Sharing Among Competing Researchers

Carolin Haeussler*
University of Passau

Lin Jiang
University of Missouri

Jerry Thursby
Georgia Institute of Technology

Marie Thursby†
Georgia Institute of Technology & NBER

May 31, 2011

Abstract

We provide a game theoretic analysis of information sharing among competing researchers in two contexts: sharing when one researcher is asked by another to share specific information or materials and sharing involving presentation of new results in an open forum. The models are tested based on a survey of German and UK bio-scientists. The theory and empirics both suggest that academia is less open than one might think, and sharing is highly context dependent. Sharing in both specific and general contexts is negatively related to competition and the importance of patents in scientific reputation. In other respects, such as career stage, they differ markedly, with nontenured faculty are less likely to respond to specific requests. Scientists in larger labs are more likely to do so, but they are less likely to share in open forums.

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

*The authors are grateful for invaluable discussions with Emmanuel Dechenaux, Dietmar Harhoff, Ilinen Kondo, Nico Lacetera, Hugo Mialon, Fiona Murray, Lynne Zucker and participants of seminars at the 2010 DRUID, 2009 NBER Summer Institute, and the 2009 Communia in Turin Italy. Haeussler acknowledges financial support from the German Research Foundation [SFB TR15] and the Munich Center of Health Sciences. Jiang and Marie Thursby acknowledge funding from the National Science Foundation (Sub-award 44771-7471 of Award 0335765) and the Mildred and Alan Peterson Foundation. Jerry and Marie Thursby acknowledge NSF Award 0965289.

†Corresponding author: marie.thursby@mgt.gatech.edu; tel-494-894-6249; fax-404-385-4894.

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 drawn considerable attention to information sharing among academic researchers (Blumenthal *et al.* 1996; Cohen and Walsh 2008; Murray 2010).

In this paper, we examine what drives competing academic researchers to share information about their work. We distinguish between sharing in two contexts: situations in which a researcher is asked by another to share specific information about the research process (data, methods, or algorithms) or materials (cell lines or mice) and situations involving presentation of new results in an open forum, such as conference presentation or webposting. The former we call specific sharing and the latter, general sharing. We derive testable hypotheses for both types of sharing and provide empirical tests using a survey of bio-scientists in the United Kingdom and Germany regarding their willingness to share research results and materials with other bio-scientists. One of the most striking features of our analysis, both theoretically and empirically is our finding that what drives researchers to share in these two types of sharing differ markedly.

Our results are based on two simple games corresponding to specific and general sharing. The game for the specific sharing has clear elements of a Prisoner's Dilemma (Dasgupta and David 1994). If a researcher shares materials or information about her techniques with another outside of her lab (or collaborative team), she increases the likelihood that the other researcher will solve the problem before she does. On the other hand, it has the potential benefit that the other researcher may share in the future. Both researchers would be better off if they shared, but in equilibrium neither shares unless the game is repeated. We specify a probabilistic horizon, which allows us to derive hypotheses regarding sharing at various stages of researchers' career cycles. In addition to the probability that the game will continue, the likelihood that sharing occurs in equilibrium depends on the value of the prize, the value of what is shared, and the researchers' respective abilities to exploit the technique or material.

In the general sharing game, we focus on the conflicting incentives facing researchers when they consider sharing intermediate research results with the entire community prior to publication. In this context, the benefit of sharing is the potential for feedback and credit for the part of the problem she has solved, but there is an expected cost because members of the community may have solved complementary parts of the problem so that sharing increases the likelihood that someone else will win the prize. In addition, researchers who share are not guaranteed that their contribution will be acknowledged. In this game, whether or not sharing is an equilibrium outcome depends on the researchers' beliefs as to whether work that is not acknowledged would be verified by others. When there are only two researchers in the community, verification is not possible and the equilibrium outcome is similar to that of the single stage specific sharing game; without sufficiently valuable feedback the researcher keeps her results to herself. With more than two researchers, whether or not sharing occurs in equilibrium depends on the value of the prize, the size of the community, basic beliefs about verification and punishment for lack of acknowledgement, as well as how close the researcher is to a solution worthy of the prize and the extent to which the shared information improves complementors' chances of winning it.

The two situations have the common feature that an increase in the value of the prize makes sharing less likely because a researcher who unilaterally shares increases the odds that another researcher will win the prize. Nonetheless, there are substantial differences highlighted by our games. In specific sharing, the existence of the information or material is already known to the research community; but in general sharing, the results are not known until the researcher presents. Thus in specific sharing, credit is not an issue, but it is in general sharing, where the researcher knows the characteristics of her audience in expectation only. Moreover, immediate feedback from general sharing can also outweigh the increased odds of a rival winning the prize when there are only two researchers. By contrast, in situations such as sharing of materials, there is no need for feedback and the potential benefit is future reciprocity, so that sharing occurs only when the researchers in question have sufficiently long expected career horizons.

Our unique survey data allow us to examine these factors empirically in both contexts. Our analysis uses as the dependent variable responses to four questions in the survey on willingness to share. Two of the questions relate to specific sharing and two to general sharing. The survey data also include responses on the level of competition in the researcher's field, career stage, their scientific team, the research profile, entrepreneurship, and attitudes about the external research environment (to include ideas about the role of the norms of science), as well as demographic effects. While we do not have a direct measure of the "prize", one would expect its value to be correlated with the level of competition in the field. In both models, we find that competition and the importance of patents to the researcher's reputation are negatively related to willingness to share. While the impact of competition is not statistically significantly different across the two models, patent reputation has a larger impact in general than specific sharing. The models also differ markedly in other respects. For example, we find that larger labs are more likely to respond to specific requests but less likely to share generally. The result for specific sharing may be a result of specialization within large labs which lowers the cost of responding. By contrast, researchers in larger labs have a larger built-in network so that feedback from general sharing is less valuable, *ceteris paribus*. Finally, academic rank does not matter for general sharing, but it does for specific sharing, where untenured faculty are less likely to share.

Our games contribute to the theoretical literature on information exchange and disclosure of research results, which with few exceptions has focused on firm behavior (Anton and Yao 2002, 2004; Lerner and Tirole 2002; Baker and Mezzetti 2005; Hellmann and Perotti 2010; Gans *et al.* 2011; Gill 2008; Stein 2008). Hellmann and Perotti (2010) and Stein (2008) are similar in relating sharing of ideas to complementarity among the players, but their players are not competing for a priority-based prize and they assume an extreme form of complementarity. In their work, further production of ideas or inventions by researchers requires the skills or ideas of complementors. In our model of general sharing, complementors have solved complementary parts of the problem, but all researchers have a positive probability of solving the complete problem.

There is an emerging theoretical literature on information exchange in academia which largely abstracts from specific sharing, focusing rather on disclosure of research results and the trade-off between publication and secrecy (Mukherjee and Stern 2009, Gans and Murray 2010). Several papers have also examined the impact of academic misconduct on research and publication decisions (Hoover 2006; Lacetera and Zirulia 2009). Our insights on verification in general sharing borrow from Lacetera and Zirulia's intuition. We differ from this stream of work by focusing on disclosure during the research process.

Our empirical results contribute to an emerging literature on the ways in which academics disclose their work (Murray 2010). While there has been little empirical analysis of academic information sharing during the research process, significant withholding has been documented (Blumenthal *et al.* 1996; Campbell *et al.* 2002). Factors identified as influential include the cost, involvement in entrepreneurial or other business activities, the ability of students to publish, and scientific competition (Hong and Walsh 2009; Walsh *et al.* 2007). For both industry and academic researchers Haeussler (2010) found expected reciprocity and the extent to which researchers perceive that their community adheres to the scientific norm of communalism to be important. These studies, however, concentrate on sharing in the specific context, where researchers have received requests for information.

To our knowledge, ours in the first study to formally model general as well as specific sharing, an exercise which reveals the importance of context in understanding scientific information sharing. This allows us to highlight career cycle effects which, for this sample, are significant only for specific sharing. It also allows us to examine the factors that influence general sharing in the presence of potential for misappropriation, something which is considered a major problem in science (Bailey *et al.* 2001; Enders and Hoover 2004; Birnholtz 2006; Couzin-Frankel and Grom 2009).

The remainder of the paper is structured as follows. In section 2, the specific and general sharing games are developed. These games are then empirically tested in section 3. Section 4 concludes with a discussion of the findings.

2 Games of Information Sharing

In both games, we consider researchers working to solve a common problem, which if completely solved earns a prize, W. The prize, such as publication, a Fields Medal or Nobel Prize could have academic value, or it could have commercial value, such as a patent, or it could have both. We further suppose that each of the researchers has solved a portion of the problem and/or developed materials of use in solving the problem. If a researcher shares her solution or materials, she makes it easier or more likely for the recipient(s) to earn the prize. The information in question could be materials (cell line, reagents), data or methods (software, lab technique), or intermediate research results useful for solving the research problem.

2.1 Specific Sharing

We first consider situations in which a researcher is asked by another to share specific materials, data, or information about techniques she has used. In this game, we abstract from issues of misappropriation and focus only on the effect of sharing on the probability of winning and the role of reciprocity in the sharing decision. In this regard, it is useful to think in terms of a request for materials, data, or methods which the research community knows the researcher has developed.

For simplicity, we also abstract from the researchers' decisions to ask for information and focus only on their choice as to whether to share. In part, this is because our data describe only the sharing decision, but also, as shown in Appendix A.1, the results of the game described below are not changed by considering the asking decision. All proofs are in Appendix A.2.

2.1.1 The Stage Game

Figure 1 presents a single stage of a game between two researchers. The researchers move sequentially, with each deciding to share or not when it is her/his turn to move. Researcher 1's expected payoff is given on the top line of each bracket and researcher 2's is given on the bottom line. Researcher i has materials, data, or research methods represented by $r_i \geq 0$ (i=1,2). The ability of researcher j to exploit the information shared by i is represented by $e_j \geq 0$, so that the value to researcher j of the shared information or materials is given by $V(e_j, r_i) \geq 0$. The value of a researcher's own information is not shown in her payoffs because it does not affect the relative returns to each strategy. The game is "winner take all" so that each researcher gets W with probability less than one.

We model the probabilities of winning such that unilateral sharing by a researcher lowers her/his probability of winning the prize. In the absence of sharing, $x_i(r_i, r_j) \in (0, 1)$ is the probability that researcher i wins the prize and $1 - x_i$ is the probability that the other researcher wins, where x_i is increasing in r_i and decreasing in r_j. If researcher i shares but j does not, researcher i's probability of winning is reduced by $\delta_i = \delta(e_j, r_i)$ and j's probability is increased by δ_j. We assume δ is increasing in both arguments since unilateral sharing leads to a greater reduction in the sharing researcher's probability of winning the greater the information shared and the greater the other researcher's ability to exploit it.

There is also a cost c_i for researcher i to prepare the materials or information requested. Thus when researcher i shares and j does not, i incurs both the costs of preparation and competition, yielding the lowest expected payoff, $(x - \delta_i)W - c_i$. We assume that this payoff is positive for both researchers, so that all expected payoffs are positive.

Under these assumptions, there is a gain to each researcher from not sharing given by $\varphi_i(NS) = \delta_i W + c_i$ regardless of whether the other researcher shares. Thus not sharing is a dominant strategy for each researcher and the unique subgame perfect Nash equilibrium is that neither researcher shares her/his information.[1] Even though there is no risk of misappropriation, sharing is not an equilibrium outcome for the stage game. Nonetheless, sharing by both researchers Pareto dominates the Nash as long as there is a net benefit from the exchange, i.e. $V(e_2, r_1) + V(e_1, r_2) - c_1 - c_2 > 0$. Thus the stage game is a classic Prisoner's Dilemma.

2.1.2 The Probablistic Horizon Repeated Game

Except in extreme cases (such as immediately before retirement), however, the opportunities to interact with colleagues and share information are not single events. A researcher who

[1] Notice this would be the unique Nash equilibrium in a simultaneous move game of specific sharing.

denies a request for information today may find herself desiring information from the other researcher in the future. Thus it is more natural to consider researchers' decisions in the context of a series of repeated stage games. There are, of course, many variants of repeated Prisoner's Dilemma games in which cooperative strategies (those with payoffs that Pareto dominate those of the stage game Nash strategies) can be supported as subgame perfect equilibria, but one that lends itself to our analysis is one with a probabilistic horizon game such as that of Arribas and Urbano (2005). In such a game, the stage game is played repeatedly an unknown, but finite number of times, and the researchers have a common probability distribution over the length of the repeated game. This structure will allow us to consider how the stage of researchers' careers affects the decision to share. For example, the expected horizon of untenured faculty is likely to be different than that of midcareer researchers with tenure.

Thus we consider a game of unknown, but finite, length T, in which the researchers assign a probability $p_t \geq 0$ to the game ending in period t. We consider trigger strategies, in which each researcher shares as long as the other has shared but once the other researcher refuses to share, she refuses to share in subsequent periods. In deciding whether to share in period t, researcher i weighs her gain against her expected loss if the game continues and researcher j does not share in future periods. In order for sharing to be an equilibrium, the expected loss to each researcher from punishment (the inability to gain access to the other's information in the future), $\pi_i(NS) = (\delta_j - \delta_i)W + V(e_i, r_j) - c_i$, must outweigh the maximum gain from not sharing in period t. Intuitively, this is more likely to occur the longer the expected length of the game. Put somewhat differently, the lower the probability the game will continue, the less weight the researchers place on their loss from not obtaining information in the future.

The condition for existence of sharing as a subgame perfect equilibrium is

$$\max_i \left\{ \frac{\varphi_i(NS)}{\pi_i(NS)} \right\} \leq E[T \mid T \geq t] - t. \tag{1}$$

An equilibrium involving sharing exists when the researchers expect the game to last long enough. Further, the gain from not sharing in period t relative to the loss incurred from the punishment in any future period determines the minimum number of additional periods the researchers must expect for such cooperation.

Arribas and Urbano (2005) characterize the expected time of play (i.e. the right hand side of condition 1) in terms of a parameter α which represents the extent to which players expect the stage game to continue beyond the current period.[2] They show that when $\alpha = 0$, the expected length of the game converges to 0 (i.e. $\lim_{t\to\infty} E[T \mid T \geq t] - t = 0$). That

[2] Mathematically, $0 \leq \alpha \leq 1$ is defined as $\lim_{l\to\infty} \frac{p_{t_{l+1}}}{p_{t_l}}$ where $\{p_{t_l}\}$ is a subsequence of $\{p_t\}$.

is, when the players' confidence in repetition of the game decreases rapidly, they expect the game to end soon, in which case the condition in (1) cannot hold. In contrast, when $\alpha = 1$, $\lim_{t\to\infty} E[T \mid T \geq t] - t = \infty$, and the players expect infinite repetition of the game so that the condition in (1) holds and the players cooperate. Finally, when $\alpha \in (0,1)$, $\lim_{t\to\infty} E[T \mid T \geq t] - t = \frac{\alpha}{1-\alpha}$, in which case cooperation for a number of periods is a subgame perfect equilibrium (SPE) when $\max_i \left\{ \frac{\varphi_i(NS)}{\pi_i(NS)} \right\} < \frac{\alpha}{1-\alpha}$.[3] In particular, applied to our game, sharing for a number of periods is a SPE when

$$\alpha > \overline{\alpha} = \max_i \left\{ \frac{\varphi_i(NS)}{\pi_i(NS) + \varphi_i(NS)} \right\} \tag{2}$$

where

$$\frac{\varphi_i(NS)}{\pi_i(NS)+\varphi_i(NS)} = \frac{\delta(e_j,r_i)W+c_i}{\delta(e_i,r_j)W+V(e_i,r_j)} \text{ for researcher } i.$$

Proposition 1 *Consider the probabilistic horizon game in which condition* (2) *characterizes the existence of sharing for some length of time as a subgame perfect equilibrium. Then the likelihood of sharing in equilibrium increases with (i) an increase in α or a decrease in c_i and (ii) a decrease in W if $V(e_i, r_j) > \frac{\delta(e_i,r_j)}{\delta(e_j,r_i)} c_i$. The effects of e_j, r_i, e_i, and r_j are ambiguous.*

With an increase in α the expected length of the game increases so that the weight the researchers attach to future punishment increases. That is the expected number of periods in which they can be punished increases. A decrease in c_i decreases researcher i's single period gain from not sharing, thus increasing 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 $\delta_i = \delta_j$, it simply says that the value of the material gained in the exchange exceeds the researcher's cost of supplying his own material.

Finally, on the one hand an increase in r_j or i's ability to exploit it, e_i, increases the expected loss to researcher i from not sharing, making researcher i more willing to share. On the other hand, r_j and e_i decreases j's incentive to share with i. Thus depending on whether i's incentive effect outweighs j's disincentive effect, the effects of r_j and e_i on sharing are ambiguous.

[3] See Arribas and Urbano (2005) for a proof for this case as well as when $\max_i \left\{ \frac{\varphi_i(NS)}{\pi_i(NS)} \right\} = \frac{\alpha}{1-\alpha}$.

2.2 General Sharing

In this section, we consider the decision to share information publicly, for example, by presenting intermediate results at a conference. Again researchers face conflicting incentives. Sharing such information allows the researcher to gain feedback, but if it provides useful insights for others it may increase the probability that someone else beats her to completely solving the problem. General sharing also has the benefit of announcing her progress which will afford credit for that work, but only if others acknowledge it. To examine this situation, we consider a game with more than two researchers working in the same area and allow for misappropriation of results. Our interest is in the conditions under which preliminary work is shared with a general audience and appropriately acknowledged as an equilibrium outcome.

Although this game has certain elements in common with the specific sharing game (and in the special case of two researchers, has a similar outcome to the single stage specific sharing game), the context is quite different. In specific sharing, the existence of the information or material (e.g. cell line data or reagents) is already known to the community, while in general sharing it is not known until the researcher presents (e.g. in a conference, working paper, or website).

We assume there are $M \geq 2$ researchers trying to solve the same research problem and as before W is the prize for the solution. To distinguish this from the one-on-one situation, we represent the portion of the problem researcher 1 has solved as $\sigma \in (0, 1)$. The $M - 1$ other researchers are trying to solve the same problem, but none has completely solved it; if any researcher has totally solved it, the game ends.

We consider the 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. For simplicity, in this situation we abstract from the cost of preparing for the presentation.[4] 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 feedback, which we represent as adding value τ to the prize, but it also 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 with at least one complementor.

[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, in the absence of sharing, the researcher need not allocate effort to preparation (e.g., preparing samples for shipment or instructions for database use).

Further, our main results for general sharing would not be significantly altered by including it.

If she shares with an audience without any complementors, she gets neither feedback nor reduces her probability of winning.

The game is represented in extensive form by Figure 2. In stage zero Nature chooses 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 (e.g. they can attend her presentation or access her working paper or website posting). Researchers obtain information from researcher 1, and all (including researcher 1) continue working on the problem. In stage two, nature decides which researcher first solves the problem. If the winner is researcher 1, the game ends. If the winner is not researcher 1, he decides whether to acknowledge 1's work (A) or not acknowledge it (NA).[5] If the winner acknowledges researcher 1's work, he earns only partial credit, $(1-\sigma)W$. If the winner does not acknowledge researcher 1's work, he earns the full credit of W. 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 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. Whether acknowledging researcher 1's work is in his interest depends on the probability that another researcher will verify the originality of his work, the reputational loss 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

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

For acknowledgement to be worthwhile for the winner, he has to expect the likelihood of verification to be sufficiently high. Recall that v is related to the number of researchers working on the same problem (and by our assumption privy to the working paper or having come to the presentation). Using the definition of v, the condition in (3) can be rewritten as

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

Thus, one of the implications of the model is that if only two researchers are working on

[5] Thus we have implicitly assumed that Scientist 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 problem, then the winner will never acknowledge researcher 1's work. Thus the only reason that researcher 1 would share is for feedback since credit for her progress will not be forthcoming. Although we have not examined researcher 1's decision yet, we will find that unless there is a third person who can verify, she will not share unless there is sufficient feedback. If $\tau = 0$ and $M = 2$, the unique equilibrium of this model is (NP, NA).

In making her decision, researcher 1 considers these two factors (verification and feedback) but also the impact of sharing on her likelihood of winning. The difference between her expected utility from sharing and not sharing is

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

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 is the positive aspect of complementors in the audience and depends on their feedback. The last term is the negative impact of complementors in the audience and depends on the extent to which sharing improves their chances of winning. This fits our intuition that for sharing to dominate not sharing, the effects of announcement and feedback need to outweigh the negative impact from complementors among the M researchers.

More precisely, the condition for $U^P - U^{NP} > 0$ can be written as

$$C > \frac{\lambda[\delta W - \tau]}{[(1-x) + \lambda\delta]\sigma W}. \tag{6}$$

Lemma 1 and Proposition 2 characterize the pure strategy equilibria and comparative statics for this game.

Lemma 1 *There are four potential pure strategy equilibria of the game* $(P, A), (P, NA), (NP, A)$, *and* (NP, NA). *Let* $\overline{v} = \frac{\sigma W}{R+W}$ *and* $\overline{C} = \frac{\lambda[\delta W - \tau]}{[(1-x)+\lambda\delta]\sigma W}$
(i) (P, A) *is an equilibrium for* $v > \overline{v}$ *and* $C > \overline{C}$.
(ii) (P, NA) *is an equilibrium for* $v < \overline{v}$ *and* $C > \overline{C}$.
(iii) (NP, A) *is an equilibrium for* $v > \overline{v}$ *and* $C < \overline{C}$.
(iv)(NP, NA) *is an equilibrium for* $v < \overline{v}$ *and* $C < \overline{C}$.

Proposition 2 *(i) The likelihood that acknowledgement by the winner is an equilibrium strategy is increasing in M, ρ, R and decreasing in σ. It is increasing in W if $\upsilon > \sigma$.*
(ii) The likelihood that researcher 1 *will share in equilibrium is increasing in τ, ρ, and R 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) are quite intuitive. The likelihood of verification increases with an increase in either the number of individuals working on the problem or the belief that a random selected researcher will verify the role of the sharing researcher's work in the winner's solution. Recall that R is the loss or penalty for misappropriation so this result is intuitive as well. An increase in R decreases the right hand side of (3) thus increasing the likelihood that the winner will acknowledge the sharing researcher's contribution. On the other hand an increase in σ, the portion of the problem that researcher 1 has solved, increases the right hand of (3).

The results in (ii) highlight the conflicting effects of sharing. An increase in feedback, τ, increases the positive effect from sharing with complementors, while increases in W, x or δ increase the potential loss from sharing with them. An increase in the size of the audience, M, increases the likelihood of at least one complementor in the audience which increases both the positive effect associated with feedback and the negative effect from increasing their chances of winning the prize, W. If $\tau > \delta W$, the feedback effect dominates so that sharing in equilibrium is more likely. Finally, increases in both ρ and R increase the probability that she will receive credit of σW, whether or not the winner, if not herself, acknowledges her contribution.

3 Econometric Analysis

We exploit a unique survey of public sector bio-scientists' willingness to share. The researchers are employed in a university or a public research organization in either the United Kingdom or Germany. Industry researchers are excluded since their willingness to share is related to motives not found among public sector researchers (see, for example, Haeussler 2010). 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). Appendix A.3 provides details of the survey.

Of greatest importance to the present study is a series of four questions regarding a researcher's willingness to share information. The questions along with our shorthand notation are in Table 1. Willingness to share is measured on a five-point Likert scale ranging from disagree strongly to agree strongly. With the exception of the third question, agreement implies some degree of unwillingness to share. For purposes of this analysis, we coded all responses so that higher scores imply a greater willingness to share or fewer restrictions on sharing. A strongly disagree response for questions 1, 2 or 4 is coded as a 5 and a strongly agree response is coded as a 1. For question 3 we code strongly disagree as a 1 and strongly agree as a 5.

The four sharing questions fall into the two distinct types of sharing discussed above. One pair (questions 1 and 2) covers specific sharing, and the other pair (questions 3 and 4) addresss general sharing. Arguably, question 4, $Withhold$, and question 1, $NotPass$, are not clear measures of the type of sharing we model above. Initially, we use all questions in Table 1 in our econometric models; in our robustness checks we drop questions 1 and 4 from the analysis; however, the results change very little.

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. Our econometric approach is to "stack" responses to the four questions so that we consider a single econometric model explaining Likert scores for the general and specific sharing questions as a function of a set of independent variables. That is, we have created a panel where the first person in the sample provides the first four observations (assuming that an answer is provided for each sharing question). The second person provides observations 5 through 8, etc. Since each respondent can appear in the data up to 4 times, we use cluster standard errors to account for within individual correlations across the disturbances.

Since we have information on a given researcher's opinions on both types of sharing, it is appropriate to consider the same set of regressors as explanations for both. However, as one might expect from the theory, the marginal effects of certain regressors on specific and general sharing may be quite different. To deal with different marginal effects for general versus specific sharing we create two binary variables: $Specific$ is equal to one if the question relates to specific sharing (that is, questions 1 or 2 in Table 1) and it is equal to zero otherwise, and $General$ is equal to one if the question relates to general sharing (that is, questions 3 or 4 in Table 1) and it is equal to zero otherwise. Each of our independent variables appears as an interaction with $Specific$ and as an interaction with $General$; marginal effects for the types of sharing are then immediately obtained as the coefficients of the interactions. 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) allows us to use of information contained in the cross question disturbances for each respondent.[6]

3.1 Specific Sharing

The independent variables capture information about life cycle or career stage attributes, the scientific team, the research profile, entrepreneurship, and attitudes about the external research environment (to include ideas about the role of the norms of science), as well as some demographic effects.

Under the conditions of Proposition 1 specific sharing is less likely the larger the prize for solving the problem. 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 competition is in their field. *Competition* takes on the value respondents attach to the level of competition where higher values indicate greater competition. The theory does not distinguish between prizes of commercial value and those that reinforce scientific reputation since both may be relevant. As measures of the importance of scientific recognition, we use two variables reflecting the extent to which respondents believe the scientific reward structure operates in their field. Respondents were asked to rate, on a five-point Likert scale, to what extent they agree that the first to find new research results is highly esteemed among peers. Higher values of $FirstEsteemed$ indicate greater esteem. Respondents were also asked to rate on a 5 point Likert scale the importance for their reputation among peers of the number of articles published in peer reviewed journals (*PubReputation*). As a measure of prizes of commercial value, 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. *Competition*, $FirstEsteemed$, *PubReputation* and *PatentReputation* are expected to be negatively associated with specific sharing.

We also include the respondent's success at publishing and patenting as controls. *Publications* is the total number of respondent 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 the commercial potential of the respondent's research, we use the number of technically unique patent applications (*Patents*) which the respondent claims list them as an inventor. Thus, not only do we include perceptions of the importance of patents and publications, but also the numbers of patents and publications. *Patents* are associated with the commercial

[6] An alternative would be seemingly unrelated regressions. However, to implement that would require four regressions with cross equation coefficient restrictions.

orientation of the researcher which we expect to be negatively associated with sharing.

The cost to a researcher of responding to information requests, which theoretically decreases her willingness to share, is also unobserved. As suggested by Guimera *et al.* (2005) larger teams enable specialization and effective division of labor, and empirically Wuchty *et al.* (2007) and Adams *et al.* (2005) find that larger teams are more productive. To the extent there are economies of scale, larger teams may have lower costs of complying with a request. Thus we include $TeamSize$, the number of researchers with an academic degree who are currently working in the respondent's research group. While scale economies should yield a positive coefficient, $TeamSize$ is also likely to reflect another unobservable factor, the ability to exploit information coming into the lab, which in the theory has an ambiguous effect.

The longer the length of time a researcher expects the game to continue, α, the greater the likelihood of sharing. Given the structure of the survey this only can be captured for the researcher to whom the request is made. An increase in age reduces the number of periods in which a researcher can be punished for not sharing. Studying sharing in the context of a specific, identified request, Haeussler (2010) finds that an older researcher is less likely to share information as predicted by our theory. We include the age of the researcher, Age. We also include $Professor$ which is an indicator variable equal to one if the respondent is a professor (and hence has tenure) and it is equal to zero if the rank is less than professor (and the respondent does not in general have tenure).[7] While untenured faculty generally have a longer life cycle horizon, they also have a horizon defined by the date they are considered for tenure. We argue that the latter dominates in determining the expected length of any game involving at least one untenured faculty member. An argument also can be made that the size of the prize from research is higher for untenured faculty since the awarding of tenure is a part of the prize. This would reinforce the positive effect of $Professor$ since the size of the prize for tenured faculty is less than that for untenured faculty.

We include *Responsible*, the number of full time employees who currently report directly to the respondent. In experimental settings Charness *et al.* (2007) and Fei Song (2008) find that cooperation is less likely in repeated Prisoner's Dilemma games when individuals view themselves as representing members of a group. While their setting is one of cooperation and ours is specific sharing, we nonetheless expect higher values of *Responsible* to be associated with less sharing.

The greater respondent's beliefs that the norms of science operate in their field the greater is the expected level of sharing. Respondents were asked to rate, on a five-point Likert scale,

[7] At the time of the survey, in the German academic system, it was almost always the case that those faculty who had a lower rank than professor were not tenured.

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. For a sample of researchers in academia and industry, Haeussler (2009) reports the likelihood of sharing information with an inquirer increases with the extent to which researchers perceive that their community adheres to the scientific norm of communalism. 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 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. Finally, *OwnResearch* is the percentage of the respondent's time that is spent on their 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 *PatentReputation* and *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 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. 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. In a recent study, Haeussler (2010) indeed finds that researchers with an entrepreneur in their family are less likely to fulfill a request for information.

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. Empirical evidence on the effect of gender on information-sharing is mixed. Whereas Campbell *et al.* (2002) find that men are more likely to refuse requests for information, Walsh *et al.* (2007) report women to be more likely to deny a request for information. Haeussler (2010) find no significant effect of gender on the willingness to share information. *UK* is an indicator variable equal to one if the respondent is a researcher working in the United Kingdom, otherwise they are working in Germany.

Respondents were asked to indicate in which of 13 subfields of biological sciences they worked. Multiple subfields were permitted. They were also provided with an "other" category. Indicator variables for subfield are included in the regressions; however we do not provide the estimated coefficients in our results. We also include, but do not report in Table 4, indicator variables for the four questions.

The odds ratios for the specific sharing coefficients are in Panel A of Table 4; the general sharing results are in Panel B and we discuss those results below. Note that 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. The variables *OwnResearch*, *Consult*, *FamilyEnt*, *Married* and the indicator for a UK researcher (*UK*) are included based on our priors rather than the theoretical models. None is statistically significant and with only the exception of *Consult* in Panel A of Table 4 all have t statistics less than one. For this reason we drop those variables from the regression and the parsimonious model results are in Table 5. We consider Table 5 to be the base model and we discuss results and conduct robustness checks using this base model. Note that the results in Tables 4 and 5, with the exception of *Patents* in Panel A and *TeamSize* in Panel B, are very similar.

Consider the specific sharing results in Panel A of Table 5. Three of the variables associated with the size of the prize, *Competition*, *FirstEsteemed*, and *PatentReputation* have the expected negative signs (odds ratio less than one) but *FirstEsteemed* is not significantly different from zero. *PubReputation* has an unexpected positive sign, but it is not significantly different from zero. The coefficient of *Publications* is not significantly different from zero, and the coefficient of *Patents* is negative (as expected) but it is also not significantly different from zero. Consistent with the comparative static effect of cost, *TeamSize* has a positive association with specific sharing, and the coefficient is significantly different from zero.

The time horizon is captured by *Age* and *Professor*. *Age* is not significantly different from zero and, hence, does not provide support for Proposition 1. In our sample the average age of respondents is a fairly young 46 and only 15% of respondents are older than 55 and 6% are older than 60, thus the insignificance of *Age* may be due to having few observations on researchers close to the end of their career. *Professor* has the predicted positive sign and is significantly different from zero.

The coefficient of *Responsible* has the anticipated negative sign and it is significantly different from zero. *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 scientists' specific sharing increases when the community is perceived to follow the norm of communalism. *ExploitLose* has an unexpected negative sign, but it is not significantly different from zero. *Basic* is positive and significantly different from zero, as expected.

The literature on the effects of team size on the productivity of the team 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). As a robustness check we included the square of the size of the team, $TeamSq$, and results for specific sharing are in Panel A of Table 6. $TeamSize$ and $TeamSq$ are not individually significantly different from zero, but they are jointly different from zero (p-value = 0.036). The other results are very similar to the base case.

Both $TeamSize$ and *Responsible* are highly skewed (see Table 2). As a robustness check we dropped observations if either $TeamSize$ and *Responsible* is greater than 99. Eleven respondents (44 observations are dropped). The coefficients of these variables for specific training, which are significant in Table 5 only at a 10% level, are now 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 are there significant effects on specific sharing. Other results are little changed. We do not present the detailed results.

Earlier we noted that the questions $Withhold$ and $NotPass$ are questionable as measures of the sharing we model. In Table 7 are results after dropping these questions. The specific sharing results are little changed, as expected. The only noteworthy results is the non significance of *Basic* and $TeamSize$.

3.2 General Sharing

To the extent that the level of competition is positively associated with the size of the prize, then Proposition 2 suggests that *Competition* should be negatively associated with general sharing, just as it was in the case of specific sharing. But the level of competition is also likely to be positively related to the number of researchers working on the same problem, M, which according to the proposition has an ambiguous effect on the likelihood of sharing. When general presentation is highly risky in terms of increasing other researchers' odds of winning, then an increase in M makes sharing less likely. In this case, we expect a negative relation between *Competition* and general sharing. However, when this risk is outweighed by the value of feedback, the overall effect of *Competition* is ambiguous.

As in the specific sharing case, we use additional measures of the prize, *per se*, $FirstEsteemed$, *PubReputation* and *PatentReputation*. These effects are not confounded by other factors, hence they are expected to be negatively associated with general sharing according to Proposition 2. In addition, to the extent that the awarding of tenure is a prize, then the effect of

$Professor$ should be positively associated with sharing.

Proposition 2(ii) implies that a researcher is more likely to present in an open forum, the higher value she attaches to feedback, τ. Respondents in small labs may place a higher value on external feedback than those in large labs to the extent that the latter have better internal access to feedback. This suggests that *TeamSize* should have a negative effect on general sharing. On the other hand, respondents are more likely to expect feedback on working papers and conference presentations when open exchange is the norm in their field, so that we expect a positive effect of *OpenExchange*.

Proposition 2(ii) also implies researchers are more likely to present the higher the likelihood of verification, ρ, and the higher is the penalty for misappropriation, R. We expect ρ and R to be positively correlated with a respondent's belief that the norms of science are operative in their field, which reinforces our expectation that higher values of *OpenExchange* will have a positive impact on sharing. Moreover, previous research suggests that the strength of a norm is associated with the anticipated consequence of violating the norm (e.g. Bendor and Swistak 2001; Henrich and Boyd 2001). As another measure of loss of reputation, we asked respondents to rate on a five-point Likert scale the extent to which they believe that 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.

Other controls included in the model are the demographic variables UK, *Age*, $Married$ and $Male$ and the variable *Responsible*. We do not have priors on the effects of these variables. We also include *Publications* and *Patents*. As the case of specific sharing, we expect that the number of patents will reflect the commercial orientation of the researcher and thus be negatively associated with general sharing. We also include *Basic*, *OwnResearch*, *Consult* and $FamilyEnt$. The research profile measures are expected to be positively associated with sharing and the measures of academic entrepreneurship are expected to be negatively associated with sharing.

The general sharing coefficients are in Panel B of Table 4. As noted in the section on specific sharing we drop from this regression the variables *OwnResearch*, *Consult*, $FamilyEnt$, $Married$, and UK. Results for the base model, which drops these regressors, are in Table 5.

The coefficient of *Competition* is negative and significant (odds ratio is less than one). This is consistent either with the effect of W, the size of the prize, outweighing any positive effects of M when feedback effects dominate or with the feedback effect in M being dominated by the potential loss from presenting to complementors (see Proposition 2(ii)). $FirstEsteemed$ and *PubReputation* have unexpected positive coefficients and $Professor$ has the expected positive effect, however, only *PubReputation* is significant and only at the 10% level. The significant positive effect of *PubReputation* may well reflect the fact that when publications are important for reputation, presentation is an important communication mechanism. This measure of the prize most surely reflects this as well as the value of the prize. On the other hand, *PatentReputation* has the expected negative sign associated with the value of the prize and it is significantly different from zero. *OpenExchange* has the anticipated positive coefficient and it is significantly different from zero. *ExploitLose* has a counterintuitive negative sign but it is not significantly different from zero. $Teamsize$, as predicted by our model, has a negative effect on general sharing and it is significantly different from zero.

As was the case with the specific questions, those who conduct more basic research are more willing to generally share. *Responsible* and *Patents* have expected negative coefficients and each is significantly different from zero. Women and researchers with more publications are more willing to generally share, and men share less than women. *Age* is not significantly related to general sharing.

In Panel B of Table 6 are results when $TeamSq$ is included. $TeamSize$ and $TeamSq$ are jointly significantly different from zero (p-value = 0.0); the other results are very similar to the base model. In Panel B of Table 7 are results after dropping $WithHold$ and $NotPass$. The only notable differences is that *Competition* is not longer significantly different from zero; its p-value, however, is 0.102.

In our discussion of specific sharing results we noted the skewness in both $TeamSize$ and *Responsible* so that we consider our regression after dropping observations where either $TeamSize$ and *Responsible* is greater than 99. *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. We do not present the detailed results.

3.3 Are the Models Different?

A comparison of the results in Panels A and B of Table 5 suggests that the two forms of sharing are empirically quite different. 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 5. For this set of variables there are three regressors with coefficients that are not significantly different across the two types. These are *Competition*, *Responsible* and *Basic*. Thus the greater the perceived level of competition and the more individuals the researcher is responsible for the less likely is sharing and the effect does not depend on the type of sharing. On the other hand, the more basic is a researchers agenda the more likely they are to share and the effect again does not vary by type of sharing.

The coefficient of $TeamSize$ for specific sharing is positive and significantly different from zero, whereas for general sharing $TeamSize$ has a significant, negative effect. The opposite effects of $Teamsize$ on the two types of sharing are expected based on our theoretical models. To the extent that $Teamsize$ reflects economies of scale, the specific sharing model predicts an increased sharing by larger teams, whereas in general sharing members of larger teams need to rely less on external feedback thus the effect is predicted to be negative.

The coefficient of $Professor$ is significantly different from zero in the case of specific sharing, but it does not have a statistically significant effect on general sharing. This is also consistent with our theoretical predictions. In the context of specific sharing, tenured professors have a longer time horizon for future sharing than untenured professors, and thus are more likely to engage in sharing. In the context of general sharing, however, tenured professors are less likely to rely on open forum to get credit or feedback than untenured professors.

Three variables, *OpenExchange* and *PatentReputation* have statistically significant coefficients for both types of sharing, but each shows a statistically significantly larger impact on general sharing than on specific sharing. In the case of *OpenExchange*, the larger positive effect on general sharing may reflect the fact that *OpenExchange* is positively related both to expected feedback and the expectation of verification in a situation of general sharing. Furthermore, $Male$ does not have a statistically significant effects on specific sharing whereas it has a negative, significant effect on general sharing.

4 Conclusion

Information-sharing provides the basis for cumulative knowledge production and thus for scientific progress. While sharing of information is desirable from a communal point of view, researchers endogenously choose whether they share or not, with their decision depending on competitive incentives in the research process.

Our game-theoretic analysis captures some of the main characteristics of the scientific research process and the academic research community. Our analysis of specific sharing suggests that the likelihood a researcher will comply with a request for information is negatively related to the size of the reward, or prize, for solving the problem the two researchers are investigating and the cost of providing materials, but positively related to the probability of the game continuing. For general sharing, our analysis indicates that the likelihood of presenting intermediate research results to the scientific community is increasing with the benefits from announcing preliminary results in terms of credit and feedback but decreasing with the danger that presenting might increase the chance that other researchers will solve the entire research puzzle and win the prize. In addition, general sharing depends on how likely it is that a contribution will be acknowledged and, if not, that others will verify.

In general, our empirical results support both models, and in particular, they support our contention that the question "to share or not" is not as simple as one might think and is highly context dependent. Among the statistically significant coefficients, only a measure of competition, the amount of basic research conducted by the respondent and the number of individuals reporting to the researcher are not significantly in the specific and the general sharing models. Moreover, the empirical differences are, in general, predicted by the theoretical models. For example, rank does not matter for general sharing, but it does for specific sharing, where untenured faculty are less likely to share. For specific sharing, large teams are more likely to share but less likely to share generally. In addition, the empirical results imply that the stronger beliefs that the norms of science operate, the more willing researchers are to specifically share, and their willingness to share is even greater for general sharing.

Several limitations of the analysis suggest useful directions for research in this area. First, to derive testable hypotheses 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. This would be the case for generally circulated working papers or presentations for conferences where papers are posted on the internet. Thus we did not examine salient issues on more targeted sharing (e.g. to a few trusted colleagues) or the stage at which one might want to share information.

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.

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 collaborative and competitive field, it is an open question as to the extent our empirical findings operate in other scientific fields.

5 References

Adams, James D., Grant C. Black, Roger Clemmons, and Paula E. Stephan. 2005. "Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981-1999." *Research Policy*, 34(3): 259-285.

Anton, James and Dennis Yao. 2002. "Sale of Ideas: Strategic Disclosure, Property Rights, and Contracting." *Review of Economic Studies*, 69: 513-531.

________. 2004. "Little Patents and Big Secrets: Managing Intellectual Property." *RAND Journal of Economics*, 35: 1-22.

Arribas, Ivan and Amparo Urbano 2005. "Repeated Games with Probabilistic Horizon." *Mathematical Social Sciences*, 50: 39-60.

Bailey, Charles D., Patrizia J. Euzent, Thomas L. Martin, and John A. List. 2001. "Academic Economists Behaving Badly? A Survey on Three Areas of Unethical Behavior." *Economic Inquiry*, 39(1): 162-170.

Baker, Scott and Claudio Mezzetti. 2005. "Disclosure as a Strategy in a Patent Race." *Journal of Law and Economics*, 48: 173-194.

Bendor, Jonathan and Piotr Swistak. 2001. "The Evolution of Norms." *American Journal of Sociology*, 106(6): 1493-1545.

Birnholtz, Jeremy 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(13): 1758-1770.

Blumenthal, David, Nancyanne Causino, Eric Campbell, and Karen Seashore Louis. 1996. "Relationships between Academic Institutions and Industry in the Life Sciences." *The New England Journal of Medicine*, 334: 368-374.

Campbell, Eric G., Brian R. Clarridge, Manjusha Gokhale, Lauren Birenbaum, Stephen Hilgartner, Neil A. Holtzman, David Blumenthal. 2002. "Data withholding in academic genetics." *Journal of the American Medical Association*, 287(4): 473-480.

Charness, Gay, Luca Rigotti, and Aldo Rustichini. 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(6): 467-487.

Cohen, Wesley and John Walsh. 2008. "Real Impediments to Academic Research." *Innovation Policy and the Economy* 8: 1-30.

Couzin-Frankel, Jennifer and Jackie Grom. 2009. "Plagiarism Sleuths." *Science*, 324: 1004-1007.

Dasgupta, Partha and Paul David. 1987. "Information Disclosure and at the Economics of Science and Technology." In *Arrow and at the Ascent of Modern Economic Theory*, ed. George R. Feiwel, 519-542. New York: University Press, New York.

_____________. 1994. "Toward a New Economics of Science." *Research Policy*, 23: 486-521.

Diaz-Frances, Emilio, S. Ruiz-Velasco, and Jaime Jimenez. 1995. "Relationship between Publication Rate and Research Unit Size in Mexico." Proceedings of the fifth biennial Conference of the International Society for Scientometrics and Informetrics. Learned Informationm, 137-146, Medford.

Enders, Walter, Gary A. Hoover. 2004. "Whose Line is it? Plagiarism in Economics." *Journal of Economic Literature*, 42: 487-493.

Gans, Joshua, Fiona Murray and Scott Stern. 2011. "Contracting over the Disclosure of Scientific Knowledge: Intellectual Property and Academic Publication." SSRN Working Paper.

Gill, David. 2008. "Strategic Disclosure of Intermediate Research Results." *Journal of Economics and Management Strategy*, 17: 733-758.

Guimera, Roger, Brian Uzzi, Jarrett Spiro, Luis A. Nunes-Amaral. 2005. "Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance." *Science*, 308: 697-702.

Haeussler, Carolin. 2010. Information-Sharing in Academia and Industry: A Comparative Study." Research Policy, 40(1): 105-122.

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

_________. 1974. "Competition in Science." *American Sociological Review* 39: 1-18.

Hellmann, Thomas and Enrico Perotti. 2010. "The Circulation of Ideas in Firms and Markets."

Henrich, Joseph and Robert Boyd. 2001. "Why People Punish Defectors: Conformist Transmission Stabilizes Costly Enforcement of Norms in Cooperative Dilemmas." *Journal of Theoretical Biology*, 208: 79-89.

Hong, Wei, and John Walsh. 2009. "For Money or Glory? Commercialization, Competition, and Secrecy in the Entrepreneurial University." *The Sociological Quarterly 50: 145-171.*

Hoover, Gary A. 2006. "A Game-Theoretic Model of Plagiarism." *Atlantic Economic Journal*, 34: 449-454.

Kretschmer, Hildrun. 1985. "Cooperation Structure, Group Size and Productivity in Research Groups." *Scientometrics*, 7(1-2): 39-53

Lacetera, Nicola and Lorenzo Zirulia. 2009. "The Economics of Scientific Misconduct," *Journal of Law, Economics, and Organization,* forthcoming.

Lerner, Josh and Jean Tirole. 2002. "Some Simple Economics of Open Source." *Journal* *of Industrial Economics*, 50(2): 197-234.

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

Mukherjee, Arijit and Scott Stern. 2009. "Disclosure or Secrecy? The Dynamics of Open Science." *International Journal of Industrial Organization,* 27(3): 449-462.

Murray, Fiona. 2010. "The OncoMouse that Roared: Hybrid Exchange Strategies as a Source of Productive Tension at the Boundary of Overlapping Institutions." *American Journal of Sociology*, 116(2), 341-388.

Murray, Fiona and Siobhan O'Mahony. 2007. "Exploring the Foundations of Cumulative Innovation: Implications for Organization Science." *Organization Science*, 18: 1006-1021.

Powell, Walter W., Douglas R. White, Kenneth W. Koput, Jason Owen-Smith. 2005. Network Dynamics and Field Evolution: The Growth of Inter-organizational Collaboration in the Life Sciences. *American Journal of Sociology*, 110(4): 1132-1205.

Song, Fei. 2008. "Trust and Reciprocity Behavior and Behavioral Forecasts: Individuals versus Group Representatives." *Games and Economic Behavior*, 62: 675-696.

Stein, Jeremy. 2008. "Conversations Among Competitors." *American Economic Review*, 98: 2150-2162.

Stephan, Paula. 1996. "The Economics of Science." *Journal of Economic Literature*, 34: 1199-1235.

Walsh, John P., Wesley M. Cohen, and Charlene Cho. 2007. "Where Excludability Matters: Material Versus Intellectual Property in Academic Biomedical Research." *Research Policy*, 36: 1184-1203.

Walsh, John P. and Wei Hong. 2003. "Secrecy is Increasing in Step with Competition." *Nature*, 422: 801-802.

Walsh, John P., Mujuan Jiang, Wesley Cohen. 2007. "Publication in the Entrepreneurial University." Presentation at the American Sociological Society, New York.

Wuchty, Stefan, Benjamin Jones, and Brian Uzzi. 2007. "The Increasing Dominance of Teams in the Production of Knowledge." *Science*, 316: 1036-1038.

A Appendix

A.1 Specific-sharing Game with Asking Decisions

In this game, we incorporate the decision of "asking for data" into the specific-sharing game described in Section 2. We show that the unique subgame perfect Nash equilibrium (SPNE) that is trembling-hand perfect (THP) involves asking and not sharing. Trembling-hand perfection rules out subgame imperfect equilibria that are unstable and thus, we only focus on THP equilibria.

In this set-up, scientist 1 and 2 decide whether to ask each other for data and whether to share their data when asked. AS stands for "asking" while NAS for "not asking"; S stands for "sharing" while NS stands for "not sharing".

There are six SPNEs shown by the bold paths in Figure 3: (2 AS, 1 NS, 1 AS, 2 NS), (2 AS, 1 NS, 1 NAS, 2S), (2 AS, 1 NS, 1 NAS, 2NS), (2 NAS 1 AS, 2 NS); (2 NAS, 1 NAS, 2 S), (2 NAS, 1 NAS, 2 NS). It is easy to show that none of the SPNEs with NAS is a trembling-hand perfect equilibrium. To show this, we prove that in subgames 1.3 and 1.4, the SPNE with NAS are not THP, while those with AS are THP.

We first show that in subgame 1.3 (1 AS, 2 NS) is THP. A THP strategy would remain to be in the equilibrium path even when the players involved have a small probability of deviating from the strategy. Assume scientist 1 plays a mixed strategy $(1-\varepsilon, \varepsilon)$ for (AS, NAS), for $0 < \varepsilon < 1$ where ε stands for a small error or deviation from 1's equilibrium strategy AS. Scientist 2's expected payoffs from playing S and NS are given by

$\pi_2^S = (1-\varepsilon)(1-x-\delta_2)W - (1-\varepsilon)c_2 + \varepsilon(1-x)W = (1-x)W - (1-\varepsilon)(\delta_2 W + c_2)$

$\pi_2^{NS} = (1-x)W.$

For a small deviation ε, scientist 2 maximizes his expected payoff by choosing NS.

Similarly, we assume scientist 2 plays strategy S with a probability of ε and plays NS with a probability of $(1-\varepsilon)$, for $0 < \varepsilon < 1$. Scientist 1's expected payoffs from playing AS and NAS are given by:

$\pi_1^{AS} = \varepsilon[(x+\delta_2)W + V(e_1, r_2)] + (1-\varepsilon)xW = xW + \varepsilon[\delta_2 W + V(e_1, r_2)]$

$\pi_1^{NAS} = xW.$

For all positive values of ε, scientist 1 maximizes her expected payoff by placing a minimal weight on NAS. Hence among the three SPNE in the subgame 1.3, only (1 AS, 2 NS) is trembling-hand perfect because the two scientists maximize their expected payoff by staying with (1 AS, 2 NS) even if there is a small chance of error. We thus rule out the SPNE (1 NAS, 2S) and (1 NAS, 2 NS).

For the same reasoning, the two SPNEs in subgame 1.4 with NAS are not THP because the payoff structure of subgame 1.4 is exactly the same as that of subgame 1.3. As a result, the only SPNE that is THP is (2 AS, 1 NS, 1 AS, 2 NS). Both scientists ask each other for information, but neither would share.

A.2 Proofs of Propositions

A.2.1 Proposition 1

Note that researcher i shares if and only if the condition for the cooperative equilibrium of the game exists. This condition is $\alpha > \max\{\overline{\alpha}_1, \overline{\alpha}_2\}$,where $\overline{\alpha}_i \equiv \frac{\delta(e_j,r_i)W+c_i}{\delta(e_i,r_j)W+V(e_i,r_j)}$, $i = 1, 2$. If the condition does not hold, for instance, $\alpha > \overline{\alpha}_1$ and $\alpha < \overline{\alpha}_2$, neither researcher would share. This is because, observing researcher 2's cutoff point $\overline{\alpha}_2$, researcher 1 knows that researcher 2 would not share in the future even if researcher 1 shares in this period. Thus researcher 1 knows he/she would be better off not to share as well. As such, the probability that researcher i would share information is the probability that the above equilibrium condition holds: $\Pr(i \; shares) = \Pr(\alpha > \max\{\overline{\alpha}_1, \overline{\alpha}_2\})$.

To examine the comparative statics on the probability of sharing by researcher i, we assume that α follows a uniform distribution $U(0, 1)$. Then we can rewrite the probability that researcher i shares as $(1-\overline{\alpha}_1)(1-\overline{\alpha}_2)$. Additionally, since we consider a typical repeated prisoner dilemma, we only consider the case when there is social loss when players deviate from cooperative equilibrium (i.e., $\pi_i(NS) > 0$). As such, $\overline{\alpha}_i = \frac{\varphi_i(NS)}{\pi_i(NS)+\varphi_i(NS)} < 1$, $i = 1, 2$.

It is easy to see the following:

1)$\frac{\partial \Pr(1 \, shares)}{\partial \alpha} > 0$.

2) Since $\frac{\partial \overline{\alpha}_i}{\partial c_i} > 0$, we have $\frac{\partial \Pr(1 \, shares)}{\partial c_1} = \frac{\partial \overline{\alpha}_1}{\partial c_1}(\overline{\alpha}_2 - 1) < 0$.

3) $\frac{\partial \overline{\alpha}_i}{\partial W} > 0$ if $V(e_i, r_j) > \frac{\delta(e_i,r_j)}{\delta(e_j,r_i)}c_i$. We assume the costs of sending inquired materials to other researchers arbitrally small since we consider the interesting case in which mutual exchange of materials increases social welfare. As such, $\frac{\partial \Pr(1 \, shares)}{\partial W} = \frac{\partial \overline{\alpha}_1}{\partial W}(\overline{\alpha}_2 - 1) + \frac{\partial \overline{\alpha}_2}{\partial W}(\overline{\alpha}_1 - 1) < 0$.

4) Additionally, $\frac{\partial \overline{\alpha}_1}{\partial y} > 0$,where $y = e_2, r_1$, and $\frac{\partial \overline{\alpha}_1}{\partial z} < 0$, where $z = e_1, r_2$. Similarly, we have $\frac{\partial \overline{\alpha}_2}{\partial y} < 0$ and $\frac{\partial \overline{\alpha}_2}{\partial z} > 0$. Thus we have $\frac{\partial \Pr(1 \, shares)}{\partial y}$ and $\frac{\partial \Pr(1 \, shares)}{\partial z}$ be negative or positive depending on the size of the parameters. Intuitively, if e_1 (or r_2) is large, although 1 has stronger incentive to share with 2, 2 has little incentive to share with 1. Expecting no sharing from 2, 1 would not share as well.

A.2.2 Proposition 2

Proposition (2i) follows from differentiating $(v - \overline{v})$ with respect to the parameters and (2ii) follows from differentiating $(C - \overline{C})$.

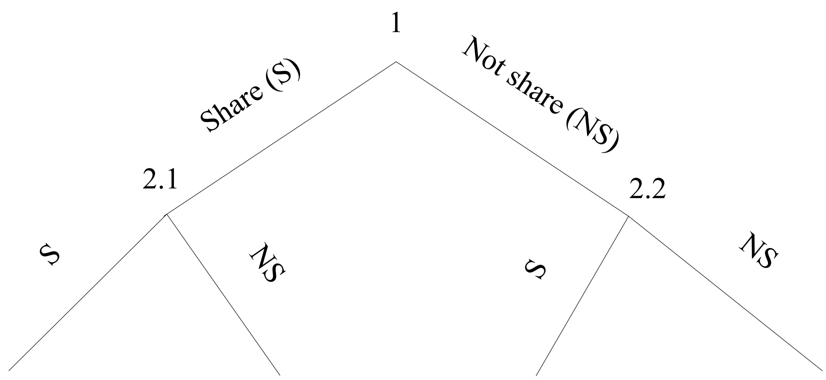
A.3 Survey Design

The bio-sciences provide an attractive testing ground for our propositions. Compared with many other scientific and technological fields, in the bio-sciences, research has developed dramatically in the last few decades. The building of collective knowledge is a key strategic task for the success of scientists (Powell *et al.* 2005).

We developed and administered a survey in 2007 to bio-scientists in Germany and the UK, the two leading countries in the bio-sciences in Europe. To identify bio-scientists we first sampled bio-scientists listed as authors in PubMed, the most prominent database of bio-scientific and medical abstract citations. From this we identified 9,074 German researchers and 8,189 British researchers who had published an article 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 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 European Patent Database filled out the questionnaire. This translates into a response rate of 13% of publishing researchers and 20% of inventors.

The search categories we 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 caught in our sample were not in fact involved in bio-scientific research. In PubMed, as well as in the European Patent Database (Epoline), there are no search categories or IPC classes that explicitly identify bio-scientific research. When designing the study, we therefore decided to use rather broad categories. In the invitation letter to researchers we pointed out that our target respondents are researchers involved in the bio-scientific field. Once we had corrected for the percentage of people who had received an invitation but were not involved in the bio-sciences (30% for publishing researchers and 25% for inventors), we ended up with a response rate of 17% in the case of publishing researchers and 26% in that of inventors.

Figures and Tables



$$\begin{pmatrix}(x-\delta_1+\delta_2)W+V(e_1,r_2)-c_1\\(1-x-\delta_2+\delta_1)W+V(e_2,r_1)-c_2\end{pmatrix}\begin{pmatrix}(x-\delta_1)W-c_1\\(1-x+\delta_1)W+V(e_2,r_1)\end{pmatrix}\begin{pmatrix}(x+\delta_2)W+V(e_1,r_2)\\(1-x-\delta_2)W-c_2\end{pmatrix}\begin{pmatrix}xW\\(1-x)W\end{pmatrix}$$

Figure 1: Specific sharing game

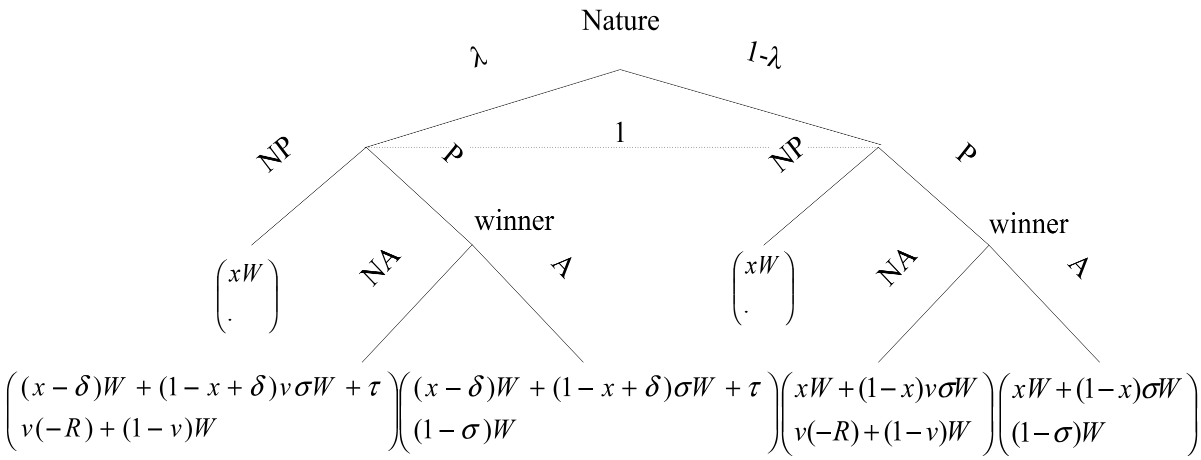


Figure 2: General sharing game

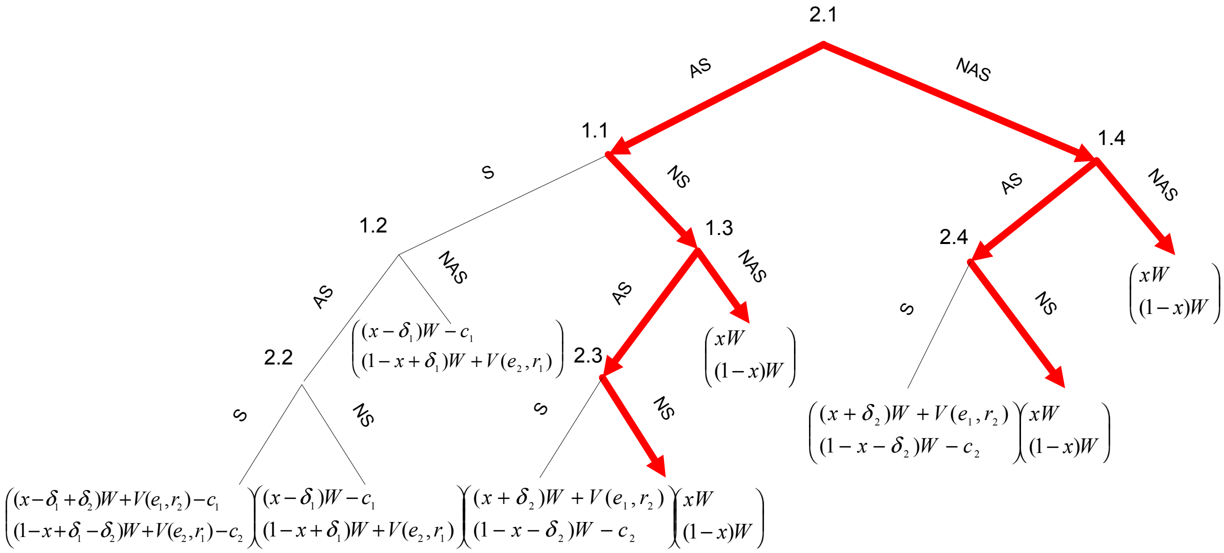


Figure 3: Specific-sharing game with asking decisions in the Appendix

Table 1. Sharing Questions

	Question	Question Shorthand	Type of Sharing
1	I only discuss unpublished or yet to be patented research results with people who will for sure not pass on this information.	*NotPass*	Specific
2	Before I share unpublished or yet to be patented research results, I first consider whether or not I will get valuable information from this researcher in the future.	*ExpectFutInfo*	Specific
3	I present unpublished or yet to be patented research results at conferences.	*PresentUnpub*	General
4	When I discuss unpublished or yet to be patented research results, I often withhold crucial parts	*Withhold*	General

Table 2. Summary Statistics

Variable	No. Obs.	Mean	St. Dev.	Min	Max
Dependent variables					
NotPass	1157	2.743	1.155	1	5
ExpectFutInfo	1131	3.308	1.105	1	5
PresentUnpub	1160	3.559	1.181	1	5
Withhold	1149	3.080	1.170	1	5
Life cycle or stage of career					
Age	1176	45.964	7.715	29	65
Professor	1176	0.517	--	0	1
Scientific team					
Responsible	1159	10.004	23.779	0	572
TeamSize	1165	6.741	11.871	0	300
Research profile					
Publications	1158	70.707	67.484	0	550
PubReputation	1173	4.060	0.875	1	5
PatentReputation	1144	1.955	0.825	1	5
Basic	1175	3.962	1.132	1	5
OwnResearch	1172	0.186	--	0	1

Table 2. Summary Statistics (con't)

Academic entrepreneurship					
Patents	1131	2.841	9.324	0	131
Consult	1172	0.189	4.981	0	80
FamilyEnt	1158	0.242	--	0	1
External research environment					
Competition	1173	4.051	0.993	1	5
OpenExchange	1173	3.304	0.911	1	5
FirstEsteemed	1169	4.053	0.898	1	5
ExploitLose	1160	3.956	1.117	1	5
Other controls					
Married	1154	0.816	--	0	1
Male	1172	0.799	--	0	1
UK	1173	0.209	--	0	1

Table 3. Correlations Among Sharing Question Responses*

	PresentUnpub	*Withhold*	*NotPass*
Withhold	0.400		
NotPass	0.430	0.441	
ExpectFutInfo	0.246	0.462	0.406

* All are significantly different from zero at the 1% level.

Table 4. Ordered Logit Results: Full Model

	A. Specific Sharing			B. General Sharing		
	Odds Ratios	**t-Stats**		**Odds Ratios**	**t-Stats**	
Competition	0.8883	-2.25	**	0.8592	-2.55	**
FirstEsteemed	0.9583	-0.78		1.0562	0.96	
PubReputation	1.0374	0.67		1.1053	1.71	*
PatentReputation	0.7302	-4.46	***	0.6156	-7.14	***
Publications	1.0010	1.06		1.0030	2.65	***
Patents	0.9885	-1.66	*	0.9673	-2.71	***
Teamsize	1.0068	2.06	**	0.9948	-1.48	
Age	0.9931	-0.83		0.9874	-1.51	
Professor	1.4406	3.18	***	1.0822	0.71	
Responsible	0.9963	-2.42	**	0.9948	-2.98	***
OpenExchange	1.2282	3.27	***	1.4549	6.12	***
ExploitLose	0.9467	-1.23		0.9538	-0.97	
Basic	1.1233	2.16	**	1.1471	2.64	***
OwnResearch	0.9983	-0.51		1.0014	0.47	
Consult	0.9889	-1.46		0.9934	-0.71	
FamilyEnt	1.0658	0.57		0.9400	-0.54	
Married	1.1825	1.22		1.0713	0.54	
Male	0.8425	-1.35		0.7412	-2.52	**
UK	0.9130	-0.73		0.8827	-0.99	
Field Fixed Effects	Yes					
Question Fixed Effects	Yes					
No. Observations	3992					
Pseudo r-square	0.0678					

*** Significant at 1% ** Significant at 5% * Significant at 10%

Table 5. Ordered Logit Results: Base Model

	A. Specific Sharing			B. General Sharing		
	Odds Ratios	**t-Stats**		**Odds Ratios**	**t-Stats**	
Competition	0.8770	-2.52	**	0.8566	-2.67	***
FirstEsteemed	0.9592	-0.78		1.0507	0.88	
PubReputation	1.0310	0.57		1.1025	1.7	*
PatentReputation	0.7248	-4.83	***	0.6095	-7.61	***
Publications	1.0012	1.19		1.0033	2.99	***
Patents	0.9904	-1.43		0.9680	-2.83	***
Teamsize	1.0063	1.91	*	0.9936	-1.98	**
Age	0.9925	-0.94		0.9881	-1.48	
Professor	1.4157	3.24	***	1.0386	0.36	
Responsible	0.9971	-1.77	*	0.9949	-2.93	***
OpenExchange	1.2232	3.3	***	1.4437	6.15	***
ExploitLose	0.9529	-1.1		0.9686	-0.69	
Basic	1.1145	2.09	**	1.1468	2.72	***
Male	0.8697	-1.12		0.7607	-2.4	***
Field Fixed Effects	Yes					
Question Fixed Effects	Yes					
No. Observations	4102					
Pseudo r-square	0.067					

*** Significant at 1%   ** Significant at 5%   * Significant at 10%

Table 6. Orderd Logit Results: Base Model Plus *TeamSq*

	A. Specific Sharing			B. General Sharing		
	Odds Ratios	**t-Stats**		**Odds Ratios**	**t-Stats**	
Competition	0.8774	-2.51	**	0.8548	-2.69	***
FirstEsteemed	0.9598	-0.77		1.0490	0.85	
PubReputation	1.0309	0.57		1.1032	1.71	*
PatentReputation	0.7241	-4.81	***	0.6111	-7.52	***
Publications	1.0012	1.18		1.0031	2.76	***
Patents	0.9905	-1.41		0.9672	-2.67	***
Teamsize	1.0047	0.53		0.9986	-0.14	
TeamSq	1.0000	0.23		1.0000	-0.67	
Age	0.9923	-0.95		0.9886	-1.42	
Professor	1.4185	3.25	***	1.0320	0.3	
Responsible	0.9972	-1.68	*	0.9947	-2.87	***
OpenExchange	1.2240	3.3	***	1.4406	6.1	***
ExploitLose	0.9529	-1.1		0.9686	-0.69	
Basic	1.1140	2.08	**	1.1484	2.74	***
Male	0.8692	-1.12		0.7620	-2.39	**
Field Fixed Effects	Yes					
Question Fixed Effects	Yes					
No. Observations	4102					
Pseudo r-square	0.0671					

*** Significant at 1%   ** Significant at 5%   * Significant at 10%

Table 7. Orderd Logit Results: Base Model Less *Withhold* and *NotPass*

	A. Specific Sharing			B. General Sharing		
	Odds Ratios	**t-Stats**		**Odds Ratios**	**t-Stats**	
Competition	0.8735	-2.31	**	0.8896	-1.63	
FirstEsteemed	0.9203	-1.27		1.1146	1.53	
PubReputation	0.9781	-0.33		1.2137	2.78	***
PatentReputation	0.6971	-4.59	***	0.5681	-6.38	***
Publications	1.0010	0.74		1.0041	3.02	***
Patents	0.9929	-0.97		0.9398	-2.28	**
Teamsize	1.0021	0.65		0.9837	-2.63	***
Age	0.9980	-0.21		0.9910	-0.86	
Professor	1.5809	3.48	***	0.9983	-0.01	
Responsible	0.9971	-1.78	*	0.9953	-2.56	**
OpenExchange	1.1673	2.02	**	1.5444	5.63	***
ExploitLose	0.9970	-0.05		0.9219	-1.37	
Basic	1.0847	1.28		1.1766	2.44	**
Male	0.8046	-1.44		0.7733	-1.73	*
Field Fixed Effects		Yes				
Question Fixed Effects		Yes				
No. Observations		2045				
Pseudo r-square		0.0665				

*** Significant at 1%   ** Significant at 5%   * Significant at 10%