General and Specific Information Sharing Among Academic Scientists

Carolin Haeussler*
Ludwig-Maximillians-University Munich

Jerry Thursby
Georgia Institute of Technology

Marie Thursby
Georgia Institute of Technology & NBER

Please do not quote or cite without permission.

June 15, 2009

Abstract

We provide theoretical and empirical evidence on the factors that influence the willingness of scientists to share research results. We distinguish between two types of sharing, specific sharing in which a researcher shares her data or materials with another and general sharing in which scientists report results to the entire community (as in conference presentations). We present two simple games in which scientists research a problem of scientific merit (with an associated prize, reputation, or perhaps commercial value). In both cases, the scientists have intermediate research results but none have solved the entire puzzle.

These games have implications for the design of policies to enforce the norms of science, such as the rules of database storage, punishment by journal editors for lack of acknowledgement, as well as the value of institutions which promote conferences and working papers. We use a unique survey of bio-scientists in the UK and Germany regarding their willingness to "share" to test these implications for specific and general sharing. Our results generally support both models and the fact that determinants of the two types of sharing are quite different. One important policy implication is that policies designed to enhance sharing must be tailored to the type of sharing.

*The authors are grateful for invaluable discussions with Hugo Mialon and Nico Lacetera and excellent research assistance from Lin Jiang. Marie Thursby thanks the National Science Foundation for support.

1 Introduction

Sharing of information 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, however, a tension between communal sharing and the competitive incentives for scientists during the research process itself (Dasgupta and David 1994). This tension, as well as the realization of commercial potential for much academic work over the past thirty years has drawn considerable attention to information sharing among academic scientists (Blumenthal et al. 1996, Campbell et al. 2000, Walsh et al. 2007).

In this paper, we examine what drives scientists to share information. We present two simple games in which scientists decide whether or not to share unprotected research results and derive testable hypotheses regarding their behavior. We then provide empirical tests using a survey of university bio-scientists in the United Kingdom and Germany regarding their willingness to share research results with other bio-scientists. One of the most striking features of our analysis, both theoretically and empirically, is our finding that what drives scientists to share and the extent to which they share depends on the context. To do this, we distinguish two types of sharing: one-on-one situations in which a scientist is asked by another to share data or materials and public sharing, such as conference participation, where researchers present work that is neither published nor patented at the time of presentation. The former we call specific sharing and the latter, general sharing.

For our theory, we consider scientists competing for a prize awarded to the first to solve a research problem. We model specific sharing as a simultaneous move game in which two scientists consider whether or not to share their materials or current research results. This situation has clear elements of a Prisoner's Dilemma (Dasgupta and David 1994). Sharing may increase the likelihood that the other scientist will solve the problem before she does. On the other hand, it has the potential benefit that the other scientist may share his materials in the future. Both scientists would be better off if they shared, but in equilibrium neither shares unless the game is repeated a sufficient number of times. We specify a probabilistic horizon, which allows us to derive hypotheses regarding sharing at various stages of scientists' career cycles. The likelihood that sharing occurs in equilibrium depends on the value of the prize, the value of their results to date, their ability to exploit the information shared, and the probability of the game continuing. The model also predicts that, all else equal, scientists with similar probabilities of winning the prize are more likely to share.

We model general sharing as a sequential move game with two or more scientists in which one scientist decides whether or not to report her results to the entire community. The benefit of presentation is credit for the part of the problem she has solved as well as potential feedback, but there is an expected cost because members of the community may have solved complementary parts of the problem so that presentation increases their likelihood of winning the prize. As well, scientists who present are not guaranteed that their contribution will be acknowledged. In this game, whether or not sharing is an equilibrium outcome depends on the scientists beliefs as to whether others will verify work that is not acknowledged. When there are only two scientists in the community, verification is not possible so the situation is similar that our game of specific sharing, except for the possibility of feedback. However, with more than two scientists, depending on a scientist's belief that someone will verify, the size of the prize, punishment for receivers not acknowledging her contribution, and the extent her preliminary results solve the problem, the equilibrium can involve conference presentation and acknowledgement or the refusal to present.

Our empirical analysis uses as the dependent variable responses to six questions in the survey on willingness to share. Three of the questions relate to specific sharing and three to general sharing. Separate empirical models are estimated for each of the sets of questions. 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. The econometric analysis shows clearly that the determinants of specific sharing are very different from the determinants of general sharing. Among the statistically significant coefficients, only a measure of competition that is insignificantly different in the two models. In addition, the empirical differences are, in general, the differences implied by the game theoretic models.

The insights from our games contribute to the theoretical literature on information exchange and disclosure of research results, which with few exceptions has focused on exchanges among firms (Anton and Yao 2002, 2004, Baker and Mezzetti 2005, Lerner and Tirole 200?, Gill 2008, Hellmann and Perotti 2007, Stein 2008). Although they focus on a different context, Hellmann and Perotti (2007) 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 scientists have a positive probability of solving the complete problem. In our model, sharing results with complementors has the upside of feedback but it also has the negative effect of increasing their chance of winning.

There is an emerging theoretical literature on information exchange in academia which focuses on an economic analysis of publication. Mukherjee and Stern (2009) examines the trade-off between disclosure through publication and secrecy, and show that the feasibility of open science as an equilibrium depends on the costs of future researchers accessing information and the relative benefits of secrecy. Closer to the spirit of our research, several papers have examined the impact of academic misconduct on research and publication decisions (Hoover 2006 and Lacetera and Zirulia 2008). Our work differs in its focus on sharing during the research process. Although we do not endogenize verification in our general sharing model, our approach borrows heavily from the intuition of Lacetera and Zirulia (2008).

While information-sharing among academic scientists *per se* has not received much attention in the empirical literature, significant withholding has been documented (e.g., Blumenthal et al 1996, Campbell et al. 2002). The main factors that have been identified as influencing sharing include the cost, involvement in business activities, protecting the ability of students to publish, and scientific competition (Campbell *et al.* 2002, Walsh and Hong 2003, Walsh *et al.* 2007). For both industry and academic scientists Haeussler (2009) has found expected reciprocity and the extent to which scientists perceive that their community adheres to the scientific norm of communalism to be important. These studies concentrate on sharing in the specific context, where scientists have received requests for materials. We consider both this case and the more proactive one of general sharing of research progress in order to gain visibility and receive feedback. By considering both specific and general sharing we have been able to highlight important differences in the context. In addition, our theoretical results provide hypotheses for our analysis which have not examined before, such as specific sharing at various stages of scientists' career cycles.

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 5 concludes with a discussion of the findings and derives implications for public policy.

2 Games of Information Sharing

2.1 Specific Sharing

We first consider a simultaneous move game in which two scientists working on the same problem must decide whether or not to share research materials or results with each other. In this game, we abstract from issues related to possible misappropriation and focus only on the effect of sharing on the probability of winning and the value of information. There is a prize, W, for solving the research problem and each scientist has intermediate results useful for solving the problem. A prize, such as publication, a Field Medal or Nobel Prize could have academic value, or it could have commercial value, such as a patent, or it could have both.

If either scientist shares her/his results with the other, but the other does not share, the scientist receiving information gains an advantage in the competition. Suppose in the absence of sharing, $z \in (0, 1)$ is the probability that scientist 1 wins the prize and $1 - z$ is the probability that scientist 2 wins; then if scientist 1 shares but 2 does not, the first scientist's probability of winning is reduced to $q \in (0, 1)$, and the other scientist's probability of winning increases to $1 - q$. For simplicity, we assume that it is the probability of winning relative to each other that matters. Thus z is the probability of scientist 1 winning when they both share the information and when neither shares; and q represents the probability of winning for a scientist who shares when the other scientist does not.

2.1.1 The Stage Game

Figure 1 presents a single stage of this game in both extensive and normal form.Scientist 1's expected payoff is given on the top line of each bracket or cell and scientist 2's is given by the bottom line. Scientist i has research materials or intermediate results represented by $r_i \geq 0$ $(i = 1, 2)$. The ability of scientist j to exploit materials shared by i is represented by $e_j \geq 0$, so that the value to scientist j of materials shared by i is given by $e_j r_i \geq 0$. Admitting the possibility that the two scientist's differ in their ability to exploit shared information will allow us to derive comparative static results of interest for scientists in different labs. For simplicity we delete the value of a scientist's own research results from her payoffs because it does not affect the relative returns to each strategy.

If sharing reduces a scientist's probability of winning when the other does not share $(q < z)$ and not sharing when the other does increases the scientist's probability of winning $(1 - q > z)$, then there is a gain to both scientists from not sharing. That is, the gain to the scientist 1 from not sharing is

$$G_1(NS) = \left\{ \begin{array}{ll} (1 - z - q)W & \text{if scientist 2 shares} \\ (z - q)W & \text{if scientist 2 does not share.} \end{array} \right\} \quad (1)$$

Similarly, if scientist 2 decides not to share, he gains

$$G_2(NS) = \left\{ \begin{array}{ll} (z - q)W & \text{if scientist 1 shares} \\ (1 - z - q)W & \text{if scientist 1 does not share.} \end{array} \right\} \quad (2)$$

Thus not sharing is a dominant strategy for each scientist and the unique Nash equilibrium is that neither scientist shares her/his materials. Even though misappropriation of results is not a risk, sharing is not an equilibrium outcome for the stage game. Nonetheless, sharing by both scientists Pareto dominates the Nash since $zW + e_1 r_2 > zW$ and $(1 - z)W + e_2 r_1 > (1 - z)W$. This game of specific sharing is thus 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 scientist who denies a request for materials today may find herself desiring materials from the other scientist in the future. Thus it is more natural to consider scientists' 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 scientists have a common probability distribution over the length of the repeated game. This structure will allow us to consider how the stage of scientists' careers affects their decisions to share. For example, the expected horizon for untenured faculty is likely to be different than midcareer scientists with tenure.

Thus we consider a game of unknown, but finite, length T,in which the scientists assign a probability $p_t \geq 0$ to the game ending in period t.We consider trigger strategies, in which each scientist shares as long as the other has shared but once the other scientist refuses to share, she refuses to share in subsequent periods. In deciding whether to share in period t, scientist i weighs her gain against her expected loss if the game continues and scientist j does not share in future periods. In order for sharing to be an equilibrium, the expected loss to each scientist from punishment (the inability to gain access to the other's materials in the future) 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 scientists place on their loss from not obtaining materials in the future.

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

$$\max_i \left\{ \frac{\varphi_i(NS)}{\pi_i(NS)} \right\} \leq E[T \mid T \geq t] - t \tag{3}$$

where $\varphi_i(NS) = \max_i G_i(NS)$ and $\pi_i(NS) = e_i r_j$. An equilibrium involving sharing exists when the scientists 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 scientists must expect for such cooperation. Arribas and Urbano (2005) characterize the expected time of play in terms of behavior in the limit of $\left\{ p_{t_l} \right\}_{l=1}^{\infty}$, a subsequence such that $p_{t_l} > 0$. They show that when $\lim_{l \to \infty} \frac{p_{t_{l+1}}}{p_{t_l}} = 0$, then the conditions in (3) can't hold, while for the opposite extreme, $\lim_{l \to \infty} \frac{p_{t_{l+1}}}{p_{t_l}} = 1$, trigger strategies of any duration will support cooperation. The more interesting case is when $\lim_{l \to \infty} \frac{p_{t_{l+1}}}{p_{t_l}} = \alpha$ for $(0 < \alpha < 1)$ so that the expected time of play is $\frac{\alpha}{1-\alpha}$. In this case, whether (3) holds depends on the underlying parameters of the stage game. In particular, applied to our game, sharing for some number of periods can be a subgame perfect equilibrium when, either

$$\alpha > \max_i \left\{ \frac{\varphi_i(NS)}{\pi_i(NS) + \varphi_i(NS)} \right\} \tag{4}$$

or there is an interger $l_o > 0$ such that

$$\frac{p_{t_{l+1}}}{p_{t_l}} > \max_i \left\{ \frac{\varphi_i(NS)}{\pi_i(NS) + \varphi_i(NS)} \right\} \text{ for all } l > l_o \text{ and } \alpha = \max_i \left\{ \frac{\varphi_i(NS)}{\pi_i(NS) + \varphi_i(NS)} \right\} \tag{5}$$

and $\max_i \{\varphi_i(NS)/\pi_i(NS)\}$ is the minimal duration of punishment that will support sharing in equilibrium.

Proposition 1 *Consider the probabilistic horizon game described above in which conditions* (4) *and* (5) *characterize the existence of sharing for some length of time as a subgame perfect equilibrium. Then*
(i) The likelihood of sharing in equilibrium increases with a decrease in W or an increase in α,q,e_i, r_j for $i \neq j$.
(ii) The likelihood of scientist i sharing in equilibrium increases with z if $z \leq 0.5$ and decreases with z if $z > 0.5$.

Proof. Clearly (4) and (5) are more likely to hold the higher is α.Define $\overline{\alpha} = \max_i \left\{ \frac{\varphi_i(NS)}{\pi_i(NS)+\varphi_i(NS)} \right\}$.Given α, the conditions in (4) and (5) are more likely to hold the lower is $\overline{\alpha}$. It is straightforward to show that $\partial\overline{\alpha}/\partial W > 0$ and $\partial\overline{\alpha}/\partial x < 0$ for $x = q$, e_i , r_j for $i \neq j$ which proves (i). Similarly, $\partial\overline{\alpha}/\partial z < 0$ if $z \leq .5$ and $\partial\overline{\alpha}/\partial z > 0$ if $z > .5$. ■

The higher is the size of the prize, W, the less likely are the two scientists to share. This is because the maximum single period gain to each scientist from not sharing increases with the size of the prize. On the other hand, the single period loss to scientist i during periods of punishment is increasing in both r_j, the amount of material that scientist j has to share, and her own ability to exploit the materials, e_i

With an increase in α the expected length of the game increases so that the weight the scientists attach to future punishment increases. That is the expected number of periods in which they can be punished increases. To understand the result for q, recall that it represents the probability of winning for a scientist who shares when the other does not. Thus, unambiguously, as q increases, the single period gains to each scientist from not sharing fall.

Recall that z is the probability that scientist 1 wins when 1 and 2 take the same action and $(1 - z)$ is the probability that scientist 2 wins when they take the same action. Thus, all else equal, the likelihood of sharing is greatest when the scientists are equally matched (i.e., $z = .5$). As z deviates from .5, the scientists are less equally matched in the competition. The scientist with the advantage is increasingly less willing to share as her probability of winning increases, while the scientist with a disadvantage is more likely to share as her probability of winning increases. Thus when $z > .5$, scientist 1 has the advantage so that an increase in z decreases the likelihood she will share because her gain from not sharing increases (i.e., $\varphi_1(NS) = \max G_1(NS) = (z - q)W$. When $z < .5$, scientist 1 has the disadvantage so that an increase in z increases the likelihood she will share because her gain from not sharing decreases (i.e., $\varphi_1(NS) = \max G_1(NS) = (1 - z - q)W$.

2.2 General Sharing

In this section, we focus on sharing in a different context: conference or workshop presentation of intermediate results from ongoing research. Again scientists face conflicting incentives. Presentation of preliminary results allows a scientist to gain feedback, but if the presentation provides useful insights for others it may increase the probability that someone else beats her to completely solving the problem. It 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 present a simple sequential move game in which misappropriation is possible. Our interest is in the conditions under which preliminary work is presented and appropriately acknowledged as an equilibrium outcome.

Suppose there are M scientists trying to solve the same research problem and as before W is the prize for the solution. To distinguish this situation from one in which a scientist considers sharing research materials, we represent the portion of the problem Scientist 1 has solved as, $\sigma \in (0,1)$. The $M-1$ other scientists are trying to solve the same problem, but none has completely solved it; if any scientist has totally solved it the game ends.

We consider the decision of a single scientist, scientist 1, who is deciding whether to present her results to the entire community in an effort to get credit σW for her progress. We let $\gamma \in (0,1)$ be the probability that a randomly chosen scientist has solved a different part of the problem, and call that scientist a complementor. Then $\lambda = .1-(1-\gamma)^{M-1}$ is the probability that at least one of the $M-1$ scientists is a complementor. Presenting to a complementor has two effects. It allows for feedback from the complementors, which we represent as adding τ to W, but it also reduces her probability of winning. We denote scientist 1's probability of winning as $x \in (0,1)$ if she does not present or presents and there are no complementers and $(x-\delta) \in (0,1)$ if she presents to at least one complementor If she presents to an audience without any complementors, she gets neither feedback that increases her potential prize 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, scientist 1 chooses between sharing her results (P) with the community or not (NP). If she presents, she makes the $M-1$ scientists aware of her progress (e.g., they can attend her presentation or access her working paper). Scientists obtain information from 1, and all (including scientist 1) continue working on the problem. In stage two, nature decides which scientist first solves the problem. If the winner is not scientist 1, he decides whether to acknowledge 1's work (A) or not acknowledge it (NA). If the winner acknowledge 1's work, he earns only partial credit, $(1-\sigma)W$. If the winner does not acknowledge 1's work, he earns the full credit of W. But with probability v one of $M-2$ scientists, observing both the winner's work and scientist 1's, will verify that the winner has used 1's idea without acknowledging it. In this case, the winner suffers a loss of reputation R and earns no credit. We denote scientist 1's belief that a randomly chosen scientist will provide verification as

$\rho \in (0,1)$ and assume that the M scientists share this belief. Then we can write each scientist's belief that at least one of the $M-2$ (other than 1 and the winner) as $v = 1-(1-\rho)^{M-2}$

Consider the winner's decision. Whether acknowledging scientist 1's work is in his interest depends on the probability another scientist verifies the originality of the his work, the reputational loss misappropriation is verified, and how much of the problem was solved by scientist 1; that is, acknowledgement is worthwhile for the winner if

$$v > \frac{\sigma W}{R+W}. \tag{6}$$

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 sicentists 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, this condition in (6) can be rewritten as

$$M > \frac{\ln\left(1-\frac{\sigma W}{R+W}\right)}{\ln(1-\rho)} + 2 \tag{7}$$

Thus, one of the implications of the model is that if only two scientists are working on the problem, then the winner will never acknowledge scientist's 1's work. This the only reason that scientist 1 would present is for feedback since credit for her progress will not be forthcoming. Although we have not examined scientist 1's decision yet, we will find that unless there is a third person who can verify, she will not present unless there is sufficient feedback. If $\tau = 0$ and $M = 2$, the unique equilibrium of this model is (NP, NA). Thus verification and feedback are signficant differences between our specific and general sharing games.

In making her decision, scientist 1 considers these two factors (verification and feedback) but also the impact of presenting on her likelihood of winning. Her expected utility from presenting is

$$U^P - U^{NP} = (1-x)\sigma W C + \lambda\tau + \lambda(\sigma C - 1)\delta W \tag{8}$$

where $C = \Pr(A) + v\Pr(NA)$ is the probability that she will receive credit regardless of whether or not the winner acknowledges. The first term on the right hand is the announcement effect and reflects the credit she hopes to get from presentation. 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 presentation improves their chances of winning. This fits our intuition that for presenting to dominate not presenting, the effects of announcement and feedback need to outweigh the negative impact from complementors among the M scientists.

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

$$C > \frac{\lambda[\delta W - \tau]}{[(1-x) + \lambda\delta]\sigma W}. \tag{9}$$

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, *and* R *and decreasing in* σ. *It is increasing in* W *if* $v > \sigma$.
(ii) The likelihood that scientist 1 *will present 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.*

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

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 scientist will verify the role of the presenting scientist'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 (6) thus increasing the likelihood that the winner will acknowledge the presenting scientist's contribution. On the other hand an increase in σ, the portion of the problem that scientist 1 has solved, increases the right hand of (6).

The results in (ii) highlight the conflicting effects of presentation. An increase in feedback, τ, increases the positive effect from presenting to complementors, while increases in W, x or δ increase the potential loss from presenting to 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 presentation in equilibrium is more likely. Finally, increases in ρ, and R both increase the probability that she will receive credit, δW, whether or not the winner, if not herself, acknowledges her contribution.

3 Econometric Analysis

As noted in the introduction we exploit the findings of a unique survey of public sector bio-scientists' willingness to share. The scientists are employed in a university or a public research organization in either the United Kingdom or Germany. Industry scientists are excluded since their willingness to share relate to additional motives not found among public sector scientists (see, for example, Haeussler 2009). We exclude questionnaires from scientists who were older than 65 years. The final sample has 1173 observations that met our criteria (approximately 21% are employed in the United Kingdom). Survey details are found in the Appendix.

Of greatest importance to the present study is a series of six questions regarding a scientist's willingness to share information. The questions are provided in Table 1 along with our shorthand notation for the questions. Willingness to share is measured on a five-point Likert scale ranging from disagree strongly to agree strongly. With the exception of the first question, agreement implies some degree of unwillingness to share. For purposes of the analysis responses are coded so that that higher scores imply a greater willingness to share or less restrictions on sharing. Thus, for question one "agree strongly," which implies a willingness to share, receives a value of 5 and "disagree strongly" receives a value of 1. The opposite coding is used for the other questions so that, for example, disagree strongly, which implies a strong willingness to share, is coded as a 5.

The six sharing questions fall into the two distinct types of sharing discussed above. One group is composed of questions regarding specific sharing; these are questions 4, 5 and 6 in Table 1. Questions 1, 2, and 3 in Table 1 address general sharing. Arguably, question 2, $Withhold$, could refer to a specific sharing situation and question 5, $ExpectFeed$, could refer to a situation of general sharing. Initially, we use the separations in Table 1; in our robustness checks we account for alternative interpretations for questions 3 and 5.

Summary statistics are in Tables 2 and 3. The correlations in Table 3 are all significantly different from zero at a 1% level with the exception of the correlation between $Delay$ and $ExpectFeed$ which is not significant at standard levels. Of particular note is the fact that the largest simple correlation in Table 3 is less than 0.5 suggesting that the six questions address distinct issues within and across different types of sharing.

3.1 Specific Sharing

We start with an ordered logit model to explain the responses provided for the three specific sharing questions. Each respondent can provide up to three responses regarding specific sharing, and our econometric approach is to "stack" responses to the three questions so that we consider a single econometric model explaining Likert scores for the 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 three observations (assuming that an answer is provided for each sharing question). The second person provides observations 4 through 6, etc. Since each respondent can appear in the data up to 3 times, we use cluster standard errors to account for within individual correlations across the disturbances.

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.

According to Proposition 1 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 is the competition in their field. *Competition* takes on the value respondents attach to the level of competition in their field where higher values indicate greater competition. A second measure of the size of the prize is the extent to which respondents believe the reward structure of open science operates in their field. Respondents were asked to rate, on a five-point Likert scale, to what extent they agree that the first to come up with new research results is highly esteemed among peers. Higher values of $FirstEsteemed$ indicate greater esteem.

The ability of scientist i to exploit information received from another researcher, e_i, is also unobserved, but we argue that the size of the researcher's team should be positively correlated with the unobserved e_i. 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. $TeamSize$ is the number of scientists with an academic degree who are currently working in the respondent's research group.

The longer the length of time the scientists expect the game to continue, α, the greater the likelihood of sharing. Given the structure of the survey this only can be captured for the scientist of whom the request is made. We include the age of the scientist, Age, as well as the square of Age ($AgeSq$). The square is included to allow for non-linear age effects. 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). 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. In addition, an increase in age reduces the number of periods in which scientists can be punished for non-sharing. Studying sharing in the context of a specific, identified request, Haeussler (2009) finds that older scientists are less like to share information.

Additional regressors include the number of full time employees, *Responsible*, 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.

We include a variable to capture respondent's beliefs that the norms of science operate in their field. Respondents were asked to rate, on a five-point Likert scale, to what extent they agree that open exchange of information is usually being practiced among researchers. Higher values of $OpenExchange$ indicate more openness is practiced. Haeussler (2009) reports for a sample of scientists in academia and in industry, that the likelihood of sharing information with an inquirer increases by the extent to which scientists perceive that their community adheres to the scientific norm of communalism.

Three regressors are used to capture the respondent's research profile. First, *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. Second, Respondents 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. 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 include three regressors that capture what might be referred to as academic entrepreneurship. *Patents* is the number of technically unique patent applications (not double-counting congeneric patents at distinct national patent offices) that respondents claim list them as an inventor. *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), Walsh *et al.* (2007) report that academic scientists involved in business activities show a lower willingness to fulfill an information request than scientists 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. Scientists 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 (2009) indeed finds that scientists 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 with regard to 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 more likely deny a request for information. Haeussler (2009) 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 scientist 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.

Finally, we include an indicator variable $NotPass$ if the question is question 4 in Table 1 and an indicator variable $ExpectFeed$ if the question is question 5 in Table 1. $ExpectFutInfo$ is the omitted category.

Results for the specific sharing questions are Panel A of Table 4. The variables associated with the size of the prize, *Competition* and $FirstEsteemed$, have the expected negative signs and both are significantly different from zero. The ability of a scientist to exploit information received from another researcher is measured by $TeamSize$. Our theoretical model predicts a positive coefficient for $TeamSize$; that is, larger teams are associated with greater sharing. The coefficient of $TeamSize$ is both positive and significantly different from zero.

The time horizon is captured by *Age*, *AgeSq,* and $Professor$. *Age* and *AgeSq* are neither individually nor jointly different from zero and, hence, do not provide support for Proposition 1. We note that this may be due to the fact that the average age of respondents is a fairly young 46; only 15% of respondents are older than 55 and only 6% are older than 60. The insignificance of the age variables may be a matter of not enough observations on scientists 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, but it is not 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. Hence, the likelihood of scientists' specific sharing increases when the community is perceived to follow the norm of communalism.

Earlier we noted that the question $Withhold$ could arguably be included as a specific sharing question. We included this question as specific sharing and the results are given in Panel B of Table 4. Compared to our base model, the coefficient of *Responsible* is now significant, while the coefficient of $FirstEsteemed$ is no longer significant but it has the expected sign. We further noted that $ExpectFeed$ could arguably be a general sharing rather than a specific sharing question. Results when only $NotPass$ and $ExpectFutInfo$ are treated as specific sharing questions are in Panel C. Compared to the base model the coefficient of *Responsible* is significant and the coefficient of $FirstEsteemed$ is not significant. The fraction of time spend consulting is negative and significantly different from zero. 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). In our final robustness check we included the square of the size of the team, $TeamSq$, and results are in the final paneo of Table 4. $TeamSize$ and $TeamSq$ are not individually significantly different from zero but they are jointly different (p-value = 0.056). Results do not change in any meaningful way from the base case.

3.2 General Sharing

As in the specific sharing econometric model, we stack the three general sharing questions (questions 1, 2 and 3 in Table 1) to form a panel and then use those responses as the dependent variable in an ordered logit model with cluster standard errors. We include independent variables that capture information about life cycle or career stage attributes, the scientific team, the research profile, entrepreneurship, and attitudes about the external research environment, as well as some demographic effects.

According to Proposition 2 the number of scientists, M, working on the problem increases the likelihood of verification which, all else equal, increases the likelihood of presentation. An increase in M also increases the chance that at least one complementor is a part of the audience. This has conflicting effects. An increase in M will increase the chance a scientist will present, if the positive effect of feedback from complementors outweighs the potential loss from her giving them her part of the solution. In our data we do not have measures that would allow us to control for these latter effects, so that, empirically, the effect of an increase in M is ambiguous. We do not have a direct measure of M but it should be positively correlated with the level of competition, *Competition*, in the field. However, as noted above, the level of competition is expected to be positively correlated with the size of the prize W. According to Proposition 2 (*ii*) the effect of an increase in W decreases the likelihood that a scientist will present her work. Consequently, the overall effect of a change in *Competition* is ambigious.

A second measure of the size of the prize is $FirstEsteemed$ which measures the extent to which the respondent believes that greater esteem is attached to the first to discover. According to Proposition 2 (*ii*) the effect of $FirstEsteemed$ should be negative.

The probability that a given scientist will verify is ρ and our measure is the respondent's belief that the norms of science are operative in their field. 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). Higher values of *OpenExchange* indicate more openness and verification is practiced so that higher values should be associated with more sharing.

R is the loss or penalty for misappropriation and higher values lead to greater sharing. In the survey respondents were asked 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.

The other controls included in the model are the demographic variables UK, *Age*, *AgeSq*, $Professor$, $Married$ and $Male$. Team effects, *TeamSize* and *Responsible*, are also included. Three regressors are used to capture the respondent's research profile (*Publications*, *Basic* and *OwnResearch*) and four regressors are included to measure academic entrepreneurship (*Consult*, *FamilyEnt* and *Patents*).

Finally, we include indicator variable $Withhold$ and *Delay* if the questions are 2 and 3 in Table 1, respectively. *PresentUnpub* is the omitted category.

Results for the general sharing questions are in the first output columns in Table 5. The coefficient of *Competition* is negative and significant. This is consistent either with the effect of W, the size of the prize, outweighing a positive effect of M when feedback effects dominate or with the feedback effect in M being dominated by the potential loss from presenting to complementors. $FirstEsteemed$ has the expected negative sign but it is not statistically signficantly different from zero. *OpenExchange* has the anticipated positive coefficient, and it is signficantly different from zero. *ExploitLose* has a counterintuitive negative and signficant (10% level) coefficient. It may be the case that *ExploitLose* also captures whether respondents have witnessed idea theft. Those who have observed idea theft may be more concerned about idea theft and are therefore less likely to share.

As was the case with the specific questions, those who conduct more basic research are more willing to generally share. *Responsible* has the expected negative coefficient which is significantly different from zero. $Patents$, $Consult$ and $FamilyEnt$ have the expected negative signs and each is significantly different from zero. The time horizon is captured by *Age*, *AgeSq*, and $Professor$; none are significantly different from zero. Men share less than women.

Question 3 in Table 1, *Delay*, is quite different from the other general sharing questions in that it refers to a delay in sharing for the purpose of securing patent protection. We dropped the *Delay* responses and re-estimated the model. Results are in Panel B of Table 5. Results are very similar to the base model except that the coefficient of *Publications* is now positive and significantly different from zero and $Consult$ and $FamilyEnt$ are no longer significant.

Earlier we noted that the question $ExpectFeed$ could arguably be included as a general sharing question. We included the question as a general sharing question and results are in Panel C. The results are very similar to the base model results reported in Panel A except that $FamilyEnt$ is not longer significant. the coefficient of *Responsible* turns significant, while the coefficient of $FirstEsteemed$ looses significance. We further noted that $WithHold$ could arguably be a specific sharing question. Results when only *PresentUnpub* and *Delay* are included as general questions are in Panel D. Results are very similar to those of the base model except that $TeamSize$ is now negative and significantly different from zero. Finally, in Panel E we have added $TeamSq$ to the base model. $TeamSize$ and $TeamSq$ are jointly significantly different from zero (p-value = 0.0); otherwise the results are very similar to the base model.

3.3 Are the Models Different?

A comparison of the specific model results in Table 4 with the general model results in Table 5 suggests that the two forms of sharing are empirically quite different. In order to test for differences in the coefficients we merged the specific data with the general data in a single regression model. All regressors from the base models are included along with each of those regressors interacted with an indicator variable *General* equal to 1 if the observation is a general sharing question, and it is 0 for specific questions. An ordered logit model with cluster standard errors is used. Tests of whether the interaction coefficients are significant (noted in the following as "$G\_$" followed by a variable name) will reveal any statistically significant differences between the two forms of sharing.[1] For the sake of parsimony the detailed results are not presented since the combined model does not provide any additional evidence beyond that found in Tables 4 and 5; it only provides a convenient and appropriate mechanism for testing differences across the models.

Before discussing the statistical tests, we use this combined model to ask whether respondents are more likely to share generally than specifically. The estimated model is used to predict the probabilities of each of the five levels of agreement separately for the case of $General = 1$ and for the case of $General = 0$. The average probabilities are found in Figure 3. In that figure, for example, we see that based on the characteristics of a randomly chosen individual the probability that the person will respond with the highest level of sharing is 0.26 if the question is a general sharing one. However, if the question is a specific sharing question then the corresponding probability is only 0.12. Respondents are much more likely to engage in general sharing.

The two types of sharing are significantly different with respect to the coefficient of $FirstEsteemed$ and the coefficient of $G\_FirstEsteemed$ (the interaction of $General$ and $FirstEsteemed$); the sum of these coefficients is not significantly different from zero. Thus $FirstEsteemed$ has a negative and significant effect on specific sharing, while, for general sharing, the effect is negative but the coefficient is not statistically significantly different from zero. $TeamSize$ has a positive and significant effect on specific sharing, but it has a negative and statistically significant effect on general sharing. Moreover, the coefficient of $Responsible$ is not significantly different from zero, but the coefficient of $G\_Responsible$ is significantly different from zero. Thus the number of individuals directly reporting to the respondent has no effect on specific sharing, but it does have a negative effect on general sharing. The coefficient of $Professor$ and $G\_Professor$ are both significantly different from zero. However, the coefficients are nearly identical though opposite in sign and their sum is not significantly different from zero. Thus professors, who are tenured, are more willing than other faculty to engage in specific sharing, but there is no effect on general sharing. Based on the results for $OpenExchange$, our measure for ρ in the general sharing model, we can conclude that the more respondents believe that the norms of science operate in their field the more willing they are to specifically share, and their willingness to share is even greater for general sharing. The coefficient of $Basic$ and G_Basic are both significantly different from zero. The more a respondent is engaged in basic research the more willing she is to specifically share, and the effect is even stronger for general sharing. The academic entrepreneurship variables, $Patents$, $Consult$, and $FamilyEnt$, each have a negative and significant effect on general sharing, but only $Patents$ is significantly related to specific sharing and its effect is also negative. $Patents$ has a negative effect on specific sharing, and its effect general sharing is both negative and statistically significantly larger than the specific sharing effect.

[1] Simple tests of significant differences between the coefficients in Tables 4 and 5 would not account for the non-independence of the two panels. Non-independence with the merged data is handled using the cluster standard errors.

The two types of sharing are not significantly different with respect to *Competition*, *Age*, *AgeSq*, *Publications*, *OwnResearch* and *Married*. Of course, with the exception of *Competition*, none of these is significant in the base model results reported in Tables 4 and 5. Coefficients are significantly different for *FirstEsteemed*, *TeamSize*, *Professor*, *Responsible*, *OpenExchange*, *Basic*, *Patents*, *Consult*, *FamilyEnt*, *Male* and *UK*. What is most striking about the comparison is that *Competition* is the only regressor that is significant in one or both models and which statistically has the same effect on both types of sharing.

4 Conclusion

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

The objective of this paper was twofold: first, to provide a theoretical framework for specific sharing which refers to one-on-one situations in which a scientist shares data or material with another scientist, and for general sharing, in which a scientist presents intermediate results to the entire community; and second, to test these frameworks empirically.

Our game-theoretic models of sharing capture some of the main characteristics of the scientific research process and the scientific community. The model for specific sharing suggests that the likelihood of complying with a request for information is negatively related to the size of the prize for solving a research problem and positively related to the value of the inquirers results to date, the ability to exploit shared materials from the inquirer and the probability of the game continuing. Furthermore, all else equal, scientists with similar probabilities of winning the prize are more likely to share. For general sharing, our model 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 scientists 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. Our empirical analysis, in general, supports both models. Among the statistically significant coefficients, only a measure of competition is insignificantly different between the specific and the general sharing model. The empirical differences are, in general, predicted by the theoretical models.

Our theoretical and empirical analysis suggests that information sharing depends on the microeconomic conditions in which the research process is embedded, and these conditions themselves depend upon public policy. Our results imply that an increase in the size of the prize for solving the entire research puzzle decreases the willingness to share information. This is the only common finding that is not significantly different across the two types of sharing and hence the only factor which, when targeted by public policy, has a similar impact on general and specific sharing. As outlined above, we expect the size of the prize to be positively correlated with the level of competition. Hence, some policies aimed at increasing competition among scientists, such as promoting rankings of scientists, can backfire because they result in a reduced willingness to share scientific research results. Instead, a rise in the government funding for research exerts two positive effects with regard to cumulative knowledge production: it increases research and softens competition among scientists.

Furthermore, our results suggest that scientists are more likely to engage in general sharing, *i.e.*, to present intermediate results to the scientific community, when there is a high probability that others will verify work that is not acknowledged. However, although misattribution of authorship and plagiarism occur and are a major problem in science (e.g., Bailey *et al.* 2001, Birnholtz 2006), verification is far from being institutionalized. Enders and Hoover (2004) even report that there is a considerable scope of interpretation regarding the definition of the term "plagiarism". They show that editors of economics journals are evenly split on whether or not the use of an unattributed idea constitutes plagiarism. In addition, if an instance of plagiarism is detected, few journals publicize it. Our results call for mechanisms that encourage verification. For example, plagiarism policies could be institutionalized at the level of journals and public research organizations and designed to reward those who verify and penalize fraudulent scientists. A positive example is the journal *Nature* which has a strict policy to support verification and encourage readers "who encounter a persistent refusal by the authors to comply with these policies [...] to contact the chief editor" (*Nature* 2009). In addition, more journals could provide a platform to publicize instances of fraud as is done by the *German Laborjournal* in the life sciences. Another idea is a contributorship system which which has been debated within the biomedical community. For example, Rennie *et al.* (1997) recommend that papers should acknowledge the work that is done by all contributors, where a contributor is a person who "has added usefully to the work". Such a system might replace or complement the current authorship system. An additional feature of this system is that it softens competition among scientists, and thus, encourages both specific and general sharing, as we have demonstrated.

In addition, our empirical results imply that the stronger beliefs that the norms of science operate the more willing scientists are to specifically share, and their willingness to share is even greater for general sharing. The power of such norms depends on the reward for complying and the punishment for not complying. Institutions that encourage conferences and provide platforms to post and update working papers enable scientists to claim ideas early on and disseminate their research results rapidly. These forums help by establishing a type of "priority" long before a final paper has made it through the peer review process of a scientific journal. Technological advances in communications and information technology provide support for this, as well as for verification (Couzin-Frankel and Grom 2009). For example, starting a blog might become an attractive and accepted way to get ideas out early, claim priority and gain potential feedback to solve the entire research puzzle.

Nonetheless, in the absence of any binding requirement, sharing of information or materials is at the discretion of the scientists themselves. So far, the majority of information and data is not subject to data requirements, sharing is seldom rewarded and usually cannot be enforced. However, where they are possible, requirements can help to move science forward. For example, requirements of funding organizations (among them the National Science Foundation, the National Institutes of Health (NIH), European Research Council) to provide an appropriate and adequate data-sharing plan go in the right direction.

Several limitations of the analysis should be noted. First, regarding our empirical findings, as well as others, which have generally examined sharing of life scientists. Caution should always be exerted when generalizing from a study in a specific context. While the bio-scientific field is a prominent example of a highly collaborative and competitive field, we believe that the microeconomic conditions underlying the decision whether information is shared or not operate in other scientific fields as well. Second, these results are based on survey data. We cannot exclude the possibility that "common method bias" is an issue in this study. However, we completed a large number of pre-tests and validation tests and are therefore confident in our data. Third, given the structure of our survey, we only capture the characteristics for the scientist to whom the request was made. An analysis based on a specific, identified request, could also provide valuable insights into the characteristics of the inquirer and in addition allows better capturing the time horizon of one-on-one sharing.

This paper provides theoretical and empirical insights into the factors that influence the willingness of scientists to share research results. In this respect, our analysis advances our understanding of specific and general sharing and points to subtle differences. With this paper, we hope to encourage discussion and policies that facilitate the specific and general type of information-sharing.

5 References

Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific Teams and Institutional Collaborations: Evidence from U.S. universities, 1981-1999. *Research Policy*, 34(3), 259-285

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

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

Bendor, J., P. 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. J*ournal of the American Society for Information Science and Technology* 57 (13), 1758-1770.

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

Campbell, E., B. Clarridge, M. Gokhale, L. Birenbaum, S. Hilgartner, N. Holtzman,D. 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, No. 6, 467-487.

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: G. Feiwel (Editor), *Arrow and at the Ascent of Modern Economic Theory* (New York University Press, New York).

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

Diaz-Frances, E., S. Ruiz-Velasco, J. Jiminez. 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 Information. Medford, pp. 137-146.

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

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

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. 2009. A Comparison of Information Sharing between Scientists in Academia and Industry. mimeo.

Hellmann, Thomas, and Enrico Perotti. 2005. The Circulation of Ideas: Firms Versus Markets, working paper, University of British Columbia.

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

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

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

Kretschmer, H. 1985. Cooperation Structure, Group Size and Productivity in Research Groups. *Scientometrics* 7, No. 1-2, 39-53

Lacetera, Nicola and Lorenzo Zirulia. 2008. The Economics of Scientific Misconduct, mimeo

Tirole, Jean and Josh Lerner. 2002. Some Simple Economics of Open Source. J*ournal of Industrial Economics* 50, No. 2, 197-234.

Mukherjee, Arijit and Scott Stern. 2009. Disclosure or Secrecy? The Dynamics of Open Science. *International Journal of Industrial Organization,* forthcoming.

Nature 2009. Guide to Publication Policies of the Nature Journals. URL: http://www.nature.com/authors/gta.pdf [accessed on June 12, 2009]

NIH 2003. NIH Data Sharing Policy and Implementation Guidance. URL: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#rev [accessed on June 12, 2009].

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.

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

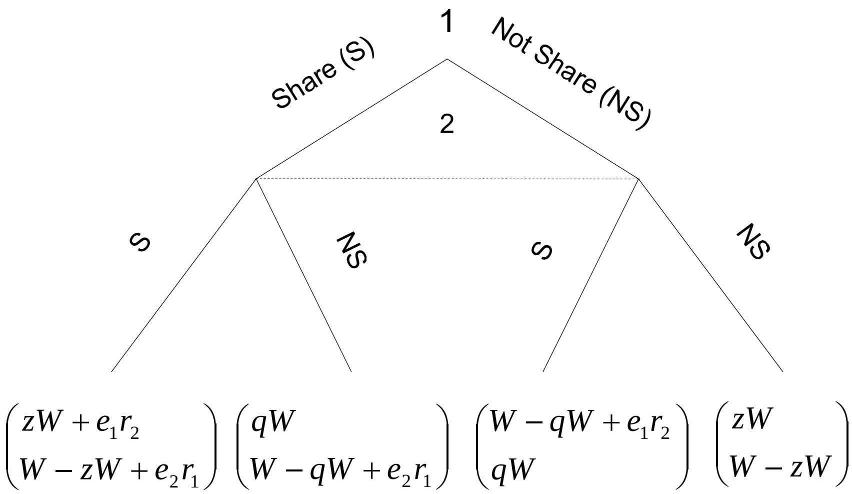
6 Appendix: Survey Design

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, White, Koput, & Owen-Smith, 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 scientists and 8,189 British scientists 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 scientists were invited to participate in an online questionnaire. About 22% of the German and British invitations did not reach the scientists, mostly because of incorrect data in the public databases, and because the addresses of scientists who had left the country or retired had not been updated. Where scientists had changed employers, we asked the former employer for the current address which was provided in about 88% of the cases.

The search categories we used for identifying scientists 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 scientists we pointed out that our target respondents are scientists involved in the bio-scientific field. A total of 2,169 scientists identified through PubMed and 2,452 identified through the European Patent Database filled out our questionnaire. This translates into a response rate of 16% of publishing scientists and 25% of inventors. 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 scientists and 25% for inventors), we ended up with a response rate of 23% in the case of publishing scientists and 33% in that of inventors.



		2: Share	2: Not share
1	Share	$zW + e_1 r_2$ $W - zW + e_2 r_1$	qW $W - qW + e_2 r_1$
1	Not share	$W - qW + e_1 r_2$ qW	zW $W - zW$

Figure 1
Specific Sharing Stage Game

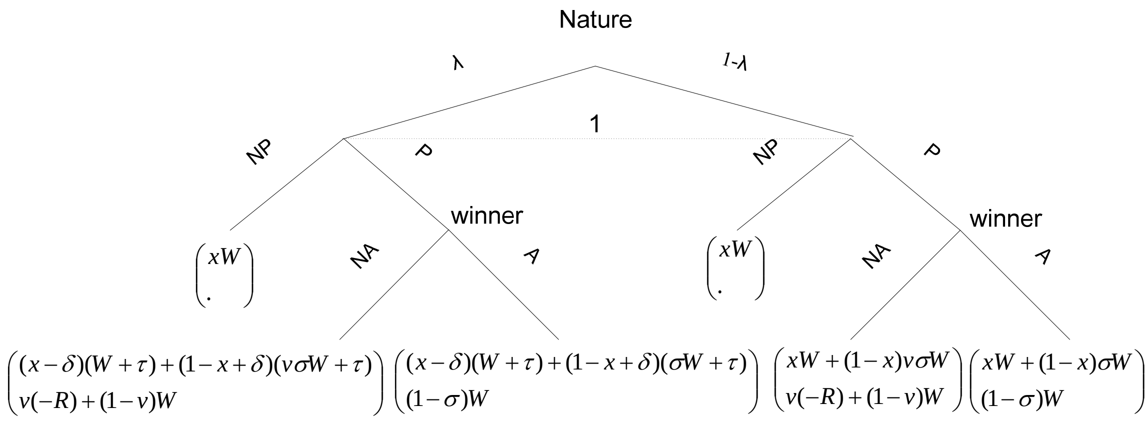


Figure 2
General Sharing Game

Table 1. Sharing Questions

	Question	Question Shorthand	Type of Sharing
1	I present unpublished or yet to be patented research results at conferences.	*PresentUnpub*	General
2	When I discuss unpublished or yet to be patented research results, I often withhold crucial parts	*Withhold*	General
3	In the past I have delayed or had to delay publication of my research in order to secure patenting the research results.	*Delay*	General
4	I only discuss unpublished or yet to be patented research results with people who will for sure not pass on this information.	*NotPass*	Specific
5	I only discuss unpublished or yet to be patented research results with people from whom I expect valuable feedback.	*ExpectFeed*	Specific
6	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

Table 2. Summary Statistics

Variable	No. Obs.	Mean	St. Dev.	Min	Max
Dependent variables					
PresentUnpub	1160	3.559	1.181	1	5
Withhold	1149	3.080	1.170	1	5
Delay	1124	3.679	1.556	1	5
NotPass	1157	2.743	1.155	1	5
ExpectFeedback	1156	2.790	1.089	1	5
ExpectFutInfo	1131	3.308	1.105	1	5
Life cycle or stage of career					
Age	1176	45.964	7.715	29	65
Professor	1176	0.517	0.500	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
Basic	1175	3.962	1.132	1	5
OwnResearch	1172	0.186	0.175	0	1
Academic entrepreneurship					
Patents	1131	2.841	9.324	0	131

Table 2. Summary Statistics (con't)

Variable	**No. Obs.**	**Mean**	**St. Dev.**	**Min**	**Max**
Consult	1172	1.891	4.981	0	80
FamilyEnt	1158	0.242	0.428	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
Other controls					
Married	1154	0.816	0.387	0	1
Male	1172	0.799	0.400	0	1
UK	1173	0.209	0.407	0	1
Subfield controls					
Biochemistry	1176	0.268	0.443	0	1
Bioinformatics	1176	0.050	0.218	0	1
Bioprocess engineering	1176	0.025	0.155	0	1
Cell biology	1176	0.317	0.465	0	1
Clinical medicine	1176	0.163	0.370	0	1
Developmental biology	1176	0.050	0.218	0	1
Genetics/Proteomics	1176	0.152	0.359	0	1
Immunology	1176	0.145	0.352	0	1
Microbiology	1176	0.146	0.353	0	1
Neuroscience	1176	0.190	0.392	0	1
Oncology	1176	0.112	0.316	0	1
Pharmaceutical sciences	1176	0.069	0.253	0	1
Plant sciences	1176	0.069	0.253	0	1
Other	1176	0.264	0.441	0	1

Table 3. Correlations Among Sharing Question Responses*

	PresentUnpub	*Withhold*	*Delay*	*NotPass*	*ExpectFeedback*
Withhold	0.400				
Delay	0.388	0.366			
NotPass	0.430	0.441	0.211		
ExpectFeedback	0.094	0.239	0.039	0.475	
ExpectFutInfo	0.246	0.462	0.202	0.406	0.414

* All are significantly different from zero at the 1% level with the exception of the correlation between *Delay* and *ExpectFeedback*.

Table 4. Specific Sharing Ordered Logit Results

	A. Base Model			B. Base Model Plus *Withhold*			C Base Model Less *ExpectFeed*			D. Base Model Plus *TeamSq*		
Variable	**Odds Ratio**	**t-Stat**		**Odds Ratio**	**t-Stat**		**Odds Ratio**	**t-Stat**		**Odds Ratio**	**t-Stat**	
Competition	0.8850	-2.54	**	0.8744	-2.90	***	0.8913	-2.19	**	0.8830	-2.57	***
FirstEsteemed	0.9088	-1.82	*	0.9243	-1.58		0.9311	-1.29		0.9073	-1.85	*
TeamSize	1.0054	2.52	**	1.0042	1.81	*	1.0062	2.23	**	1.0119	1.49	
TeamSq										1.0000	-1.04	
Age	1.0842	1.34		1.0683	1.14		1.0622	0.9		1.0817	1.29	
AgeSq	0.9992	-1.29		0.9993	-1.20		0.9993	-1		0.9992	-1.24	
Professor	1.3663	2.92	***	1.3273	2.82	***	1.4027	2.97	***	1.3577	2.85	***
Responsible	0.9981	-1.33		0.9971	-2.12	**	0.9966	-2.28	**	0.9979	-1.59	
OpenExchange	1.1988	3.09	***	1.2522	4.18	***	1.2469	3.53	***	1.1960	3.04	***
Publications	1.0002	0.23		1.0008	0.82		1.0007	0.64		1.0000	0.03	
Basic	1.1268	2.36	**	1.1295	2.63	***	1.1509	2.61	***	1.1284	2.39	**
OwnResearch	0.9979	-0.71		0.9998	-0.08		0.9982	-0.56		0.9979	-0.69	
Patents	0.9887	-2.26	**	0.9861	-2.85	***	0.9836	-2.44	**	0.9879	-2.27	**
Consult	0.9925	-1.05		0.9891	-1.55		0.9875	-1.74	**	0.9925	-1.06	
FamilyEnt	1.0570	0.53		1.0244	0.25		1.0295	0.26		1.0526	0.49	
Married	1.2357	1.67	*	1.2315	1.78	*	1.2315	1.53		1.2258	1.6	
Male	0.9395	-0.54		0.8861	-1.12		0.8860	-0.97		0.9414	-0.52	
UK	1.1769	1.37		1.0837	0.72		1.0094	0.08		1.1727	1.34	
NotPass	0.9174	-1.41		0.9206	-1.38		0.3881	-14.2	***	0.9169	-1.42	
ExpectFeed	2.4045	13.68	***	2.3862	13.77	***				2.4050	13.69	***
Withhold				1.6690	7.12	***						
Field fixed effects	Yes			Yes			Yes			Yes		
r-Square	0.0360			0.0334			0.0430			0.0357		
Obs.	3103			4140			2061			3103		

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

Table 5. General Sharing Ordered Logit Results

	A. Base Model			B. Base Model Less			C. Base Model Plus			D. Base Model Less			C. Base Model Plus		
Variable	**Coef.**	**t-Stat**		**Coef.**	**t-Stat**		**Coef.**	**t-Stat**		**Coef.**	**t-Stat**				
Competition	0.8887	-2.50	**	0.8609	-2.58	***	0.8908	-2.86	***	0.9026	-2.03	**	0.8858	-2.56	**
FirstEsteemed	0.9952	-0.10		1.0370	0.64		0.9672	-0.78		1.0066	0.12		0.9929	-0.15	
Teamsize	0.9937	-1.52		0.9950	-1.10		0.9968	-0.83		0.9885	-2.42	**	1.0024	0.24	
TeamSq							1.0000			1.0000			1.0000	-1.14	
Age	0.9855	-0.26		0.9508	-0.77		1.0170	0.35		0.9692	-0.49		0.9820	-0.32	
AgeSq	1.0000	0.00		1.0004	0.52		0.9998	-0.46		1.0002	0.29		1.0001	0.08	
Professor	0.9761	-0.25		1.0798	0.70		1.0364	0.43		0.8939	-1.03		0.9684	-0.34	
Responsible	0.9953	-3.73	***	0.9955	-3.07	***	0.9967	-2.88	***	0.9958	-3.68	***	0.9950	-4.15	***
OpenExchange	1.3048	5.30	***	1.4819	6.59	***	1.2630	5.21	***	1.2576	4.11	***	1.3007	5.21	***
Publications	1.0013	1.30		1.0028	2.37	**	1.0008	0.85		1.0012	1.07		1.0010	1.01	
Basic	1.1990	3.92	***	1.1786	3.14	***	1.1673	3.92	***	1.2338	4.12	***	1.1999	3.94	***
OwnResearch	1.0018	0.67		1.0008	0.26		1.0010	0.44		0.9996	-0.13		1.0019	0.73	
Patents	0.9468	-2.30	**	0.9598	-2.75	***	0.9662	-3.61	***	0.9172	-2.38	**	0.9451	-2.28	**
Consult	0.9770	-2.60	***	0.9907	-1.08		0.9825	-2.49	**	0.9767	-2.27	**	0.9771	-2.56	***
FamilyEnt	0.8251	-1.98	**	0.8967	-0.97		0.8838	-1.51		0.8012	-2.07	**	0.8197	-2.05	**
Married	1.0900	0.84		1.1303	0.98		1.1092	1.11		1.0603	0.51		1.0806	0.76	
Male	0.7389	-3.06	***	0.7682	-2.27	**	0.8071	-2.44	**	0.7374	-2.68	***	0.7415	-3.02	***
UK	0.8777	-1.25		0.9903	-0.08		1.0021	0.02		0.8945	-0.93		0.8717	-1.3	
ExploitLose	0.9270	-1.85	*	0.9301	-1.53		0.9329	-1.98	**	0.9094	-2.10	**	0.9267	-1.86	*
Delay	1.5844	5.32	***				1.5737	5.11	***	1.5725	5.58	***	1.5854	5.32	***
Withhold	0.5112	-11.87	***	0.4447	-11.81	***	0.4967	-12.02	***				0.5115	-11.85	***
ExpectFeed							0.3211	-15.88	***						
Field fixed effects	Yes			Yes			Yes			Yes			Yes		
r-Square	0.0706			0.0605			0.0633			0.0731			0.0712		
Obs.	3063			2058			4093			2038			3063		

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

Figure 3. Willingness to Share: General vs. Specific