Cite Unseen: Theory and Evidence on the Effect of Open Access on Cites to Academic Articles Across the Quality Spectrum

Online Appendices

Mark J. McCabe
Christopher M. Snyder

This document contains two online appendices, S1 and S2, omitted from the published paper for space considerations.

Online Appendix S1: Extension of Model to Smart Insiders

To recap the model with smart insiders, let $\sigma \in (0,1]$ denote the proportion of insiders that are "smart," able to observe q before accessing the article; the remaining $1-\sigma$ proportion are ordinary insiders who as before only know the distribution of q but not the realized value for the article before accessing it. The sets $B_u^I(a)$ and $B_f^I(a)$ become functions of q for readers who can see q before deciding on their strategy. Some general results for extreme values of q are still available, stated in the next proposition. As in the main text, OA stands for open access.

Proposition 1. *Assume equation (5) holds. For the lowest-quality articles, the OA effect for insiders is the same negative value as for generic readers:* $\Delta^I(0,a_c,a_o) = \Delta(0,a_c,a_o) < 0$. *For the highest-quality articles, the OA effect for insiders is lower than that for generic readers (i.e.,* $0 < \Delta^I(1,a_c,a_o) < \Delta(1,a_c,a_o)$*) unless* $\Delta(1,a_c,a_o) = 0$ *in which case* $0 = \Delta^I(1,a_c,a_o) = \Delta(1,a_c,a_o)$.

Proof. Letting $x^S(q,a)$ and $x(q,a)$ denote the number of cites from smart and ordinary insiders, respectively, we have

$$x^I(q,a) = \sigma x^S(q,a) + (1-\sigma)x(q,a). \tag{S1}$$

Intuitively, no smart insider cites an article with $q = 0$, implying $x^S(0,a) = 0$. This intuition can be verified in Figure 1 by replacing $\bar{q}$ with q and setting $q = 0$. We then see from the figure that $B_n(a)$ occupies the whole measurable space. Substituting $x^S(0,a) = 0$ into (S1) yields $x^I(0,a) = (1-\sigma)x(0,a)$. Thus,

$$\Delta^I(0,a_c,a_o) = \frac{x^I(0,a_o) - x^I(0,a_c)}{x^I(0,a_c)} = \frac{(1-\sigma)[x(0,a_o) - x(0,a_c)]}{(1-\sigma)x(0,a_c)} = \Delta(0,a_c,a_o).$$

We next examine the other extreme of article quality, $q = 1$. Intuitively, all smart insiders cite an article with $q = 1$ since citing unseen provides a benefit with no risk of sanction, so the reader prefers this to ignoring the article. Acquiring the full text instead always generates a cite. This intuition can be verified in Figure 1 by replacing $\bar{q}$ with q and setting $q = 1$. We then see from the figure that (i) is the relevant case, and sets $B_u(a)$ and $B_f(a)$ span the whole measurable

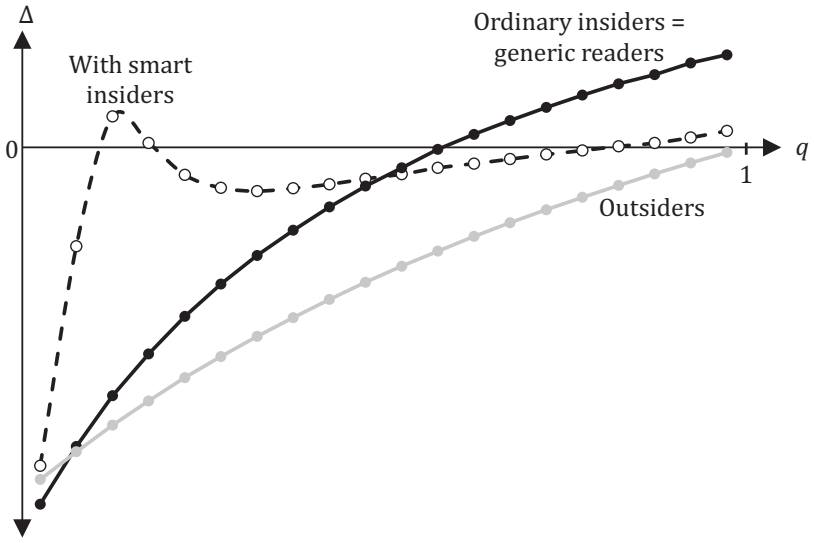

Figure S1: Monte Carlo Exercises Illustrating Smart Insiders. As in Figure 3, dots are averages of the OA effect $\Delta(q,a_c,a_o)$, plotted at the midpoint of each of twenty equal-sized quality bins constructed from 10 million Monte Carlo draws. As in that figure, reader behavior is governed by the insider/outsider model, here extended to allow for smart insiders. The distributions and parameters are the same as in the previous figure except that s has been adjusted slightly, from $s = 0.25$ to $s = 0.3$, to emphasize certain features of the curves.

space. Hence $x^S(1,a) = P^S(B_u^S(a)) + 1 \cdot P^S(B_f^S(a)) = 1 - P^S(B_n^S(a)) = 1$, implying $x^I(1,a) = \sigma + (1-\sigma)x(1,a)$. Substituting,

$$\Delta^I(1,a_c,a_o) = \frac{\sigma+(1-\sigma)x(1,a_o)-[\sigma+(1-\sigma)x(1,a_c)}{\sigma+(1-\sigma)x(1,a_c)} = \frac{x(1,a_o)-x(1,a_c)}{\sigma/(1-\sigma)+x(1,a_c)}. \tag{S2}$$

Since $\sigma > 0$ and $\Delta(1,a_c,a_o) = [x(1,a_o)-x(1,a_c)]/x(1,a_c)$, the last claim of the proposition follows. *Q.E.D.*

Figure S1 provides bin-scatter plots for a new Monte Carlo exercise illustrating possible outcomes of the insider/outsider model. Details behind the exercise are provided in the figure notes. The black curve shows the OA effect for ordinary insiders, equivalent to generic readers in this variant of the model. They experience a negative effect for low q and positive effect for high q. The grey curve represents the OA effect for outsiders. They are formally identical to ordinary insiders except that the sanction has been reduced from $s = 0.3$ to $s = 0$ for them. As expected from Propostion 3, the black curve is everywhere below the horizontal axis and approaches the axis as q approaches 1.

The dashed curve represents the OA effect for insiders when some of them are smart. The sanction for these readers has been returned to the original positive level for regular insiders but now a fraction $\sigma = 0.75$ of them are smart and can see the value of q for articles before acquiring the full text. As expected from Proposition 1, the curve approaches that for the regular insiders as q approaches 0 and is between the generic reader's curve and the horizontal axis as q approaches 1. For values of q between 0 and 1, the figure illustrates the possibility of a highly non-monotonic OA effect, in this example rising above the horizontal axis for a "pocket" of some relatively low values of q, dipping back below for larger q, and rising above the horizontal axis again for the highest values of q. While such a drain-pipe shape is not guaranteed—indeed, the curve reverts to

the monotonic gray curve in the limit $\sigma \to 0$—Figure S1 documents the possibility. The theoretical possibility that the OA effect is highest for a "pocket" of moderate rather than the highest quality articles hinges on the presence of smart insiders in the model. Such insiders are smart enough to avoid citing mediocre articles unseen. For them, the main effect of a move from closed to OA is to increase the measure obtaining full access, which can translate into a large OA effect Δ since Δ is measured as a percentage increase over a potentially very small base of cites that a mediocre article would receive from smart insiders under closed access.

Overall, our insider/outsider analysis has several empirical implications. We expect the OA effect for outsiders to be negative across the quality spectrum. The effect should also be negative for insiders citing the lowest-quality articles. Insiders may exhibit positive OA effects for higher quality articles; the effect may exhibit non-monotonicities and may be highest for articles of intermediate rather than the highest quality.

Online Appendix S2: Supplementary Exhibits

Table S1: OA Results Binning by Citation for Alternative Citation Bins

	Cites in selection period			
	0 cites	1–2 cites	3-9 cites	10+ cites
Partial OA	0.000	−0.064	−0.009	0.033***
	(0.041)	(0.046)	(0.020)	(0.012)
Full OA	−0.046	−0.089**	−0.022	0.077***
	(0.060)	(0.034)	(0.025)	(0.015)
Articles	31,008	35,212	32,986	19,589
Panel observations	162,735	177,150	160,561	95,108
Article fixed effects	Yes	Yes	Yes	Yes
Publication × citation year fixed effects	Yes	Yes	Yes	Yes
Partial online-access indicator	Yes	Yes	Yes	Yes
Full online-access indicator	Yes	Yes	Yes	Yes
Journal-specific age profile	Linear	Linear	Linear	Linear

Notes: Specification is identical to Table 4 except uses an alternative partition to for bins. Each column is a separate regression including observations for articles having the specified number of cites in selection period. Observations in selection period are omitted from the regressions, reducing the sample size relative to that reported in Table 1. Additional notes from Table 4 apply.

Table S2: OA Results Binning by Publication-Year Citation Percentiles

	Publication-year percentiles				
	0–50%	50–62.5%	62.5–75%	75–87.5%	87.5–100%
Partial OA	–0.025	–0.083	–0.038	–0.002	0.030**
	(0.032)	(0.052)	(0.055)	(0.020)	(0.013)
Full OA	–0.059	–0.124***	–0.062**	–0.014	0.071***
	(0.047)	(0.032)	(0.059)	(0.027)	(0.015)
Articles	41,239	18,079	19,492	19,957	20,028
Panel observations	205,940	90,840	97,978	100,239	100,557
Article fixed effects	Yes	Yes	Yes	Yes	Yes
Publication × citation year fixed effects	Yes	Yes	Yes	Yes	Yes
Partial online-access indicator	Yes	Yes	Yes	Yes	Yes
Full online-access indicator	Yes	Yes	Yes	Yes	Yes
Journal-specific age profile	Linear	Linear	Linear	Linear	Linear

Notes: Each column is a separate regression including observations for articles whose cites in the selection period (first two years after publication) fall into that percentile compared to other articles published in the same year. Observations in selection period are omitted from the regressions, reducing the sample size relative to that reported in Table 1. Additional notes from Table 4 apply.

Table S3: OA Results for Insiders vs. Outsiders Forming Citation Bins Using All Citations

	Cites in selection period				
Variable	0 cites	1 cite	2–5 cites	6–10 cites	11+ cites
	A. Insider cites				
Partial OA	–0.151**	–0.136***	–0.007	0.025	0.105*
	(0.062)	(0.045)	(0.025)	(0.035)	(0.059)
Full OA	–0.187***	–0.150***	0.013	0.064	0.087
	(0.047)	(0.027)	(0.028)	(0.041)	(0.056)
Articles	15,715	12,695	20,027	4,522	1,756
Panel observations	83,146	64,871	98,972	21,287	8,168
	B. Outsider cites				
Partial OA	–0.021	–0.033	–0.099**	0.021	–0.122*
	(0.045)	(0.056)	(0.039)	(0.070)	(0.055)
Full OA	–0.045	–0.032	0.190***	–0.121***	–0.127**
	(0.080)	(0.054)	(0.025)	(0.041)	(0.049)
Articles	16,439	12,789	19,873	4,490	1,750
Panel observations	87,227	65,613	98,642	21,181	8,150
Article fixed effects	Yes	Yes	Yes	Yes	Yes
Publication × citation year fixed effects	Yes	Yes	Yes	Yes	Yes
Partial-OA indicator	Yes	Yes	Yes	Yes	Yes
Full-OA indicator	Yes	Yes	Yes	Yes	Yes
Journal-specific age profile	Linear	Linear	Linear	Linear	Linear

Notes: Results are comparable to Table 7 except that, instead of using just insider cites during selection period to form bins in panel A or just outsider cites in panel B, all citations during selection period are used in both panels. Specification is otherwise identical to that in Table 7; see that table for applicable notes.