Online Appendix
to:
Of Mice and Academics: Examining the Effect of Openness on Innovation

Fiona Murray, Philippe Aghion, Mathias Dewatripont,
Julian Kolev, and Scott Stern

Theoretical Appendix: The Effect of Openness on Follow-On Research

This appendix extends ADS (2008) to consider the specific impact of openness on the composition of follow-on research, comparing the traditional linear view with a model in which researchers have the freedom to pursue their preferred projects, and a single upstream discovery can spawn multiple new research lines.

We broadly define openness as a researcher's ability to access the ideas or materials of other researchers. Openness also allows a researcher to provide access to her own ideas and share them as she sees fit. There are two main approaches to analyze the effects of openness on basic research, and on innovation. The traditional approach stresses the idea that openness should favor more applied research, generally at the expense of more basic research, as it reduces the extent to which early-stage researchers can appropriate the returns from their own research. As developed in ADS (2008), the control rights approach predicts instead that increased openness should foster basic research and the creation of new lines, in particular by reducing researchers' cost of accessing other researchers' ideas, thereby making it more likely that the alternative strategies pursued by researchers with high levels of freedom will actually lead to new lines. This approach emphasizes the complementarity between openness and freedom and the resulting effect of openness on the diversity of subsequent research lines.

Basic set-up

Consider a multi-stage research line. The line starts with an initial idea I_0. Then the idea is elaborated upon in stages. If stage 1 is successful, there is a refined idea I_1; this refined idea can be pursued further to potentially generate an even-more-refined idea I_2, etc. There are a total of k stages after the initial idea. If and only if all k stages are successful, there is a final idea I_k which generates a terminal value V, e.g. in the form of a marketable product.

A manager presides over the research line, and must hire researchers in order to advance it. Each stage requires one researcher, and that researcher succeeds with probability p if she follows a (success-maximizing) "practical" strategy at that stage. Instead of the practical strategy, the researcher may choose to follow an "alternative" strategy which yields a zero probability of success. One interpretation is that the alternative strategy may be the one that the researcher enjoys more, even though it does not pay off in monetary terms. Another interpretation is that the alternative strategy may help initiate new lines but does not generate progress on the initial line.

There is an infinite supply of researchers at each stage, each of whom has an outside option R that she can obtain by working in another profession. After being exposed to idea I_{j-1}, each researcher at stage j decides whether she would better enjoy following the practical strategy or the alternative strategy. If she is able to undertake her favored strategy, she suffers no disutility from working. However, if the researcher has to undertake the strategy that she likes less, she suffers a disutility of z. The ex ante probability that a given researcher prefers to follow the practical strategy is given by α. We further assume that the choice of the practical vs. the alternative strategy is ex ante non-contractible. In other words, one cannot write a contract that promises a bonus for following the practical strategy, because the nature of work that strategy entails cannot be adequately described ahead of time.

If the researcher has control rights over the choice of research strategy (the "researcher-freedom" regime), in equilibrium, she is paid the reservation wage $w^{freedom} = R$, and always works on her preferred strategy. This implies that with probability α, the researcher works on the practical strategy, and with probability $(1 - \alpha)$, she works on the alternative strategy. Therefore, the ex-ante probability of advancing to the next stage is given by αp.

Suppose instead that the manager has control rights, and has full information on the set of possible research strategies. Then, ex-post, the manager has the authority to force the researcher to work on the practical strategy. Anticipating this, the researcher will demand a wage of $w^{control} = R+(1-\alpha)z$ in order to work under this "manager-control" regime. The $(1-\alpha)z$ markup over the researcher-freedom regime represents compensation for loss of creative freedom: the fact that the researcher now must always adopt the practical strategy, whether this turns out to coincide with her preferences or not.

Whether the researcher will or will not enjoy control rights - i.e. de-facto authority - over her research agenda will depend upon whether or not she is monitored by an informed manager. In each stage $i = 1, 2, ...$ the timing of

events is as follows. First, the manager owner hires a researcher, and agrees to pay her a wage of w_i. Next, the manager invests effort in trying to become informed about the project. For an effort cost of $\frac{1}{2}\theta\lambda^2$, the manager has a probability λ of becoming informed. If he is informed, he is then able to force the researcher to follow the practical strategy. However, if the manager is uninformed, he is unable to direct the researcher, who is thus free to pursue her preferred strategy.

Analysis

Define $\Pi_i + 1$ as the value of successfully completing stage i. Given this, the payoff to the manager if he is informed at stage i is:

$$E(\pi_i \mid \text{informed}) = p\Pi_{i+1} - w_i \tag{1}$$

The payoff to the manager if he is uninformed at stage i is:

$$E(\pi_i \mid \text{uninformed}) = \alpha p\Pi_{i+1} - w_i \tag{2}$$

It follows that the unconditional expected payoff at stage i is given by:

$$\Pi_i = E(\pi_i) = (\lambda_i + \alpha(1-\lambda_i))p\Pi_{i+1} - w_i - \tfrac{1}{2}\theta\lambda_i^2 \tag{3}$$

Therefore, the marginal value of being informed at stage i is $(1-\alpha)p\Pi_{i+1}$, and the manager's optimal investment of effort into the probability of becoming informed at this stage is:

$$\lambda_i^* = (1-\alpha)p\Pi_{i+1}/\theta \tag{4}$$

where the equilibrium wage w_i^* is itself determined as:

$$w_i^* = R + \lambda_i^*(1-\alpha)z \tag{5}$$

Further, note that $\Pi_i < \Pi_{i+1}$, so that $\lambda_{i-1}^* < \lambda_i^*$.

The above analysis indicates that as the project moves closer to completion, there is an increasing likelihood that the manager becomes informed and imposes his choice of strategy on the researcher. Thus, we predict that research freedom decreases as the research line advances from more basic to more applied stages. It then follows from (5) that the researcher's wage also increases, to compensate for the fact that she has less de facto creative control. In other words, research organizations endogenously become more authoritarian as research projects move into their later stages.

Alternative strategies create new lines

Thus far, we have assumed that at each stage there is only one economically legitimate research strategy - namely the practical strategy, which has the potential to advance the project to the next stage along the chain. In contrast, the

alternative strategy has been taken to be nothing more than worthless puzzle-solving. Now we modify this assumption. While we keep the restriction that only the practical strategy helps to advance the current line of research, we now allow the alternative strategy to yield new insights which may spawn wholly different lines of research. The interpretation is that when scientists turn away from the applied task of pushing the current line forward, they may not be shirking per se, but rather taking a useful step back that may ultimately generate fundamental breakthroughs. It is in this context that openness and exploration can have a significant impact on the process of multi-stage innovation.

To embed this notion into our model, we proceed as follows. Now, if at any stage of the original research line, a researcher works on the alternative strategy, there is a probability p_r of a revolutionary new idea which will form the basis for $\gamma \geq 1$ entirely new "offspring" research lines. Each of these offspring lines has the same properties as the single lines analyzed above. Moreover, for computational simplicity but without any major loss of insight, we assume that the offspring lines are themselves sterile, and cannot give rise to further generations of revolutionary ideas. That is, revolutionary ideas that yield offspring can only come from the alternative strategy applied at some stage of the original parent line.

The analysis remains essentially the same as before, except that z, the researcher's preference for freedom, must be replaced by $z+p_r\gamma\Omega_0^*$, where $\Omega_0^* = \Pi_0$ is the value of a new line.[1] The higher $p_r\gamma\Omega_0^*$, the more will openness enhance the creation of new lines, because it will reduce the monitoring probability in earlier stages on a research line. In other words, more scope for creating new lines under the alternative strategy tilts the balance towards more freedom for researchers, which in turns spurs the creation of new lines.

The complementarity between openness and freedom

Now, we introduce openness into the picture and show that the above model can be extended to account for the complementarity between openness and (academic) freedom. The basic idea is that when researchers cannot be forced to work on projects they dislike, openness allows a better matching between researchers and positive-NPV research opportunities.

Consider two parallel research lines, 1 and 2, each of which operates as described above. Specifically, with ex-ante probability α the researcher initially allocated to the current stage of either of these two lines prefers to pursue the practical strategy for that line. With complementary probability $(1-\alpha)$ she prefers not to pursue this practical strategy.

We model the introduction of openness by allowing for the possibility that the researcher on line 1 can learn about *and gain access to* project 2, and vice-versa; consequently, with positive probability φ, she may choose to work on the

[1] Note that this result does not depend on whether the manager or the researcher is the one who captures the value of the new research line, as the adjustment of wages will lead to the same impact on monitoring in equilibrium.

practical strategy for project 2 if nobody else does. In this framework, a higher value of φ therefore indicates a greater degree of openness.[2]

The impact of openness is to increase the net present value of a research line operated under researcher freedom from

$$E(\pi_i \mid \text{uninformed}) = \alpha p \Pi_{i+1} - w_i$$

to

$$E(\pi_i \mid \text{uninformed}) = [\alpha + (1-\alpha)\varphi] p \Pi_{i+1} - w_i$$

Importantly, this implies that an increase in openness may well increase the value of starting a new research line, if a significant portion of that research line is performed under a "researcher-freedom" regime. By contrast, openness has no value when an informed manager controls the researcher's agenda, since the researcher is forced to work on the practical strategy. Thus we still have:

$$E(\pi_i \mid \text{informed}) = p \Pi_{i+1} - w_i^p$$

Hence:

$$\Pi_i = E(\pi_i) = (\lambda_i + (1-\lambda_i)[\alpha + (1-\alpha)\varphi]) p \Pi_{i+1} - w_i - \tfrac{1}{2}\theta\lambda_i^2$$

which in turn will lead to a lower equilibrium value of manager investment in information:

$$\lambda_i^{**} = (1-\alpha)(1-\varphi) p \Pi_{i+1}/\theta$$

This reduction in the probability of the manager being informed will expand the researcher's freedom to pursue her preferred research agenda, which leads to two distinct but complementary effects. First, an expanded population of researchers will have the opportunity to move to new projects, above and beyond the direct effects of openness facilitating greater mobility for "already-free" researchers. Second, more new lines are likely to be created when researchers pursue alternative strategies. In equilibrium, this leads to a third prediction: greater openness will result in a higher fraction of researchers working on the early stages of research lines, leading to a greater share of exploratory, basic research.

Main theoretical predictions

Overall we obtain the predictions that openness: *(i) expands the population of researchers that might work on a given research line; (ii) increases the share of early-stage, long-horizon research, and (iii) enhances the creation of new and speculative research lines. In addition, openness can expand the value of new research lines when a significant portion of their development is characterized by researcher freedom.* These predictions form the foundation of our empirical analysis.

[2]It is also possible to capture changes in openness by analyzing changes in access costs for a given technology, and introducing an effort choice by the researcher when attempting to switch from one project to another. As both versions of the model yield the same conclusions on the interaction between openness and freedom, we focus on the simpler formulation to highlight the core of our analysis.

Empirical Appendix: Supplementary Figures and Robustness Tests

FIGURE A1: AVERAGE CITATION RATES BY MOUSE TECHNOLOGY

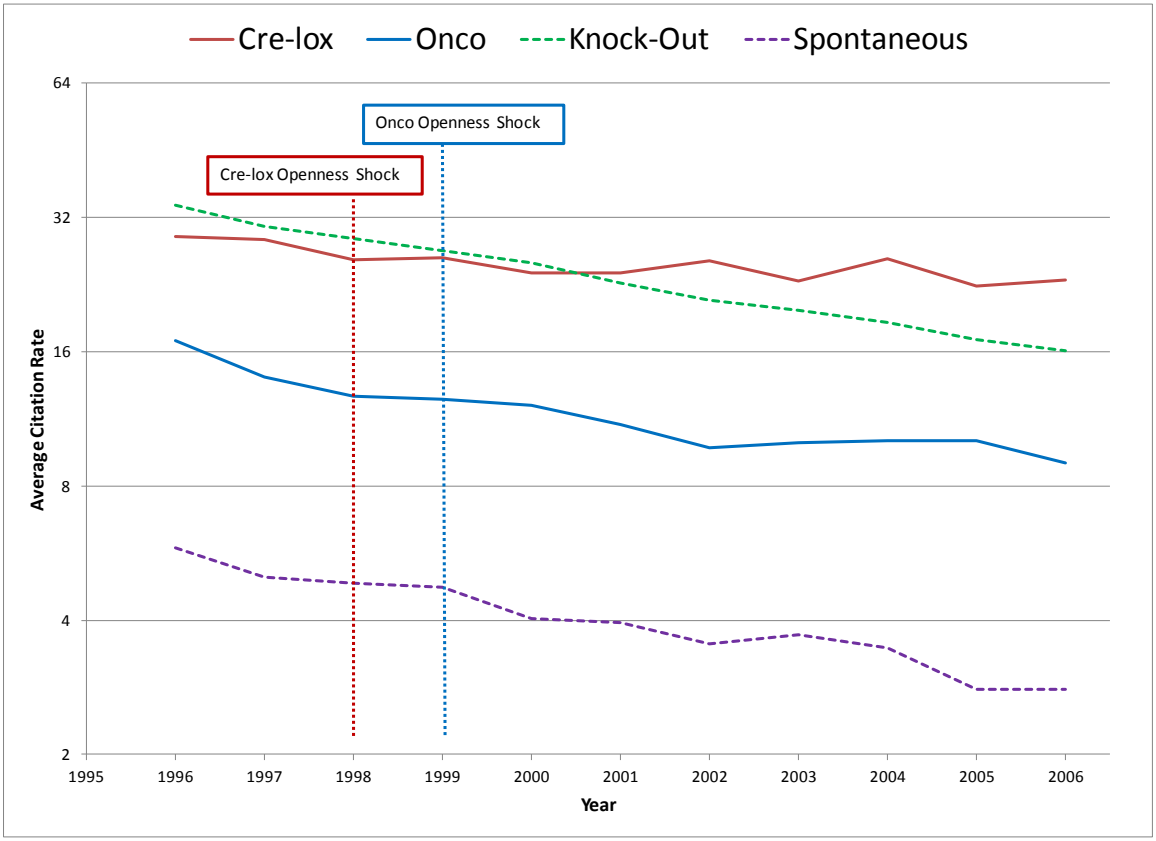

TABLE A1: RANDOM-EFFECTS TESTS FOR A PRE-SHOCK TREATMENT TREND FOR RESULTS ON OVERALL CITATIONS, NEW VS OLD AUTHORS, AND NEW VS OLD KEY WORDS

	[Incidence rate ratios reported in square brackets] Estimated coefficients in 2nd line (Block bootstrapped SEs reported in parentheses)				
	NEGATIVE BINOMIAL	**STACKED NEGATIVE BINOMIAL**			
	(A1-1) DV= Annual Citations With Treatment Trends	**(A1-2a)** DV= New Authors With Treatment Trends	**(A1-2b)** DV= Old Authors With Treatment Trends	**(A1-3a)** DV= New Key Words With Treatment Trends	**(A1-3b)** DV= Old Key Words With Treatment Trends
Post-NIH	[1.115] 0.109 (0.075)	[1.126] 0.119 (0.087)	[1.043] 0.042 (0.107)	[1.083] 0.080 (0.093	[0.824]* -0.194 (0.110)
Treatment Group Age Trend per Year	[1.002] 0.002 (0.013)	[1.005] 0.005 (0.017)	[0.991] -0.009 (0.018)	[1.006] 0.006 (0.018)	[1.002] 0.002 (0.023)
Post-NIH Change in Trend per Year	[1.051]*** 0.050 (0.016)	[1.062]*** 0.060 (0.023)	[1.054]** 0.053 (0.023)	[1.048]** 0.047 (0.023)	[1.040] 0.039 (0.024)
CONTROL VARIABLES					
NIH-Window	[1.200]*** 0.182 (0.061)	[1.207]** 0.188 (0.088)	[1.208]** 0.189 (0.083)	[1.154] 0.143 (0.095)	[0.907] -0.098 (0.087)
Age FEs	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes+		Yes+	
Article Effects	Random	Random		Random	
Log-likelihood	-69519.516	-100090.09		-197234.74	
# of Observations	22,265	44,530		44,530	

Significance levels: * 10% ** 5% *** 1%

\+ Calendar-year fixed effects include a set of indicator variables common to both margins in a given regression, and a linear difference variable which allows for a constant difference in growth rates between the two margins.

Tests of Differences Between Coefficients:

(A1-2): b(Post-NIH effect on New Authors) – b(Post-NIH effect on Old Authors):
Estimate = **0.077***; SE* = **0.075;** *Prob*>|**z**|=**0.308**

(A1-3): b(Post-NIH effect on New Key Words) – b(Post-NIH effect on Old Key Words):
Estimate = **0.274***; SE* = **0.059;** *Prob*>|**z**|<**0.001**

TABLE A2: POISSON SPECIFICATION RESULTS ON OVERALL CITATIONS, NEW VS OLD AUTHORS, AND NEW VS OLD KEY WORDS

	[Incidence rate ratios reported in square brackets] Estimated coefficients in 2nd line (Block bootstrapped SEs reported in parentheses)				
	ML POISSON	**STACKED ML POISSON**			
	(A2-1) DV= Annual Citations	**(A2-2a)** DV= New Authors	**(A2-2b)** DV= Old Authors	**(A2-3a)** DV= New Key Words	**(A2-3b)** DV= Old Key Words
Post-NIH	[1.242]**	[1.334]***	[1.108]	[1.477]***	[1.055]
	0.217	0.288	0.103	0.390	0.054
	(0.096)	(0.080)	(0.095)	(0.079)	(0.075)
CONTROL VARIABLES					
NIH-Window	[1.078]	[1.108]	[1.022]	[1.249]***	[0.962]
	0.075	0.103	0.022	0.222	-0.039
	(0.077)	(0.070)	(0.080)	(0.061)	(0.056)
Age FEs	Yes	Yes	Yes	Yes	Yes
Year FEs+	Yes	Yes+		Yes+	
Article FEs	Yes	Yes		Yes	
Log-likelihood	-59705.8	-89698.7		-274191.2	
# of Observations	22,265	42,802		44,488	

Significance levels: * 10% ** 5% *** 1%

\+ Calendar-year fixed effects include a set of indicator variables common to both margins in a given regression, and a linear difference variable which allows for a constant difference in growth rates between the two margins.

Tests of Differences Between Coefficients:
(A2−2): β(Post-NIH effect on New Authors) – β(Post-NIH effect on Old Authors):
Estimate = **0.185**; *SE* = **0.042;** *Prob*>|**z**|<**0.001**

(A2−3): β(Post-NIH effect on New Key Words) – β(Post-NIH effect on Old Key Words):
Estimate = **0.336**; *SE* = **0.043;** *Prob*>|**z**|<**0.001**

TABLE A3: NON-PARAMETRIC TESTS FOR A PRE-SHOCK TREATMENT TREND FOR RESULTS ON OVERALL CITATIONS, NEW VS OLD AUTHORS, AND NEW VS OLD KEY WORDS

	[Incidence rate ratios reported in square brackets] Estimated coefficients in 2nd line (Block bootstrapped SEs reported in parentheses)				
	NEGATIVE BINOMIAL	**STACKED NEGATIVE BINOMIAL**			
	(A3-1) DV= Annual Citations, Pre-Treatment Sample	**(A3-2a)** DV= New Authors, Pre-Treatment Sample	**(A3-2b)** DV= Old Authors, Pre-Treatment Sample	**(A3-3a)** DV= New Key Words, Pre-Treatment Sample	**(A3-3b)** DV= Old Key Words, Pre-Treatment Sample
Treatment-Specific Year	[0.978] -0.022 (0.054)	[1.012] 0.012 (0.049)	[0.948] -0.053 (0.130)	[1.024] 0.024 (0.070)	[0.989] -0.011 (0.064)
Treatment-Specific Year^2	[0.991] -0.009 (0.014)	[0.994] -0.006 (0.015)	[0.999] -0.001 (0.029)	[1.000] 0.000 (0.016)	[0.999] -0.001 (0.014)
CONTROL VARIABLES					
Age FEs	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes
Article FEs	Yes	Yes		Yes	
Log-likelihood	-5227.3991	-8600.0498		-19089.152	
# of Observations	2812	5366		5620	
# of Mouse Articles	855	820		854	
F-Test on Treatment Trend (P values)	0.74	0.46	0.50	0.74	0.96

Significance levels: * 10% ** 5% *** 1%

TABLE A4: IMPACT OF OPENNESS ON CITATIONS BY NEW-TO-MOUSE, NEW-TO-TECHNOLOGY, AND NEW-TO-SAMPLE LAST AUTHORS

	STACKED NEGATIVE BINOMIAL **[Incidence rate ratios reported in square brackets]** **Estimated coefficients in 2nd line** **(Block bootstrapped SEs reported in parentheses)**			
	(A4-1a) DV= New-to-Mouse, Old-to-Technology Authors	**(A4-1b)** DV= New-to-Technology, Old-to-Sample Authors	**(A4-1c)** DV= New-to-Sample Authors	**(A4-1d)** DV= Old Authors
Post-NIH	[0.861]*** -0.150 (0.054)	[2.081]*** 0.733 (0.073)	[3.071]*** 1.122 (0.054)	[1.298]*** 0.261 (0.066)
CONTROL VARIABLES				
NIH-Window	[0.595]*** -0.519 (0.083)	[1.655]*** 0.504 (0.065)	[3.190]*** 1.160 (0.049)	[1.240]*** 0.215 (0.063)
Age FEs	Yes			Yes
Year FEs	Yes+			
Article FEs	Yes			
Log-likelihood	-175898.63			
# of Observations	88,968			

Significance levels: * 10% ** 5% *** 1%

\+ Calendar-year fixed effects include a set of indicator variables common to both margins in a given regression, and a linear difference variable which allows for a constant difference in growth rates between the new and old margins.

Tests of Differences Between Coefficients:

(A4−1): β(Post-NIH effect on New-to-Mouse Authors) – β(Post-NIH effect on Old Authors):
Estimate = **-0.411**; *SE* = **0.067;** *Prob*>|**z**|<**0.001**

β(Post-NIH effect on New-to-Technology Authors) – β(Post-NIH effect on Old Authors):
Estimate = **0.472**; *SE* = **0.068;** *Prob*>|**z**|<**0.001**

β(Post-NIH effect on New-to-Sample Authors) – β(Post-NIH effect on Old Authors):
Estimate = **0.862**; *SE* = **0.042;** *Prob*>|**z**|<**0.001**