close to retirement. The need for positive η bears special mention since it is well known that cooperative strategies, such as sharing, can occur in repeated games without community norms. For example, if we were to model sharing as a repeated simultaneous move Prisoner's Dilemma game, sharing in the absence of such a norm could occur in equilibrium (Arribas and Urbano 2005; Mailath and Samuelson 2006).[3] However, we view the asynchronous move with a norm as more realistic. More to the point, the role of the theory is to derive testable hypotheses for the empirical analysis where the impact of norms on sharing is one of the items of interest.

Standard comparative statics on the equilibrium conditions, (1) and (2) yield the following proposition.

Proposition 1. *In the specific sharing game, the feasibility of sharing increases with (i) an increase in α and η, (ii) a decrease in c_1, c_2, or (iii) a decrease in W if $V(e_1, r_2) > \frac{\delta_2}{\delta_1} c_1$. The effects of e_1, e_2, r_1 and r_2 are ambiguous.*

The proposition has several implications for our empirical estimation. First, an increase in α, the researcher's belief that the game will continue beyond the current period, increases the likelihood of sharing. It is this result which allows us to relate α to respondent career stage in the survey . For example, researchers within their first six years in academia (i.e., untenured) are clearly at risk of exiting academia (or of exiting from an academic research position), as are those near retirement dates. Second, a decrease in the cost of sharing, c_i, decreases researcher i's single period gain from not sharing, thus increasing the likelihood she will share. As we discuss in Section 3.1, under an assumption of economies of scale in

---

[3] An appendix available from the authors derives comparative statics similar to those in Proposition 1 in the context of a repeated simultaneous move model where the period game is repeated an uncertain number of periods.

research teams, this result implies that researchers with larger teams face lower costs of preparing requested information and, thus, would be more likely to share, all else equal. Third, under the assumption stated in the proposition, a decrease in the size of the prize (W) increases the likelihood of sharing. A decrease in W decreases the single period gain to not sharing, but it also decreases the loss from future punishment for not sharing. The condition in the proposition ensures the former effect dominates. In the special case of $\delta_1 = \delta_2$, the condition simply says that the value of the material gained in the exchange exceeds the researcher's cost of supplying his own material. Fourth, in a community with high reputational cost or punishment for not reciprocating (corresponding to a high η), sharing is more likely to occur. The remaining results are ambiguous since any of these changes increase one researcher's incentive to share while decreasing the other's incentive.

2.2   General Sharing

In this section, we consider the decision to share research findings in an open forum before publication, for example, by presenting intermediate results at a conference or posting them on a website. Again researchers face conflicting incentives. Sharing such information allows the researcher to gain feedback, but as in specific sharing, if sharing provides useful insights for others it may increase the probability that someone else beats her to completely solving the problem. Unlike specific sharing, general sharing also has the benefit of announcing her progress to the public which will afford credit for that work, but only if others acknowledge it. Our interest is in the conditions under which preliminary work is shared with a general audience and appropriately acknowledged as an equilibrium outcome.

We consider a simplified sharing decision of a single researcher, researcher 1, who is deciding whether to share her results with the entire community in an effort to get credit σW for her progress. We assume the community includes M researchers trying to solve the same problem where M ≥ 2. None of the researchers has completely solved the problem; if any has, the game ends. As before, W is the prize for the solution. We also represent the portion of the problem researcher 1 has solved as $\sigma \in (0,1)$. For simplicity, we abstract from the cost of preparing for the presentation.[4]

Because in general sharing there is uncertainty about who will be in the audience, we let $\gamma \in (0,1)$ be the probability that a randomly chosen researcher has solved a different part of the problem, and call that researcher a complementor. Then $\lambda = 1 - (1 - \gamma)^{M-1}$ is the probability that at least one of the $M - 1$ researchers is a complementor. Sharing with complementors has two effects: it allows for

---

[4] Since she has to bear preparation costs for publication once she solves the entire puzzle, it is not clear there is a true opportunity cost associated with presenting preliminary results. This is in contrast to specific sharing where the researcher may share information that is not required to be shared for publication (e.g., data/material samples for shipment or instructions for database use). Including a cost to reflect travel or other expenses not related to publication would make presentation less likely all else equal.

feedback, which improves the presenting scientist's solution to the problem, adding value τ to W,[5] but it reduces her probability of winning the prize by δ. We denote researcher 1's probability of winning as $x \in (0,1)$ if she does not share (or shares and there are no complementors) and $(x - \delta) \in (0,1)$ if she shares and the audience includes at least one complementor. If she shares with an audience without any complementors, she gets neither feedback nor reduces her probability of winning.[6]

The game is represented in extensive form by Figure 1. In stage zero nature chooses the probability that there is a complementor, γ, and in stage one, researcher 1 chooses between sharing her results by presenting (P) to the community or not (NP). If she shares, she makes the $M - 1$ researchers aware of her progress. Researchers obtain information from researcher 1, and all (including researcher 1) continue working on the problem.

In stage two, nature decides which researcher is first to completely solve the problem. If the winner is researcher 1, the game ends. If the winner is another researcher, he decides whether to acknowledge 1's work (A) or not acknowledge it (NA).[7] If he acknowledges her work, he earns partial credit, $(1 - \sigma)W$, while he earns full credit, W, if he does not acknowledge it. But with probability v, one of $M - 2$ researchers, observing both the winner's work and researcher 1's, will verify that the winner has used researcher 1's idea without acknowledging it. In this case, the winner suffers a loss of reputation R_G and earns no credit. We denote researcher 1's belief that a randomly chosen researcher will provide verification as $\rho \in (0,1)$ and assume that the M researchers share this belief. Then we can write each researcher's belief that at least one of the $M - 2$ (other than researcher 1 and the winner) verifies as $v = 1 - (1 - \rho)^{M-2}$.

Consider the winner's decision. For acknowledgement to be worthwhile for the winner, he has to expect the likelihood that another researcher will verify the originality of his work to be sufficiently high.

---

[5] An alternative approach would be to allow feedback to increase the probability of solving the problem from x to *(1+τ)*. We chose the simpler approach since they both yield the same qualitative results. Note also that non-complementors may provide feedback as well. For simplicity, however, we normalize their feedback to zero since complementors' feedback is likely to be more pertinent. Our specification also treats feedback as received privately. For instance, a colleague may listen to the presentation by researcher 1 or see her paper posted on a web-site, and then provide feedback through email or one-on-one discussion. In that case, the feedback would not benefit others in the audience. Nevertheless, one may think that feedback can also be received in the public setting and thus known by all the players. Allowing feedback to benefit both researcher 1 and the other player, however, does not change our results.

[6] Even without complementors, researcher 1 does not necessarily win with certainty because she still faces competition from non-complementors. Although non-complementors presumably do not have more valuable information than researcher 1, they may still solve the problem first and win because scientific discovery is highly uncertain and there might be alternative research approaches.

[7] Thus we have implicitly assumed that researcher 1 cannot, herself, force those with whom she shares to acknowledge her work. This seems appropriate for work that is neither published nor patented. Even for results codified by publication or patent, one can argue that an external mechanism is involved.

The decision also depends on the reputational loss R_G if misappropriation is verified, as well as the size of the prize and the extent to which researcher 1 solved the problem; that is, acknowledgement is worthwhile for the winner if

$$v > \frac{\sigma W}{R_G + W}. \tag{3}$$

Since v is a function of the number of researchers working on the same problem, this condition can be rewritten as

$$M > \frac{\ln\left(1 - \frac{\sigma W}{R_G W}\right)}{\ln(1 - \rho)} + 2. \tag{4}$$

Thus, the model implies that if only two researchers are working on the problem, the winner will never acknowledge researcher 1's work. Since credit for her progress will not be forthcoming, the only reason that researcher 1 would share is for feedback. In fact, we will find that unless there is sufficient feedback, researcher 1 will not share.[8]

In making her sharing decision, researcher 1 compares her expected utility from sharing and not sharing

$$U^P - U^{NP} = (1 - x)\sigma W C + \lambda \tau + \lambda(\sigma C - 1)\delta W \tag{5}$$

where $C = \Pr(A) + v\Pr(NA)$ is the probability that she will receive credit regardless of whether or not the winner acknowledges,

$$\Pr(A) \equiv \begin{cases} 1 & \text{if (4) holds} \\ 0 & \text{otherwise,} \end{cases}$$

and $\Pr(NA) = 1 - \Pr(A)$. The first term on the right hand side of (5) is the announcement effect and reflects the credit she hopes to get from sharing. The second reflects feedback from complementors, while the third reflects the fact that their presence in the audience increases their chances of winning. Thus for sharing to dominate not sharing, the effects of announcement and feedback must outweigh the negative impact from competing with complementors.

Together the conditions in (4) and (5) allow us to completely characterize the potential pure strategy equilibria of the game in terms of $\bar{v} = \frac{\sigma W}{R_G + W}$ and $\bar{C} = \frac{\lambda[\delta W - \tau]}{[(1-x)+\lambda\delta]\sigma W}$. For given values of the parameters M, ρ, σ, W, and R, the unique equilbrium involves presentation and acknowledgement for $v > \bar{v}$ and $C > \bar{C}$; presentation without acknowledgement for $v < \bar{v}$ and $C > \bar{C}$; no presentation despite the willingness of potential audience members to acknowledge for $v > \bar{v}$ and $C < \bar{C}$; and finally, no willingness to present or acknowledge in equilibrium for $v < \bar{v}$ and $C < \bar{C}$. Comparative statics

---

[8] Note that when M=2, we have C=0 and thus $U^P - U^{NP} = \lambda(\tau - \delta W)$. This means sharing occurs only if $\tau > \delta W$.

with respect to the parameters yield the following results.

Proposition 2 *(i) The range of parameters for which acknowledgement (A) is an equilibrium strategy is increasing in M, ρ, R_G and decreasing in W and σ.*

(ii) The range of parameters for which presentation (P) is an equilibrium strategy for researcher 1 is increasing in τ, ρ, and R_G and decreasing in W, x and δ. It is increasing in M for $\tau > \delta W$. The effect of σ is ambiguous.

The results for M and ρ in Proposition 2(i) follow from the fact that an increase in either of them increases the likelihood of verification. An increase in R_G increases the cost to the winner from misappropriation should it be discovered. On the other hand, an increase in *either W or σ* increases the gain to the winner should misappropriation escape verification

The results in (ii) provide several useful insights for the empirical analysis. First, an increase in feedback, τ, increases the benefit of sharing. This result allows us to relate team size with general sharing: researchers in a larger team need less external feedback, thus they are less likely to generally share. Second, increases in the size of prize W increases the potential loss from sharing. Additionally, an increase in the size of the audience, M, increases the likelihood of having at least one competing complementor in the audience which increases both benefit associated with feedback and the risk of losing the competition. If $\tau > \delta W$, the feedback effect dominates so that sharing in equilibrium is more likely. Finally, in a community with a higher likelihood for verification and punishment for misappropriation (corresponding to a high ρ and R), risk of sharing is lower and thus sharing is more likely to occur.

3   Econometric Analysis

To examine the extent to which sharing differs in these two contexts, we exploit a unique survey of bio-scientists' decisions to share information or not. The researchers are employed in a university or a public research organization in Germany or the United Kingdom. Industry researchers are excluded since their sharing is related to motives not found among public sector researchers (see, e.g., Haeussler, 2011). We exclude questionnaires from researchers who were older than 65 years. The final sample has 1173 observations that met our criteria (approximately 21% are employed in the United Kingdom). In the Appendix we provide details of the survey.

Of greatest importance to the present study is a series of four questions regarding a researcher's decision to share information (Table 1). Questions 1 and 2 cover specific sharing, and questions 3 and 4 address general sharing. Responses are measured on five-point Likert scales ranging from disagree

strongly to agree strongly. Responses are coded so that higher scores imply a greater likelihood of sharing. A strongly disagree response for questions 1, 2 or 4 is coded as a 5 while strongly agree response is coded as a 1. For question 3 strongly disagree is coded as a 1 and strongly agree as a 5.

Summary statistics are in Tables 2 and 3. The correlations among the sharing question responses in Table 3 are positive and significantly different from zero at a 1% level, and the largest correlation is less than 0.5. Together the correlations suggest that the four questions address distinct issues within and across types of sharing.

We use an ordered logit model to examine how responses to the four sharing questions relate to a set of independent variables. A panel is created using the four questions and we use interaction terms (see below) to estimate a single econometric model explaining Likert scores for the general and specific sharing questions. That is, the first person in the sample provides the first four observations, the second person provides observations 5 through 8, etc. Since each respondent can appear in the data up to four times, we use cluster standard errors to account for within individual correlations across the disturbances.

To explore econometrically differences between general and specific sharing, the same regressors are used for both. However, as one might expect from the theory, the marginal effects on specific and general sharing may be quite different. To deal with different marginal effects by type of sharing we create two binary variables: $Specific$ is equal to one if the question relates to specific sharing questions 1 or 2 in Table 1 (zero, otherwise), and $General$ is equal to one if the question relates to general sharing questions 3 or 4 in Table 1 (zero, otherwise). Each independent variables appears as an interaction with $Specific$ and as an interaction with $General$. That is, the model is

$$Y_i = \sum_{j=1}^{k} \beta_j \, Specific_i \cdot X_{ji} + \sum_{j=1}^{k} \gamma_j \, General_i \cdot X_{ji} + \epsilon_i.$$

Marginal effects for the types of sharing are then immediately obtained as the coefficients of the interactions; that is, if the observation relates to specific sharing then the marginal effects are given by the β's whereas the γ's are the marginal effects for general sharing. Alternatively, we could have relied on separate regressions for specific and general sharing. However, a single estimating equation (rather than separate regressions for the two types of sharing, or for each question) allows us to use cluster standard errors to capture the information contained in the cross question disturbances for each respondent.[9]

3.1 Regressors

We have four sets of key regressors and an additional set of control variables. Among the

_____

[9] An alternative is seemingly unrelated regressions (SUR). Table 7 includes the SUR results as a robustness check. Comparing Table 7 to the base model in Table 4 suggests that the results are nearly identical.

regressors of interest some are expected to have similar effects on both sharing decisions (e.g., competition); however, regressors related to the expectation of continued interaction or the value of feedback are expected to differ in their effects.

3.1.1 The Size of the Prize and Competition

Both types of sharing are expected to vary with the size of the prize. We do not directly observe the prize, but it is reasonable to expect competition to be greater for prizes of higher value. In the survey, respondents are asked to rate on a five-point Likert scale how tough the competition is in their field. *Competition* is expected to be negatively related to specific sharing.[10] All else equal, we would expect the same relationship for general sharing. However, *Competition* may be positively correlated with the number of competing researchers, M, and this will conflate the effect of *Competition* on general sharing. Proposition 2 suggests that an increase in M has an ambiguous effect on sharing. When presentation is highly risky in terms of increasing other researchers' odds of winning, an increase in M makes sharing less likely. In this case, a negative relation between *Competition* and general sharing is expected. However, when this risk is outweighed by the value of feedback and an increase in the likelihood of verification (see Proposition 2), the overall effect of *Competition* on general sharing can be positive.

A potential concern with *Competition* is that a respondent's view of competition could reflect the level of sharing in the respective field. This raises the issue of endogeneity in some settings; however, in our case, each individual is such a small part of the field that reverse causality is not a concern. Further, from our interviews, scientists' definitions of competition reflected issues of funding, publication rates, and competition for students, rather than information sharing. The Appendix provides further details.

As a measure of the commercial value of the prize we include the importance that respondents attach to patents for their reputation among peers. *PatentReputation* is measured on a five-point scale where larger values indicate greater esteem from patents. As measures of prizes of scientific value, respondents were also asked to rate on a five-point Likert scale the importance for their reputation among peers of the number of articles published in peer reviewed journals (*PubReputation*). Our theoretical models do not distinguish between prizes of commercial value and those that reinforce scientific reputation since both can be relevant. We expect *PubReputation* and *PatentReputation* to be negatively associated with both types of sharing.

Finally, respondents were asked to rate, on a five-point Likert scale, to what extent they agree that

---

[10] Blumenthal *et al.* (2006) uses a similar question but considers only a four-point scale. Notably the percentage of respondents who answer 4 or 5 in our survey is 71%, which is identical to the percent answering 4 in theirs. Hong and Walsh (2009) use Hagstrom's (1974) specification of competition in order to examine changes in secrecy over time.

the first to find new research results is highly esteemed among peers. Higher values of *FirstEsteemed* indicate greater esteem and we expect it to have a negative effect on both general and specific sharing.

3.1.2 The Expectation of Continuation in the Specific Sharing Game

In our model of specific sharing, the feasibility of sharing as a subgame perfect equilibrium is increasing in α, the probability of the game having a second period in which researcher 2 has information of value to researcher 1. Intuitively, one would expect this probability to be lower when one or more of the researchers is nearing retirement. Thus we include the age of the respondent, *Age*. In the general sharing model there is no prediction regarding the effects of α and, thus, no expectation on the effects of *Age*. *Professor* is an indicator variable equal to one if the respondent is a professor (and hence has tenure) and zero if the rank is less than professor. At the time of the survey, in the German academic system, which covers the majority of our respondents, faculty who had a lower rank than professor were almost always untenured. While untenured faculty may have a longer life cycle horizon to receive future reciprocity, they may also have a horizon defined by the date they are considered for tenure. The time horizon is expected to be shorter when at least one of the researchers is untenured compared to both researchers being tenured. Hence, the expected effect of *Professor* is positive in specific sharing.

An argument can be made that, all else equal, the size of the prize from research is higher for untenured faculty since the awarding of tenure is a part of the prize. This aspect of rank should affect specific and general sharing in the same way. In the case of specific sharing, this effect should reinforce the positive effect of career horizon for *Professor*; and in general sharing, it could lead to a positive coefficient for *Professor*.

3.1.3 Size of the Respondent's Research Group

We include *TeamSize,* which is the number of researchers with an academic degree who currently work in the respondent's research group, as a regressor. *TeamSize* clearly reflects a number of aspects of the research production process. By definition the more researchers in a lab, the larger is a researcher's built-in network of colleagues who can provide feedback on projects. Thus, consistent with Guimera *et al*. (2005) researchers in small teams will be more dependent on formal presentations as a feedback mechanism (a higher value of τ). This suggests a negative effect of *TeamSize* on general sharing.

TeamSize is also well known to reflect scale economies resulting from specialization and effective division of labor, with optimal team size determined by balancing scale economies against coordination costs (Becker and Murphy, 1992). It is not surprising, then, to find that larger teams appear to be more productive (Wuchty *et al.,* 2007; Adams *et al.,* 2005), which, all else equal, yields lower costs. One of the factors influencing specific sharing in our model is the cost of responding to the requestion and prior empirical studies this cost to be a major reason for denying requests (Campbell *et al.,* 2002 and

Walsh *et al.*, 2007). This aspect of $TeamSize$ suggests a negative coefficient in the context of specific sharing. However, we cannot rule out a positive coefficient, since more productive teams are better able to exploit information coming into the lab, an unobservable factor which in the theory has an ambiguous effect.

3.1.4   Scientific Norms for Openness

The greater respondent's beliefs that the norms of science operate in their field the greater is the expected level of sharing in both contexts. Respondents were asked to rate, on a five-point Likert scale, to what extent they agree that open exchange of information is usually practiced among researchers. Higher values of $OpenExchange$ indicate that more openness is practiced.

The specific sharing game includes a reputation loss, R_S, when sharing is not reciprocated. The general sharing game includes the possibility of misappropriation and reputational loss, R_G, if misapropriation is verified by researchers in the open forum. These aspects of general sharing should reinforce the expected positive relationship of $OpenExchange$ and general sharing. Respondents were also asked on a five-point Likert scale the extent to which they believe someone who exploits the ideas of others against their will is bound to lose reputation. Higher values of $ExploitLose$ reflect a stronger belief that punishment takes place and higher values are expected to be positively associated with general sharing.

3.1.5 Control Variables

We include the respondent's prior success at publishing and patenting as controls. $Publications$ is the total number of publications as reported by the respondent. Walsh *et al.* (2007) report that among academic bio-scientists the number of publications is positively associated with the likelihood that a request for information is denied. As a measure of prior success in patenting, we use the number of technically unique patent applications ($Patents$) which the respondent claims list them as an inventor. $Patents$ are associated with the commercial orientation of the researcher which we expect to be negatively associated with both types of sharing.

Respondents also were asked to rate, on a five-point Likert scale, how strongly they pursue basic research. Higher values of $Basic$ indicate a greater concentration on basic research, and our prior is that higher values are associated with greater sharing in both games. $OwnResearch$ is the percentage of the respondent's time that is spent on own research. This is a measure of how engaged the respondent is in research rather than other activities such as administration, teaching or grant writing; we do not have a prior about the effect of $OwnResearch$ on sharing.

We include two regressors in addition to $Patents$ to capture what might be referred to as academic entrepreneurship. $Consult$ is the percentage of the respondent's time that is spent "advising companies". Using a measure for business activity (ranging from being involved in writing a business

14

plan to founding a firm), Cohen and Walsh (2008) report that academic researchers involved in business activities show a lower willingness to fulfill an information request than researchers never involved in any such activity. They might be more concerned about protecting the potential to commercialize research findings or they might use the results in their consulting activities. Finally, $FamilyEnt$ is an indicator variable equal to one if a parent or sibling of the respondent is a founder of a firm. Researchers with family members who are entrepreneurs may be more cognizant of the potential commercial value of their discoveries and hence less likely to share. Haeussler (2011) finds that researchers with an entrepreneur in their family are less likely to fulfill a request for information. $Consult$ and $FamilyEnt$ are expected to have negative effects on both types of sharing.

We include $Responsible$, the number of full time employees who currently report directly to the respondent. In experimental settings Charness *et al.* (2007) and Song (2008) find that cooperation is less likely in repeated Prisoner's Dilemma games when individuals view themselves as representing members of a group. Thus we expect higher values of $Responsible$ to be associated with less sharing in the specific sharing game. We have no expectation regarding its sign for general sharing.

Other control variables include $Married$ which is an indicator variable equal to one if the respondent is married and $Male$ which is an indicator variable equal to one if the respondent is male.

Respondents were asked to indicate in which of 13 subfields of biological sciences they worked and field fixed effects are included. We also include fixed effects for the four questions.

3.2 Results and Discussion

The coefficients of the variables $OwnResearch$, $Consult$, $FamilyEnt$, $Married$ and the indicator for a UK researcher (UK) are neither individually nor jointly statistically different from zero. Since these regressors are included based on our priors rather than the theoretical models we drop them from the regression and estimate a more parsimonious model; results are in Table 4. For the sake of brevity we do not present the results for the full model (i.e., the one that also includes $OwnResearch$, $Consult$, $FamilyEnt$, $Married$ and UK) which are very similar to those of the parsimonious model.

Results in terms of odds ratios for the specific sharing coefficients are in Panel A of Table 4, and the general sharing results are in Panel B. An odds ratio greater than one indicates a positive effect of the regressor on the level of sharing while an odds ratio less than one indicates a negative effect. Recall from above that interacting each of the regressors with the $Specific$ and $General$ indicator variables allows us to immediately obtain the specific and general coefficients, and by estimating the coefficients jointly we are able to account for cross question correlations.

3.2.1 Specific Sharing

In the specific sharing results (Panel A) two of the variables associated with the size of the prize, $Competition$ and $PatentReputation$, have the expected negative signs (odds ratio less than one) and

are significantly different from zero. *FirstEsteemed* has the expected negative sign and *PubReputation* has an unexpected positive sign, but neither is significantly different from zero.

The time horizon is captured by *Age* and *Professor*. *Professor* has the predicted positive sign and is significantly different from zero, thus supporting Proposition 1. *Age*, however, is insignificant which may be due to having few observations on researchers close to the end of their career. In our sample the average age of respondents is 46 and only 15% of respondents are older than 55 and 6% are older than 60.

Consistent with the comparative static effect of cost, *TeamSize* has a positive and significant association with specific sharing.[11]

OpenExchange, the extent to which the respondent believes the norm of open exchange is practiced, has the expected positive sign and it is significantly different from zero. The likelihood of researchers' specific sharing increases when the community is perceived to follow the norm of communalism. *ExploitLose* has an unexpected negative, though insignificant, sign.

With respect to control variables, the coefficient of *Publications* is not significantly different from zero, and the coefficient of *Patents* is negative (as expected) but it is also not significant. *Basic* is positive and significant, as expected. The coefficient of *Responsible* is significant and has the anticipated negative sign.

Robustness checks were conducted, and the results remain similar to the base case. First, we included the square of the size of the team, *TeamSq*. The literature on the effects of team size on team productivity has generally found positive effects of increasing the size of teams when the team is small. Some have found a moderating effect as teams get larger (Diaz-Frances *et al*., 1995) while others have found the effect to remain linear (Cohen, 1981; Kretschmer, 1985). The results of including the squared term for specific sharing are in Table 5, Panel A. *TeamSize* and *TeamSq* are not individually significantly different from zero, but they are jointly significant (p-value = 0.036). The other results are very similar to the base case.

Second, we dropped two sharing questions (question 4, *Withhold* and question 1, *NotPass*) and results are in Table 6. Arguably, the dropped questions may not be clear measures of the type of sharing we model. However, the specific sharing results are little changed. The only noteworthy results is the non-significance of *Basic* and *TeamSize*.

Third, both *TeamSize* and *Responsible* are highly skewed. As a robustness check we dropped

---

observations if either regressor has a value greater than 99; 44 observations are dropped. The specific sharing coefficients for *TeamSize* and *Responsible* are no longer significant. This suggests that it is only in the very largest teams and/or when individuals are responsible for a large number of other researchers that there are significant effects on specific sharing. Other results are little changed and therefore not presented in detail.

Fourth, *Age* and *Publications* are not significant in the base model (Table 4), and they are highly correlated (the simple correlation is 0.57). We replaced *Publications* with the ratio of publications to age. The new results are very similar to those reported in Table 4 and therefore not presented in detail.

Lastly, as noted earlier we estimated our models using a seemingly unrelated regression approach. Results are in Table 7 and they are nearly identical to the ones presented in Table 4.

3.2.2   General Sharing

The general sharing coefficients are in Panel B of Table 4. The coefficient of *Competition* is negative and significant. This is consistent with the effect of W, the size of the prize, outweighing any positive effects of M. *PatentReputation* has the expected negative, significant sign associated with the value of the prize. On the other hand, *PubReputation* has an unexpected positive coefficient significant at the 10% level. The positive effect may reflect the fact that when publications are important for reputation, general sharing provides the researcher with the benefit of announcing her progress to the public, and as such presentation is an important communication mechanism. *PubReputation* likely reflects this as well as the value of the prize. *FirstEsteemed* has a statistically insignificant sign.

We do not have a prior on the effects of *Age* and it is not significantly different from zero. Faculty without tenure are expected to view tenure as a part of the prize so that *Professor* may be related to the size of the prize. As expected it has a positive coefficient, but it is not significant.

The need for feedback is relevant for general sharing and members of larger teams, all else equal, are not expected to rely as much on external feedback as are members of smaller teams. As predicted, *TeamSize* has a negative and significant effect on general sharing.

OpenExchange has the anticipated positive and significant coefficient. *ExploitLose* has a counterintuitive negative sign, but it is not significantly different from zero.[12]

With respect to the control variables, those who conduct more basic research are more likely to generally share. *Responsible* and *Patents* have expected negative and significant coefficients. Researchers with more publications are more likely to generally share, and men share less than women.

---

[12] If misappropriation is strongly sanctioned in a community, misconduct is rarely seen. If such rare events have not been observed by our interviewees, they might rate *ExploitLose* at a low level. This may explain the lack of effect.

As was the case with specific sharing, we did various robustness checks. In Panel B of Table 5 are results when $TeamSq$ is included. $TeamSize$ and $TeamSq$ are jointly significantly different from zero (p-value=0.000); the other results are very similar to the base model. In Panel B of Table 6 are results after dropping $WithHold$ and $NotPass$. The only notable differences is that $Competition$ is no longer significantly different from zero; its p-value, however, is 0.102. When we drop observations for which either $TeamSize$ or $Responsible$ is greater than 99 we find that $Responsible$ is no longer significant suggesting that it is only when respondents are responsible for large numbers of other researchers that there is a significant effect on general sharing. However, $TeamSize$ continues to be significant with an odds ratio less than one. Other results are little changed, so we do not present the detailed results. When $Publications$ is replaced by the ratio of publications to age, the results are very similar to those Table 4. Finally, the SUR results in Table 7 are nearly identical to those in Table 4.

3.3   Are the Models Different?

The main purpose of our empirical analysis is to test whether the two forms of sharing are different. A comparison of Panels A and B of Table 4 suggests that it is indeed the case. We tested for significant differences in the marginal effects for each of the regressors. Attention is confined only to regressors that are significantly different from zero in at least one of the panels in Table 4. For these variables, there are only three regressors with coefficients that are not significantly different across the two types of sharing: $Competition$, $Responsible$ and $Basic$ (p-values equal to 0.529, 0.271 and 0.688, respectively). The effect of competition is expected as, theoretically, competition has similar effects in both specific and general settings.

In other respects, however, the empirical results reflect factors which differ in the two contexts. These differences are borne out in the results reported in 3.2.1 and 3.2.2. In the specific sharing game, $TeamSize$ is significant and positive which is consistent with the fact that specific sharing can incur the cost of answering information requests and larger teams have smaller costs, but its effect in general sharing is quite different; in fact it is negative and significantly different from zero (the p-value in a test of their difference is 0.027). This is consistent with the fact that general sharing is a common approach for external feedback and larger teams have less need for external feedback. $Professor$, which is our variable reflecting tenure, is positively related to specific, but it is not related to general sharing. This latter result is expected since repetition is not necessary for general sharing to take place.

Two variables are positive and significantly related to sharing in both models, but marginal impacts are statistically significantly higher in general sharing than specific. These are $OpenExchange$ (p-value in the test of differences is 0.004), which reflects the environment for public sharing, and $PatentReputation$ (p-value in the test of differences is 0.010), which reflects the extent to which the respondent's reputation is affected by patents. When patents are considered important for one's

reputation, presenting intermediate research results in the public setting can compromise patent protection, while it is not necessary the case in private sharing.[13] In terms of $OpenExchange$, the more open the exchange in the community, the higher the likelihood of verification for misconduct.[14]

Above differences are consistent with our argument that a researcher can face different tradeoffs in the two contexts. In specific sharing, respondents share information privately, thus the potential benefit of public recognition and feedback from audience does not come with specific sharing. Instead, an important driver of sharing is the possibility of future reciprocity. This implies that specific sharing occurs in equilibrium only when the researchers have sufficiently long expected career horizons to allow for future sharing. This future interaction is not a necessary condition for general sharing. Thus, the variable (*Professor*) that relates to career horizon has a greater effect on specific sharing than general sharing. In contrast, in general sharing, the potential benefit is credit for the results presented and feedback from the audience. The downside is, because the results are unknown until presentation, patents can be compromised if the result is patentable. Further, the presenter's credit may not be acknowledged. Thus, variables that relate to the risk of losing credit due to misappropriation (*Open Exchange*) and the importance of patenting (*PatentReputation*) have a greater effect on general sharing than on specific sharing. Finally, specific sharing may bear a cost of preparation that the respondent would not incur absent the request. Thus lower costs of sharing for those in larger teams promotes specific sharing, but not general sharing. In fact, a larger team is less likely to engage in general sharing during research process because it has a greater built-in team for feedback.

4   Concluding Remarks

Information-sharing during the research process provides the basis for cumulative knowledge production and thus for scientific progress. While it is desirable from a communal point of view, researchers endogenously choose whether or not to share. More importantly, we find that researchers' decisions are context specific. For example, researchers in large labs may be more receptive to specific requests for data than those in small labs because the costs of sharing are lower, while they may be less likely, *ceteris paribus*, to share generally because they have a larger network within the lab for feedback.

In terms of policies, any policy increasing competitive pressures will reduce sharing, *ceteris paribus*. Thus prizes or rankings (e.g., Thomson's Science Citation Index) may encourage entry into

---

[13]   Any public disclosure jeopardizes patent rights in first-to-file-patent systems where the first inventor to file a patent application is granted the patent. The United States' first-to-invent system is one of the few exceptions to this, in which case a provisional application provides some degree of protection.
[14]   In terms of other control variables, we find that *Male* does not have a statistically significant effect on specific sharing whereas it has a negative, significant effect on general sharing.

competitions, but they come with the unintended consequence of decreasing sharing during the research process. In contrast, requirements by funding agencies (among them the European Research Council, German Research Foundation, National Science Foundation) to provide an appropriate and adequate data-sharing plan make sharing more likely. In terms of our specific sharing game, such a requirement strengthens the norm of returning favors (i.e., the impact of R_S), where in this case, researchers are expected to provide conditions under which they would share the data obtained with public funding.

Another implication is that platforms which encourage general sharing should also be cognizant of the importance of verification that results are novel to the author. Although journals penalize plagiarism in principle, their policies are often lax. A counterexample is the journal *Nature* which has a strict policy to support verification and encourages readers "who encounter refusal by the authors to comply with these policies should contact the chief editor" (*Nature*, 2012). Journals could also provide a platform to publicize instances of fraud as is done by the German *Laborjournal* in the life sciences. Public attention clearly helps to promote detecting plagiarism. The "GuttenPlag Wiki" was started as an online Wiki in Germany in 2011 to detect with the help of many volunteers instances of plagiarisms in doctoral theses with many "Plags" following.

Several limitations of the analysis suggest useful directions for research in this area. First, we made a number of theoretical simplifications. For example, for general sharing we assumed that the researcher's decision was whether to present to the entire community as would be the case for generally circulated working papers, presentations at conferences or where papers are posted on the internet. We did not examine more targeted sharing or the stage at which one might want to share information. Endogenizing the feedback decision and including the possibility of collaboration are also interesting but nontrivial problems worth pursuit. Moreover, we did not consider presentation of intermediate results to discourage rivals from continuing research in the area, but consider this as fruitful path for future research.

Second, there is a widely held belief that sharing information is always socially beneficial. In both of our models, while we take into account the fact that the sharer increases the chances of other researchers winning the prize, it does not affect the aggregate probability that the problem is solved. Such considerations are more complex, but nonetheless quite important to pursue.

Third, while our empirical analysis builds on cross-sectional data, a longitudinal dataset would allow one to track the career of a scientist over time and to track various cohorts of researchers.

Finally, caution should be exercised in generalizing the empirical results since they pertain to bio-scientists. While the bio-scientific field is a prominent example of a highly competitive field, it is an open question as to the extent our empirical findings operate in other scientific fields.

Acknowledgements

References

Adams, J., Black, G., Clemmons, R., Stephan, P., 2005. Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981-1999. Research Policy 34, 259-285.

Anton, J., Yao, D., 2002. Sale of ideas: Strategic disclosure, property rights, and contracting. Review of Economic Studies 69, 513-531.

Anton, J., Yao, D., 2004. Little patents and big secrets: Managing intellectual property. RAND Journal of Economics 35, 1-22.

Armstrong, J.B., Overton, T.S., 1977. Estimating non-response bias in mail surveys. Journal of Marketing Research 14, 396-402.

Arribas, I., Urbano, A., 2005. Repeated games with probabilistic horizon. Mathematical Social Sciences 50, 39-60.

Bailey, C., Euzent, P., Martin, T., List, J., 2001. Academic economists behaving badly? A survey on three areas of unethical behavior. Economic Inquiry 39, 162-170.

Baker, S., Mezzetti, C., 2005. Disclosure as a strategy in a patent race. Journal of Law and Economics 48, 173-194.

Becker, G., Murphy, K., 1992. The division of labor, coordination costs, and knowledge. The Quarterly Journal of Economics 107, 1137-1160.

Birnholtz, J.P., 2006. What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. Journal of the American Society for Information Science and Technology 57, 1758-1770.

Blumenthal, D., Nancyanne C., Campbell, E., and Louis, K., 1996. Relationships between academic institutions and industry in the life sciences. The New England Journal of Medicine 334, 368-374.

Blumenthal, D., Campbell, E., Gokhale, M., Ycel, R., Claridge, B., Hilgartner, S., Holzman, N., 2006. Data withholding in genetics and other life sciences: Prevalences and predictors. Academic Medicine

81, 137-145.

Campbell, E.G., Clarridge, B., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N.A., Blumenthal, D., 2002. Data withholding in academic genetics. Journal of the American Medical Association 287(4), 473-480.

Charness G., Rigotti, L., Rustichini, A., 2007. Individual behavior and group membership. American Economic Review 97, 1340-1352.

Cohen, J.E., 1981. Publication rate as a function of laboratory size in 3 biomedical-research institutions. Scientometrics 3, 467-487.

Cohen, W., Walsh, J., 2008. Real impediments to academic research. Innovation Policy and the Economy 8, 1-30.

Couzin-Frankel, J., Grom, J., 2009. Plagiarism sleuths. Science 324, 1004-1007.

Dasgupta, P., David, P., 1987. Information Disclosure and at the Economics of Science and Technology, in: Feiwel, G.R. (Ed.), Arrow and at the Ascent of Modern Economic Theory. University Press, New York, pp. 519-542.

Dasgupta, P., David, P., 1994. Toward a new economics of science. Research Policy 23, 486-521.

Diaz-Frances, E.,Ruiz-Velasco, S., Jimenez, J., 1995. Relationship between Publication Rate and Research Unit Size in Mexico, in: Proceedings of the fifth biennial Conference of the International Society for Scientometrics and Informetrics. Learned Information Inc., Medford, pp. 137-146.

Enders, W., Hoover, G.A., 2004. Whose line is it? Plagiarism in economics. Journal of Economic Literature 42, 487-493.

Gächter, S., von Krogh, G., Haefliger, S., 2010. Initiating private-collection innovation: The fragility of knowledge sharing. Research Policy 39, 893-906.

Gans, J., Murray, F., 2010. Funding Conditions, the Public-Private Portfolio, and Disclosure of Scientific Knowledge, in: Lerner, J., Stern, S. (Eds.), The Rate and Direction of Inventive Activity Revisited. University of Chicago Press, Chicago.

Gans, J., Murray, F., Stern, S., 2011. Contracting over the Disclosure of Scientific Knowledge: Intellectual Property and Academic Publication. SSRN Working Paper, 8 April 2011.

Gill, D., 2008. Strategic disclosure of intermediate research results. Journal of Economics and Management Strategy 17, 733-758.

Guimera, R., Uzzi, B., Spiro, J., Nunes-Amaral, L.-A., 2005. Team assembly mechanisms determine collaboration network structure and team performance. Science 308, 697-702.

Haeussler, C., 2011. Information-sharing in academia and industry: A comparative study. Research Policy 40, 105-122.

Hagstrom, W.O., 1965. The Scientific Community. Basic Books, New York.

Hagstrom, W.O., 1974. Competition in science. American Sociological Review 39, 1-18.

Hellmann, T., Perotti, E., 2011. The circulation of ideas in firms and markets. Management Science 57, 1813-1826.

Hong, W., Walsh, J., 2009. For money or glory? Commercialization, competition, and secrecy in the entrepreneurial university. The Sociological Quarterly 50, 145-171.

Hoover, G.A., 2006. A game-theoretic model of plagiarism. Atlantic Economic Journal 34, 449-454.

Kretschmer, H., 1985. Cooperation structure, group size and productivity in research groups. Scientometrics 7, 39-53.

Lacetera, N., Zirulia, L., 2011. The economics of scientific misconduct. Journal of Law, Economics, and Organization 27, 568-603.

Lerner, J., Tirole, J. 2002. Some simple economics of open source. Journal of Industrial Economics 50, 197-234.

Mailath, G.J., Samuelson, L., 2006. Repeated Games and Reputations: Long-Run Relationships. Oxford University Press, Oxford.

Merton, R.K., 1973. The Sociology of Science: Theoretical and Empirical Investigations. University of Chicago Press, Chicago.

Mukherjee, A., Stern, S., 2009. Disclosure or secrecy? The dynamics of open science. International Journal of Industrial Organization 27, 449-462.

Murray, F., 2010. The onco mouse that roared: Hybrid exchange strategies as a source of productive tension at the boundary of overlapping institutions. American Journal of Sociology 116, 341-388.

Murray, F., O'Mahony, S., 2007. Exploring the foundations of cumulative innovation: Implications for organization science. Organization Science 18, 1006-1021.

Nature, 2012. Guide to Publication Policies of the Nature Journals (downloaded on 14 March 2012 from http://www.nature.com/authors/gta.pdf).

NSF, 2011. Grant Proposal Guide (downloaded on 3 January 2013 from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp).

NIH, 2012. NIH Grants Policy Statement (downloaded on 8 January 2013 from http://grants.nih.gov/grants/sharing.htm).

Song, F., 2008. Trust and reciprocity behavior and behavioral forecasts: Individuals versus group representatives. Games and Economic Behavior 62, 675-696.

Stein, J., 2008. Conversations among competitors. American Economic Review 98, 2150-2162.

Stephan, P., 1996. The economics of science. Journal of Economic Literature 34, 1199-1235.

Von Hippel, E., 1987. Cooperation between Rivals: Informal know-how trading. Research Policy 25, 291-302.

Walsh, J.P., Cohen, W.N., Cho, C., 2007. Where excludability matters: Material versus intellectual property in academic biomedical research. Research Policy 36, 1184-1203.

Wuchty, S., Jones, B., Uzzi, B., 2007. The increasing dominance of teams in the production of knowledge. Science 316, 1036-1038.

Appendix: Survey Design

We developed and administered a survey in 2007 to bio-scientists in Germany and the UK, two leading countries in the bio-sciences. To identify bio-scientists we sampled authors listed in PubMed, the most prominent database of bio-scientific and medical abstract citations. We identified 9,074 German and 8,189 British researchers who had published between 2002 and 2005, using search categories related to the bio-scientific field. We then sampled all inventors who filed patents with bio-scientific IPC codes with the European Patent Office (EPO) between 2002 and 2005. This yielded 8,265 German and 4,196 British inventors. All identified researchers were invited to participate in an online questionnaire. A total of 2,169 researchers identified through PubMed and 2,452 identified through the EPO Database responded; this is a response rate of 13% of publishing researchers and 20% of inventors. For this paper, we restricted the sample to academic researchers who received a Ph.D. and were below 66 years. This gave a sample of 1,173 observations.

The search categories used for identifying researchers in the two databases were very broad. We concluded from discussions with experts and a small telephone survey with non-respondents that about 30% of the scientific authors and about 25% of the inventors in our sample were not in fact involved in bio-scientific research. In PubMed, as well as in the European Patent Database, there are no search categories or IPC classes that explicitly identify bio-scientific research. In the invitation letter to researchers we noted that our target respondents are researchers involved in bio-scientifics. After correcting for those who had received an invitation but were not involved in bio-sciences, we ended up with a response rate of 17% in the case of publishing researchers and 26% in that of inventors.

To test for non-response bias, we followed the standard procedure of testing whether the answers to our dependent variable of sharing information differ between early and late respondents (see Armstrong and Overton 1977). We estimated a series of ordered logit models. To distinguish between the first wave and last wave of respondents, we used a quite conservative threshold by comparing the 10% of first respondents with the 10% of last respondents, and in another series of tests the 20% of the first with the 20% of the last respondents. In neither specification was the coefficient of our independent variable (early versus late respondent) significant. In addition, we tested whether our early respondents received more requests from other researchers to share information. Arguably the (early) respondents might be more willing to provide information which would affect our results. However, the coefficient of comparing the early and late 10% as well as the early and late 20% of respondents is not significant.

In preparing the survey we conducted a series of 16 pre-tests with scientists in both Germany and the UK who were part of the target population, but were not included in the final sample. As part of this process, we went through the questionnaire with the interviewees and asked for their feedback on the questions including any additions they would find important. We also placed the questions in the survey

in a manner that respondents first had to answer field level questions before asking them about sharing of information. Further, in our latest interviews with researchers, we specifically asked about different aspects of competition and how scientists would define competition. Scientists defined competition in terms of funding, publication rates, and acquiring skilled students. They did not mention the level of sharing as an indicator for the level of competition.

We further computed various correlations to check whether they would give us an indication of reverse causality, but did not find any support for such an indication. For example, the correlations between our measure for competition and our sharing variables are quite low (0.050 for *PresentUnpub*; 0.055 for *Withhold*; 0.078 for *NotPass*; and 0.065 for *ExpectFutInfo*). We also correlated our measure for competition with the number of requests a researcher received for results researchers outside their organization. When there is possibility of information exchange influencing the level of competition in a field this should also be reflected in the relation between competition and the frequency with which researchers approach each other. Our researchers received on average 8.3 requests within the twelve months preceding the survey. The correlation coefficient between competition in the field and received requests is quite low (corr=0.045).