NBER WORKING PAPER SERIES

# CERTIFIED RANDOM:
# A NEW ORDER FOR CO-AUTHORSHIP

Debraj Ray
Arthur Robson

Working Paper 22602
http://www.nber.org/papers/w22602

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2016

Ray thanks the National Science Foundation for support under grant numbers SES-1261560 and SES-1629370. Robson thanks the Canada Research Chairs Program and the Social Sciences and Humanities Research Council of Canada. We thank Dan Ariely, Joan Esteban, Itzhak Gilboa, Ed Green, Johannes Horner, Navin Kartik, Laurent Mathevet, Sahar Parsa, James Poterba, Andy Postlewaite, Phil Reny, Ariel Rubinstein, Larry Samuelson, Rakesh Vohra, and Leeat Yariv for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Certified Random: A New Order for Co-Authorship
Debraj Ray and Arthur Robson
NBER Working Paper No. 22602
September 2016, Revised December 2016
JEL No. A10,A14

## ABSTRACT

In economics, alphabetical name order is the baseline norm for joint publications. A growing literature suggests, however, that alphabetical order confers uneven benefits on the first author. This paper introduces and studies certified random order, which involves randomization of names that is institutionally ratified by a commonly understood symbol. Certified random order maintains all the ethical niceties of alphabetical order, but in addition: (a) it distributes the psychological and perceptual weight given to first authorship evenly over the alphabet, (b) it allows either author to signal credit when contributions are extremely unequal, (c) it will be willingly adopted in a decentralized manner even in an environment where alphabetical order is dominant, (d) it is robust to deviations, (e) it dominates alphabetical order on the grounds of ex-ante efficiency, and (f) barring the addition of a simple symbol, it is no more complex than the old system, and brings perfect symmetry to joint authorship.

Debraj Ray
Department of Economics
New York University
19 West Fourth Street
New York, NY 10003
and University of Warwick
and also NBER
debraj.ray@nyu.edu

Arthur Robson
Department of Economics
Simon Fraser University
8888 University Drive
Burnaby, British Columbia
CANADA V5A 1S6
robson@sfu.ca

# Certified Random:

## A New Order for Co-Authorship[1]

Debraj Ray and Arthur Robson®
New York University and Simon Fraser University

August 2016

Abstract

In economics, alphabetical name order is the baseline norm for joint publications. A growing literature suggests, however, that alphabetical order confers uneven benefits on the first author. This paper introduces and studies *certified random order*, which involves randomization of names that is institutionally ratified by a commonly understood symbol. Certified random order maintains all the ethical niceties of alphabetical order, but in addition: (a) it distributes the psychological and perceptual weight given to first authorship evenly over the alphabet, (b) it allows *either* author to signal credit when contributions are extremely unequal, (c) it will be willingly adopted in a decentralized manner even in an environment where alphabetical order is dominant, (d) it is robust to deviations, (e) it dominates alphabetical order on the grounds of ex-ante efficiency, and (f) barring the addition of a simple symbol, it is no more complex than the old system, and brings perfect symmetry to joint authorship.

## 1. Background

Our last names above seemingly appear in alphabetical order, but they don't. A coin was tossed to determine name placement. The symbol ® next to our names is a signal that the names are in *random order*. A study of *certified* random order — i.e., randomization that is institutionally ratified by a commonly understood symbol — is the object of study of this paper.

In economics, as is universally known to economists, alphabetical order is the baseline norm for name order in joint research. Table 1 reports the prevalence of alphabetical order in economics. Around 85% of two-author papers are written in alphabetical order.[2] That percentage falls with more authors, possibly capturing the fear of the notorious *et al* effect, or there could be research assistants involved.[3] Compare this to the physical sciences, in which there is often a tussle for first authorship, while other not-so-subtle signals such as lab leadership are sent through ancillary ordering conventions. It can be argued that the civility of alphabetical order lends itself to more joint work, as the possible rancor in settling on a name order at publication time is thereby avoided.

And yet, there are issues with alphabetical order that are worth taking serious note of:

1. Psychologically, names that appear first are more likely to be given "extra credit." This order effect is certainly in line with research on marketing: products presented earlier exhibit higher probabilities of selection, as the aptly ordered article by Carney and Banaji (2012) observes. Even stocks with earlier names in the alphabet are more likely to be traded; see another aptly ordered paper by Jacobs and Hillert (2016), or Itzkowitz, Itzkowitz and Rothbort (2016).

2. Earlier names appear bunched together on a bibliographical or reference list, lending additional perceptual weight to how often they are cited. They also appear earlier on the reference list. Haque and Ginsparg (2009) — aptly ordered again — note that article positioning in the ArXiv repository is correlated with citations

[1]Ray thanks the National Science Foundation for support under grant numbers SES-1261560 and SES-1629370. Robson thanks the Canada Research Chairs Program and the Social Sciences and Humanities Research Council of Canada. We thank Dan Ariely, Joan Esteban, Itzhak Gilboa, Ed Green, Johannes Horner, Navin Kartik, Laurent Mathevet, Sahar Parsa, James Poterba, Andy Postlewaite, Phil Reny, Ariel Rubinstein, Larry Samuelson, Rakesh Vohra, and Leeat Yariv for helpful comments.

[2]Certainly, alphabetical order is occasionally overturned (see Table 1 again) and when it is, it is a clear signal that the author who now appears first has done the bulk of the work. We will return to this issue below; it will be central to the theory we develop.

[3]The *et al* effect alone is possibly captured better by papers in which the first author is out of alphabetical order; this percentage is, not surprisingly, lower as Table 1 reveals.

| | Number of Authors | | | |
|---|---|---|---|---|
| | **Two** | **Three** | **Four** | **Five** |
| Total | 53858 | 17857 | 1865 | 340 |
| Alphabetical | 45337 | 13124 | 1155 | 163 |
| Non-Alphabetical | 8521 | 4733 | 710 | 177 |
| % Non-Alpha | **15.82** | **26.51** | **38.07** | **52.06** |
| First Author Non-Alpha | 8521 | 2754 | 339 | 95 |
| % First Author Non-Alpha | **15.82** | **15.42** | **18.18** | **27.94** |

**Table 1.** Alphabetical Order in Peer-Reviewed Journals in Economics. *Sources and Notes. EconLit*, 1969–2013, using the list of 69 leading economics journals in Engemann and Wall (2009).

of that article. Feenberg, Ganguli, Gaule, and Gruber (2015) demonstrate that the same bias exists in the downloading and citation of NBER "New This Week" Working Papers, which led to a change in NBER Policy.[4]

3. There is at least one major journal in economics (the *Review of Economic Studies*) which publishes articles in alphabetical order (using the last name of the first author). Because many other journals use the convention that the lead article is to be regarded as special, and because many do not know that the *Review of Economic Studies* follows this policy — did you? — this confers a potential advantage on earlier names.

4. There is, of course, the *et al* convention, which, while strictly speaking is not a corollary of alphabetical order, is widely used in citations and especially on slides in seminars, completely swamping the identity of later authors. Even if *et al* were to be banned in journal publications (which it currently is not), it cannot be banned from slides.[5] In addition, it is widespread practice in verbal presentations to mention the name of the first author and then add "and coauthors": an understandable but possibly inequitable shortcut.

There is ample evidence that these considerations matter. In a paper published in the *Journal of Economic Perspectives*, Einav and Yariv (2006) write (we quote their abstract in full):

> "We present evidence that a variety of proxies for success in the U.S. economics labor market (tenure at highly ranked schools, fellowship in the Econometric Society, and to a lesser extent, Nobel Prize and Clark Medal winnings) are correlated with surname initials, favoring economists with surname initials earlier in the alphabet. These patterns persist even when controlling for country of origin, ethnicity, and religion. We suspect that these effects are related to the existing norm in economics prescribing alphabetical ordering of authors' credits. Indeed, there is no significant correlation between surname initials and tenure at departments of psychology, where authors are credited roughly according to their intellectual contribution. The economics market participants seem to react to this phenomenon. Analyzing publications in the top economics journals since 1980, we note two consistent patterns: authors with higher surname initials are significantly less likely to participate in projects with more than three authors and significantly more likely to write papers in which the order of credits is non-alphabetical."

There are several other papers that are in line with the Einav-Yariv empirical findings; see, for instance, the appropriately hedged Chambers, Boath and Chambers (2001), or the impeccably ordered van Praag and van Praag (2008). Going beyond Einav and Yariv (2006), this last article finds "significant effects of the alphabetic rank of an economist's last name on scientific production, given that an author has already a certain visibility in academia ... Being an *A* author and thereby often the first author is beneficial for someone's reputation and academic performance." Moreover, as they go on to observe, the recognition

[4]An email from James Poterba dated September 2, 2015, states that "beginning next week, the order of papers in each of the more than 23,000 "New This Week" messages that we send will be determined randomly. This will mean that roughly the same number of message recipients will see a given paper in the first position, in the second position, and so on."

[5]Larry Samuelson advocates avoiding *et al* entirely, since the cost of listing authors in full is now negligible, at least in electronic documents. But it is harder to follow this injunction in verbal presentations.

accorded to earlier authors appears to cumulate over time: "Professor A, who has been a first author more often than Professor Z, will have published more articles and experienced a faster productivity rate over the course of her career as a result of reputation and visibility." A recent survey by Weber (2016) summarizes the literature thus: "there is convincing evidence that alphabetical discrimination exists."

## 2. Name-Order Conventions and Institutional Intervention

Social conventions — name-order norms among them — are typically equilibria that are immune to deviations by individuals or even by sub-segments of the population. Suppose, for instance, that Archimedes and Boethius, working together,[6] gallantly agree to be fair to each other by (privately) randomizing their joint authorship, perhaps over a sequence of papers. Will they agree to the randomization? There are clear difficulties. *Given* an "alphabetical society," a change in name order is a clear signal that the newly christened first author has contributed the bulk of the work. Thus, for instance, "Boethius and Archimedes" would be a statement that Boethius has done most of the research for that paper, whereas "Archimedes and Boethius" would indicate very little, any such signal being swallowed in most part by the naming convention. Therefore Archimedes gains nothing over alphabetical order when his name comes first, while Boethius gains a lot when his does.[7] Boethius will agree to the *ex-ante* randomization, but Archimedes will not. That is one reason it is hard to "invade" an alphabetical society with a different scheme. (But see Section 6.1 for more discussion.) We formalize these remarks below by showing that alphabetical order is stable under deviations — deterministic or random — given the set of alternatives available to authors today (Theorem 1).

But institutions can change that. Here is a simple variant of the randomization scheme — a mutant — which will set it apart from pure randomization. Suppose that any randomized name order is presented with the symbol Ⓡ immediately following it; e.g., Ray and Robson® (2016) is the appropriate reference for this paper. Suppose, moreover, that such a symbol is certified by an august body such as the *American Economic Association. That is, they acknowledge that this alternative is available.* There is, of course, no question of imposing it.

It is unclear that this "mutant" would successfully invade the population. But we are going to argue that it will. The key point that makes this argument possible is that economics does not *entirely* follow alphabetical order. There are exceptions, which occur when the author who is lower down the alphabetical food chain has really contributed quite disproportionately. And we know these exceptions exist, because we see them being made quite often. Table 1 shows that over 15% of two-author publications in the 69 leading economics journals identified by Engemann and Waall (2009) have their names reversed. That percentage rises significantly for three or more authors.

How are these exceptions made? Clearly, the first author (typically gracefully) concedes the order change in such circumstances. We therefore work with a model that builds in this capacity to concede — albeit in extreme situations — out of a sense of "guilt" or "fairness." But guilt isn't really needed, and the exact source of the first author's unease is unimportant. For instance, he might feel uncomfortable about the sense of resentment that his co-author might feel if the contributions are very different and go unacknowledged. The important point is that that same capacity which facilitates name reversal *also* ushers in certified random order as just described, but *not* private randomization. In short, in the presence of certified random order, alphabetical order is "unstable," even though it is "stable" against the possibility of private randomization (Theorem 2). The reverse is not true: random order is indeed immune to invasion by alphabetical order or any other existing name-order arrangements (Theorem 3).

---

[6]This perhaps stretches realism a bit. Boethius, a Roman senator and philosopher, and best known for his work, *The Consolation of Philosophy*, was a bit more than 750 years younger than Archimedes, but we're not off by much more than the average assumption in a theory paper.

[7]Engers *et al* (1999) emphasize this point, arguing that alphabetical order can *disadvantage* "early authors," because a reversal can be used to signal a higher contribution by the late author, but there is no comparable signal for the early author. That may well be true, but on the other hand Engers *et al* have no counterpart to the direct premium from first-authorship that we will posit. The empirical literature that we've discussed suggests that such a premium is a first-order consideration. The Engers *et al* model does not generate the advantage to first authorship seen in the data, because the authors' payoffs are only the Bayesian rational assignment of credit. For example, if authors are always listed alphabetically, then the credit assigned in their model will be equal.

Observe that the mere fact of successful invasion does *not* imply that the replacement is somehow better. The classic example for this is the invasion of cooperative norms by deviants: while the invasion is often successful, the resulting Prisoners' Dilemma outcome is Pareto-inferior to the cooperative convention it displaces. So the above arguments are not a normative prescription.

That said, in the current scenario, random order appears to maintain all the ethical niceties of alphabetical order. In addition, there are some other features worth consideration. The first is from the perspective of *ex ante* efficiency. When ordering is alphabetical but relative contributions are not consistent with that ordering, there will be feelings of unfairness, guilt, disappointment, or outrage. Indeed, the fact that alphabetical order is occasionally reversed is circumstantial evidence that such feelings do exist. At the same time, because alphabetical order *is* the norm and therefore can often be insisted upon by the first author, it is unclear that the "disappointment-minimizing" choice of name order is invariably made.[8] These ex-post considerations can then create inefficiencies ex-ante, as authors weigh the various payoffs conditional on name-order. In Theorem 4, we formalize and prove this result.[9]

Second, random order distributes the psychological and perceptual weight given to first authorship evenly over the alphabet. Given the efficiency result in Theorem 4, it is therefore easy enough to show that any quasiconcave Bergson-Samuelson welfare function defined over author payoffs would prefer random order to alphabetical order.

Third, random order allows "outlier contributions" to be recognized in both directions; that is, given the convention that puts "Archimedes and Boethius®" on center-stage, *both* "Archimedes and Boethius" and "Boethius and Archimedes" would acquire symmetric — and fair — meaning. Barring the addition of a simple symbol, random order is no more complex than the existing system, and brings perfect symmetry to joint authorship.

Now for the details.

## 3. A Model of Name-Order Conventions

Say a paper is worth a total credit of 1 unit. There are two authors: our eminent worthies Aristotle and Boethius.[10] Their contributions are $x$ and $y = 1 - x$ respectively, where $x$ is uniformly distributed on $[0, 1]$. Ex post, $(x, y)$ is observed by the authors but not by the public, who must infer these from the ambient social convention and the particular name order followed.

3.1. **Name-Order Conventions.** Let $n$ be a name order: e.g., alphabetical order ($n = \alpha$), reverse-alphabetical order ($n = \rho$), or random order ($n = r$). The use of a particular order sends signals about the contributions $(x, y)$, but these signals also depend on the ambient naming *convention* in place in society. A convention maps realized contributions $(x, y)$ to choices of name order. For instance, *pure meritocracy* is the convention that chooses $\alpha$ when $x > 1/2$, and $\rho$ when $y > 1/2$. *Pure alphabetical order* is the convention that chooses $\alpha$ no matter what the contributions are. Economics uses a convention that is close to pure alphabetical order, with a small modification: sometimes, the names are reversed, presumably to

[8] To some extent, these feelings can be also taken care of with merit-based ranking, but merit has its own share of problems, to which alphabetical order was presumably a response in the first place.

[9] There are other efficiency arguments. For instance, individuals put in effort into doing research. Unequal division of the credits from that research might be surplus-dominated — even Pareto-dominated — by equal division, as efforts adjust to the more equitable distribution of credits. This approach to team production with moral hazard is a possible critique worth considering, but not one we study here. It would add to our arguments, but in ways that the literature already appreciates. We note that Engers *et al* (1999) derive the contributions of the authors from their endogenous effort choices. In particular, they show that alphabetical order prevails over meritocracy in equilibrium, despite the greater efficiency of the latter (in their model). One could extend this argument *a fortiori* to random order. We do not take this route here. There are also possible efficiency losses from the strategic choice of co-authors when it is feared that lexicographic relegation to the end of the name-order (or even to the dreaded *et al* dungeon) might lead to a loss in payoff. Einav and Yariv (2004) provide evidence that individuals further down the alphabet are more averse to writing with multiple co-authors. Again, this is an efficiency loss. See Ray (2013) for notes on strategic choice of co-authors.

[10] The analysis of three or more authors is an interesting open question that we leave for future investigation in the event that this paper does not suffer an untimely demise.

signal a significant imbalance in contributions. Formally, alphabetical order is followed as a default, with a name-order reversal when $x < \epsilon$, for some threshold $\epsilon \in (0, 1/2)$.[11]

3.2. **Credits.** Let $a(n, C)$ and $b(n, C)$ denote the credits to Archimedes and Boethius respectively when convention $C$ is observed and the name-ordering $n$ is followed. (These are not overall payoffs, which will include two other components to be described below.) For the modified alphabetical convention $E$ used in Economics with threshold $\epsilon$, the use of $n = \alpha$ yields a credit of

$$a(\alpha, E) = \frac{1+\epsilon}{2} \text{ and } b(\alpha, E) = \frac{1-\epsilon}{2} \tag{3.1}$$

to Archimedes and Boethius respectively, while, if the reverse order $\rho$ is observed, the corresponding credits are

$$a(\rho, E) = \frac{\epsilon}{2} \text{ and } b(\rho, E) = 1 - \frac{\epsilon}{2}. \tag{3.2}$$

Of course, $a(n, C) + b(n, C) = 1$; that is, a total credit of 1 is always being divided.[12]

3.3. **Reputational Payoff.** We suppose that each name order yields a gain $\delta$ in the reputation of the *first* author in the order. This is due to visibility, bunching in reference lists, the *et al* effect and so on, and it accrues over and above the "direct credits" $a(n, C)$ and $b(n, C)$ that the public will estimate. This reputational payoff is central to our model.

3.4. **Guilt and Overall Payoffs.** Our final ingredient in payoffs is some notion of unfairness or guilt which comes into play when the naming order departs substantially from relative contributions. To this end, we introduce a function $\Gamma(z)$: the "guilt function." It is experienced by the author if he feels that he has treated his co-author badly.[13] It has as its domain the *difference* $z$ between the true credit due to the co-author, and the inferred credit that the co-author obtains from the announced name order and the going convention.

For example, consider the economics convention $E$. Suppose that $(x, y)$ are the true contributions, whereas the inferred contribution under $E$ is $\left(\frac{1+\epsilon}{2}, \frac{1-\epsilon}{2}\right)$. Then, in the eyes of Archimedes, the shortfall for Boethius is $z_B = y - \frac{1-\epsilon}{2}$, and the resulting sense of guilt that Archimedes experiences is given by $\Gamma(z_B)$. (As formalized below, this will only matter when $z_B > 0$.)

Overall payoffs are given as follows. For some realization of true credits $(x, y)$, Archimedes's overall utility from a choice of name order $n$ is assumed to be *the following difference*:

(i) his payoff from $n$, which is $a(n, E)$ plus $\delta$ if his name comes first under $n$; *minus*

(ii) the guilt experienced if he is unfair to Boethius $\Gamma(z_B)$ generated by $n$, as described above.

A parallel formulation holds for Boethius.

We impose the following restrictions on $\Gamma$. First, we presume that $\Gamma(z) = 0$ when $z \leq 0$. (To be sure, the *other* author is treated unfairly when $z < 0$, but we count that separately.) Second, $\Gamma$ is continuous everywhere, and increasing and strictly convex for $z \geq 0$. Finally, we impose two end-point conditions that

[11] We adopt the simplification that $\epsilon$ is the same for all author pairs. One might also consider heterogeneous standards of fairness, with the public using an average when assigning credit; presumably this would be a relatively straightforward extension.

[12] It is possible that the name order also affects the total credit. For instance, "Boethius and Archimedes" will garner less *overall* attention than "Archimedes and Boethius" in a reference list. We are assuming that this effect is small relative to the other considerations in play.

[13] This is a simple and minimal formulation, and one might proceed differently. For instance, one can think of the co-author experiencing resentment at the perceived mistreatment, which is then passed on to the author, inducing him to make a different decision. A related interpretation of $\Gamma$ is that it is a reduced-form strategic punishment cost to an individual of being "unfair," paid, say, in terms of reduced payoffs from future coauthorships. We do not expect these more complicated formulations to make any difference to the results.

play several roles in the analysis. In particular, they guarantee that a name-reversal threshold exists under $E$, but is skewed below $x = 1/2$:

$$\Gamma\left(\frac{1}{2}\right) > \frac{1}{2} + \delta, \text{ but } \Gamma\left(\frac{1}{4}\right) \leq \frac{1}{2}. \tag{3.3}$$

The first end-point condition can be interpreted as follows. Consider a pure alphabetic convention. Suppose that Boethius has done *all* the work, and Archimedes none, so that $(x, y) = (0, 1)$. By insisting on alphabetical order, Archimedes "steals" $1/2$ a unit of credit from Boethius. This generates a sense of guilt (in Archimedes's mind) equal to $\Gamma(1/2)$. The first inequality in (3.3) guarantees that Archimedes will feel bad enough to reverse authorship in this case. The data in Table 1 essentially demand that such a restriction be placed on the model.

For the second end-point condition, consider a hypothetical, purely meritocratic system, which assigns no $\delta$-premium to name order, but assigns a credit of $3/4$ to the first name and $1/4$ to the second name, which credits follow from the uniform distribution on $[0, 1]$ assumed for $x$. In such a system, assume that if Archimedes and Boethius have contributed *exactly equally*, Archimedes would prefer his name to go first, rather than reverse names. That is, we imbue both both players with a minimum degree of selfishness in the face of equal contributions. Then it is easy to see that $\Gamma(1/4) \leq 1/2$.

3.5. **Stable Conventions.** Behavior under a convention can be formalized using game theory, though the game in question is convention-dependent. Informally, we want to think of either author as being able to submit a "proposal" to do something different, to which an author suggesting the default can suitably respond. More concretely:

In Stage 1, contributions $(x, y)$ are revealed to the authors. In Stage 2, both Archimedes and Boethius independently choose either the default $d$ or one of a number of other actions $s$ (a name order, or a proposal to randomize across name orders). In Stage 3, if the same actions are chosen, they are implemented. If both parties take different *non*-default actions, then the default is implemented. Finally, if one party chooses the default, and the other $s$, then the person who chose the default action is given the opportunity to agree to $s$, or suggest a new proposal $s'$ of his own (which could include the default). If the counterproposal is accepted, $s'$ is implemented. If not, the default is implemented.[14]

Defaults and payoffs are only well-defined under the convention that holds. For instance, under the Economics convention $E$, each party can insist on alphabetical order; i.e., $d = \alpha$, and for the two name orders used by the convention, the payoffs are specified by (3.1) and (3.2), coupled with the $\delta$ and fairness terms. We can extend payoffs to mixed actions in the obvious way. A delicate consideration arises, however, for actions that are never specified by the convention in question: what payoffs are to be assigned to those cases?

Specifically, consider an isolated appearance of the Ⓡ-symbol in a society fully dominated by the Economics convention. Whether the authors in question would like to implement a deviation will depend on the social perception of relative contributions following the observation of that deviation. Our analysis therefore rests on the device of an "equilibrium deviation," which is related to neologism-proofness (Farrell 1993). Briefly, suppose that society assigns certain beliefs to the expected relative contributions of the authors when an occurrence of Ⓡ is observed. Those beliefs must then be "rational" in the sense that they must correspond to the expected value of precisely those configurations for which the authors wish to deviate to random order (in an ambient environment dominated by alphabetical order).

This is an (admittedly strong) "equilibrium refinement" which ties down the out-of-equilibrium beliefs held by society. Without some restriction on out-of-equilibrium beliefs, it is clear that a wide variety of outcomes could supported as stable outcomes. This particular refinement is intuitively attractive.

As in the large literature on evolutionary selection following a mutation, we presume that the fraction of the population that are Ⓡ mutants is "small". The useful implication of this is that the inferred contributions after observing $\alpha$ or $\rho$ are not affected by the presence of the mutant. See Section 6.2 for more discussion.

[14]This structure rules out coordination failures in which undesirable self-fulfilling outcomes emerge, such as both parties choosing $d$ even when they both want to switch to $s$.

A *stable convention*, then, is assessed within the context of some allowable set of actions, both used and unused. Given that allowable set, (a) the actions that are used in the convention must constitute a subgame perfect equilibrium (henceforth, SPE) of the game *that has only these used actions*, and (b) there must be no unused action from the allowable set which can serve as an equilibrium deviation. In particular, the requirement that a convention is an SPE is necessary for it to be stable, but not sufficient.

## 4. Stability of the Economics Convention

In this section, we analyze the *economics convention* $E$, which is alphabetical order modified by a threshold $\epsilon$ such that for $x < \epsilon$, name-order is willingly reversed by both parties. Under the Economics convention with commonly agreed reversal threshold $\epsilon$, and using (3.1), Archimedes's utility from $\alpha$ is

$$a(\alpha, E) + \delta - \Gamma\left(z_B^{\alpha}\right) = \frac{1+\epsilon}{2} + \delta - \Gamma\left(y - \frac{1-\epsilon}{2}\right), \tag{4.1}$$

while if $\rho$ is implemented, then using (3.2), his utility is

$$a(\rho, E) - \Gamma\left(z_B^{\rho}\right) = \frac{\epsilon}{2} - \Gamma\left(y - \left[1 - \frac{\epsilon}{2}\right]\right). \tag{4.2}$$

4.1. **Baseline.** In the baseline setting without Ⓡ, the only options available are alphabetical order $\alpha$ or reverse-alphabetical order $\rho$, and randomizations over these actions. With this in mind, we want to solve out for Archimedes's reversal decision (when Boethius contributes a lot) and then use a fixed point argument to make sure it coincides with society's anticipated threshold. To this end, using (4.1) and (4.2), and substituting in the values for credits, observe that Archimedes will reverse when

$$\Gamma\left(y - \frac{1-\epsilon}{2}\right) - \Gamma\left(y - \left[1 - \frac{\epsilon}{2}\right]\right) > \frac{1}{2} + \delta, \tag{4.3}$$

where the RHS is the "direct loss" to Archimedes from reversal, and the LHS is the gain he enjoys by reducing his sense that he has been unfair to Boethius. By strict convexity, the left-hand side is increasing in $y$, and so there exists a unique $y^*$ large enough (it could be 1) such that Archimedes will agree to reverse if $y > y^*$, or equivalently, $x < x^* \equiv 1 - y^*$. In a social equilibrium, $x^* = \epsilon$, and using this information in (4.3), we see that that $\epsilon$ must solve

$$\Gamma\left(\frac{1-\epsilon}{2}\right) - \Gamma\left(-\frac{\epsilon}{2}\right) = \Gamma\left(\frac{1-\epsilon}{2}\right) = \frac{1}{2} + \delta, \tag{4.4}$$

Condition (3.3), and the assumption that $\Gamma$ is increasing, guarantees that there is a unique value of $\epsilon \in (0, 1/2)$ that solves (4.4).

Notice how the strict convexity of $\Gamma$ and the end-point conditions (3.3) are necessary to get what we see in the data. For instance, if $\Gamma$ is linear, then Archimedes simply trades off units of his credit for units of Boethius's, and if he places a higher weight on own credits, he will never reverse. If he places a higher weight on Boethius's credits, he will always reverse. We see neither, which suggests that the "marginal guilt" climbs as Boethius's contribution climbs (holding fixed the name order).

Is this convention an SPE? It is. Given the "fixed-point" that pins down $\epsilon$, Archimedes will want to precisely use the threshold $\epsilon$ and no other. Boethius might want to reverse for even lower values of his contribution $y$, but Archimedes will have none of it: his sense of guilt is not piqued enough for that to happen. It should also be noted that Boethius will always be happy with the decision to reverse; i.e., he will not refuse Archimedes's gesture on the grounds that *he* is now treating Archimedes unfairly. The most compelling case for this possibility is when Archimedes has contributed $\epsilon$ but only receives a credit of $\epsilon/2$, in which case Boethius receives the overall payoff $\delta + [1 - (\epsilon/2)] - \Gamma(\epsilon/2)$ on reversal. In the status quo, he gets $(1-\epsilon)/2$. Consequently, Boethius will always be happy with reversal provided that $\delta + [1 - (\epsilon/2)] - \Gamma(\epsilon/2) \geq (1-\epsilon)/2$, which yields

$$\Gamma\left(\frac{\epsilon}{2}\right) \leq \frac{1}{2} + \delta. \tag{4.5}$$

Comparing this requirement with (4.4) and noting that $\epsilon < 1/2$, we see that (4.5) will always be satisfied.

What other options do Archimedes and Boethius have? Given $(x, y)$, they might agree to randomize the order of their names by tossing a (possibly biased) coin. But the expected utility (to Archimedes) of such a coin flip is strictly sandwiched between the two utilities from $\alpha$ and $\rho$, so that if, say, the expected utility beats that from $\alpha$, it must in turn be bettered by the utility from $\rho$. Generically (in $x$), randomization can never occur, at least not without an institutionally supported indicator such as Ⓡ.

It is easy enough to convert these arguments into a formal treatment. With alphabetical order as default, $\alpha$ will be the unique SPE outcome conditional on credit $x$ lying in $(\epsilon, 1)$ and $\rho$ the unique SPE outcome conditional on credit lying in $(0, \epsilon)$. When $x \in (\epsilon, 1)$, Archimedes can force his preferred outcome $\alpha$ by choosing $\alpha$ (the default action) to begin with, and thereafter insisting on it. When $x \in (0, \epsilon)$, $\rho$ is best for both and is therefore easily seen to be the unique equilibrium. We have thus established:

**Theorem 1.** *There is a unique value $\epsilon \in (0, 1/2)$, given by the solution to (4.4), for which the economics convention $E$ with threshold $\epsilon$ is an SPE.*

However, as already noted, being an SPE does not guarantee immunity to possible invasion by allowable actions that are unplayed in equilibrium. We now turn to an examination of this question.

4.2. **Disrupting $E$ With Certified Random Order Ⓡ.** Introduce the possibility of random order (aways with Ⓡ, in case we do not explicitly mention it). This must be an option that is institutionally provided, say by a consortium of the leading journals, so that the meaning of Ⓡ is commonly known. The question is whether it affects the existing social order defined by $\alpha$ with the occasional reversal to $\rho$.

Given that the alphabetical order norm is currently dominant, credits to the choice of $\alpha$ or $\rho$ will continue to be given by (3.1) and (3.2), where $\epsilon$ is pinned down by (4.3).[15] Should Archimedes and Boethius employ random order for some realizations of $x$? If they do, a new pair of payoffs will be generated. These will consist of $\delta/2$ to each (in expected value), plus deterministic (though possibly asymmetric) credits to each, and any additional guilt terms. Of these three components, notice that assigned credits will depend on society's view of just when the authors are agreeing to randomize. If, for instance, it is believed that they are doing so on an interval of $x$-realizations that is symmetric around $1/2$, then the credit will be split equally. But that choice is itself endogenous, and its symmetry is not assured when the background convention is $E$.

To examine the stability of $E$, therefore, we will need to define "equilibrium social beliefs" for surprise observations. We will say, then, that $E$ is *disrupted by an equilibrium deviation* to random order, if there exists an assignment of credits post-deviation that coincides with the expected credits over the set of $x$ realizations for which Archimedes and Boethius both agree to adopt random order.

**Theorem 2.** *The convention $E$ is disrupted by an equilibrium deviation to random order.*

The Appendix provides a complete proof of this central observation. This proof is somewhat involved, but it can be outlined as follows. Fix the convention $E$ and its associated reversal threshold $\epsilon$, given by (4.4). Suppose that society assigns a credit pair $(a, 1-a)$ to the observation of random order. For each such assignment, we find two thresholds defined on the domain of Archimedes's actual credit. One, which we call $x^\alpha(a)$, is such that Archimedes prefers random order to alphabetical order whenever his realized credit falls below $x^\alpha(a)$, and another, called $x^\rho(a)$, is such that Archimedes prefers random order to reverse order whenever his realized credit lies above $x^\rho(a)$. Over a subdomain of the $a$'s, the former threshold lies above the latter, so there is a zone in which Archimedes prefers random order to both alphabetical and reverse order. Moreover, there is a particular assignment of credit $a = a^*$ for which the conditional expected value of Archimedes's credit over this zone *exactly equals* $a^*$. We next verify that Boethius prefers random order in this zone over alphabetical order.

Define $x_1^* \equiv x^\rho(a^*)$ and $x_2^* \equiv x^\alpha(a^*)$. The proof is then completed by showing that in the zone $(x_1^*, x_2^*)$, Ⓡ is the unique SPE outcome. Recall that Ⓡ is Archimedes's favorite outcome in this range. Moreover, Boethius prefers random order in this range to alphabetical order. Even though Boethius might like reverse

[15] Recall that we assume that the equilibrium deviation is adopted, at first, by a "small" mutant fraction of the population of author pairs. This means that the payoffs ascribed to the author listings $\alpha$ and $\rho$ can be taken to be unaffected by the presence of the mutants.

order even more, the fact that Archimedes can retreat to the alphabetical default ensures that the latter can enforce Ⓡ as the SPE outcome in $(x_1^*, x_2^*)$. This argument shows that a small group of Ⓡ mutants can certainly invade the convention $E$, employing an "equilibrium deviation" in the sense described above. That completes the proof of the Theorem; see Appendix for details.

## 5. Stability of the Random Order Convention

We turn now to the analysis of the *certified random order convention*, one that uses Ⓡ. This is a hypothetical convention — for now at least — but we can envisage desiderata for it that run parallel to the economics convention. First, Ⓡ becomes the default choice for either author, so that *both* authors would need to be happy with any change. For instance, Ⓡ can be overridden by Archimedes in favor of $\rho$, if Boethius agrees, but neither Archimedes nor Boethius can insist on this. Likewise, alphabetical order $\alpha$ is certainly available as a choice, but with no default status, so that in the context of the new convention, neither author can unilaterally insist on it.[16]

The set of available actions is $\{d, \alpha, \rho\}$, where the default is now $d =$Ⓡ, which stands for the option in which the two players randomize with equal probability over the name orders $\alpha$ and $\rho$, but with the Ⓡ symbol attached to each realized outcome.

Recall the rules described in Section 3.5. In Stage 1, contributions are revealed. In Stage 2, both Archimedes and Boethius independently choose from $\{d, \alpha, \rho\}$. In Stage 3, if the same action is chosen, this is implemented. If both parties take different *non*-default actions, then the default is implemented. Finally, if one party chooses the default, and the other $s \in \{\alpha, \rho\}$, then the person who chose the default action is given the opportunity to say agree to $s$, or suggest a new proposal $s' \in \{d, \alpha, \rho\}$ of his own. If the counterproposal is accepted, $s'$ is implemented. If not, the default is implemented.

5.1. **The Certified Random Order Convention is an SPE and Stable.** Describe a (completely symmetric) convention as follows. There exists $\epsilon \in (0, 1/2)$, such that (i) if $x < \epsilon$ then the outcome is $\rho$, (ii) if $x > 1 - \epsilon$ then the outcome is $\alpha$, and (iii) if $x \in [\epsilon, 1 - \epsilon]$ then Ⓡ is the outcome. Call this a *random order convention with threshold* $\epsilon$.

Analogously to Theorem 1 for the economics convention, but in sharp contrast to Theorem 2, we have:

**Theorem 3.** *There is a unique value $\epsilon^* \in (0, 1/3)$ for which the random-order convention Ⓡ with threshold $\epsilon^*$ is an SPE and stable in the allowable set $\{$Ⓡ$, \alpha, \rho\}$.*

The Appendix contains a detailed proof; here is an outline. First it is shown that there exists an $\epsilon^*$ such that Archimedes prefers $\rho$ to Ⓡ if $x < \epsilon^*$ but prefers Ⓡ to $\rho$ if $x > \epsilon^*$. Since $\rho$ can be shown to be Boethius's favorite outcome when $x < \epsilon^*$, it follows that $\rho$ is an SPE outcome in this range. Analogously, $\alpha$ is an SPE outcome for $x > 1 - \epsilon^*$.

If $x \in [\epsilon^*, 1/2]$, it can be shown that $d$ is Boethius's favorite outcome, so Boethius can exploit the default status of $d$ to enforce it; that is, $d$ is the only possible SPE outcome. Similarly, Archimedes must obtain his favorite outcome $d$ in the range $x \in [1/2, 1 - \epsilon^*]$.

Finally, observe that the the entire set of allowable actions is on display under the random order convention. Because there is no unused action from the allowable set, the convention is stable.[17]

5.2. **Efficiency Gain From the Random Order Convention.** Up to this point, we have shown that the economics convention is an SPE but not stable, while the random order convention is both an SPE and stable. That is, no forceful intervention is necessary to achieve the transition (more on this in Section 6.2); just the nudge of making Ⓡ available. In this section, we go further to show that such a nudge can also be justified on the grounds of aggregate welfare and not just fairness:

[16] The game we propose gives more detailed and precise expression to this intuitive description.

[17] It is possible that the random order convention might also cease to be stable if a wider set of allowable actions is used, with fresh symbols encoding information about relative contributions. We discuss this further in Section 6.3.

**Theorem 4.** *The random order convention is more efficient than the economics convention, in the sense of having a strictly higher sum of expected overall payoffs for the two agents.*

At one level, Theorem 4 is not surprising since the number of different signals under Ⓡ is three; whereas the number of different signals under $E$ is only two. However, it is *a priori* possible for the three ranges under Ⓡ to be so badly situated that the total expected payoff under $E$ is greater, so there is something to be proved in this respect.

Indeed neither convention is efficient. The convention $E$ is not efficient within the class of two signal systems. It is not hard to see that efficiency requires that the signals have equal ranges, so that, with two signals, it should be that $\epsilon = 1/2$. However, Theorem 1 states that $\epsilon < 1/2$. The convention Ⓡ is *also* inefficient in the class of three signal mechanisms. Once again, efficiency requires that the ranges for the three signals be equal; that is, $\epsilon^* = 1/3$. However, Theorem 3 shows that $\epsilon^* < 1/3$.

Put another way, under either mechanism there is inadequate incentive for an agent to concede first authorship, which leads to inefficiency. See Section 6.3 for a discussion of other conventions. So it is indeed important to observe that the random order convention Ⓡ dominates the economics convention $E$ on these grounds. After all, we've certainly seen that Ⓡ is fairer than $E$: Ⓡ yields equal payoffs for Archimedes and Boethius, whereas $E$ favors Archimedes. A welfare criterion that was sufficiently averse to inequality — sufficiently quasiconcave, that is — might then prefer Ⓡ to $E$.

However, our efficiency result implies that the case for Ⓡ does not rest on "sufficient inequality aversion." The *sum* of the two agents' expected utilities is higher under Ⓡ than under $E$, so that *any* symmetric quasiconcave welfare criterion (including Bentham's additive utilitarianism) would strictly prefer Ⓡ to $E$.[18]

The proof of Theorem 4 involves carving up the possible gamut of contributions, summarized by $(x, 1-x)$, into different ranges that depend on the thresholds $\epsilon$ (from the economics convention) and $\epsilon^*$ (from the random order convention). Overall expected payoff is the sum of (probability-weighted) integrals from each of these ranges. The proof proceeds on the broad strategy that the greater number of ranges afforded by the random order convention allows it to generate a higher expected payoff than the economics convention, though, as already mentioned, such a comparison across conventions with distorted placement of thresholds is not immediate.

## 6. Remarks and Extensions

6.1. **Private Versus Certified Randomization.** The mechanism of *private* randomization is not a novel one. In the simplest case, this involves two researchers flipping a coin to decide the order of names and — in possibly its most effective form — including a lead footnote to that effect. This has been used previously on a number of occasions. The present mechanism of certified randomization differs from such private randomization. In particular, the exact comparison of the two mechanisms depends on the channel through which published papers enter the cognitive window of other researchers.

The present paper is motivated by the key channel of *written* citations to the paper in the subsequent literature. In this case, certified randomization outperforms private randomization. What is important here is that the symbol Ⓡ is maintained in subsequent references, but that the lead footnote with private randomization is not. Indeed, that accurately describes how private randomization has actually operated to date. This implies that private randomization can generate only two permanently observed configurations for Archimedes and Boethius—$\alpha$ or $\rho$, which limits its effectiveness.

But there are other channels, to be sure. The most obvious is direct viewing of the paper by another researcher. In this case, given that there is a lead footnote indicating private randomization, presumably noted by the researcher, as is the symbol Ⓡ under our mechanism, the two mechanisms are essentially equivalent *if direct viewing is the only channel*. That is, there are four possible outcomes of a collaboration between Archimedes and Boethius under private randomization: $\alpha$, $\rho$, $\alpha^{\text{Ⓕ}}$ or $\rho^{\text{Ⓕ}}$, where $\alpha^{\text{Ⓕ}}$, for example,

[18] We assume that such a welfare criterion is defined on the *expected* utilities of the agents. This is appropriate if the publication game is repeated often, so that there are many independent draws of $x$.

means that the authors are listed as first Archimedes, then Boethius, and there is a lead footnote. These possibilities correspond precisely to the four possibilities under the Ⓡ mechanism, with only notational differences.

The last channel that seems worthy of mention concerns verbal and written citations of the published paper in seminars. What happens here depends on how exactly this citation is made. If the symbol Ⓡ is retained (e.g., on slides), whereas the lead footnote escapes attention, the advantages of our mechanism over private randomization remain. If no mention is made of Ⓡ , the effectiveness of our mechanism will be reduced to match that of private randomization.

Certified randomization is always, then, at least as effective as private randomization, and, for at least one important channel, that of written citations in subsequent literature, strictly more effective.

6.2. **The Transition From One Convention to the Other.** We've shown that alphabetical order is not immune to an invasion by (certified) random order, while random order is protected from the reverse invasion. It is customary for game theorists — even evolutionary game theorists who embrace a dynamic model — to rest their case at this point, and as far as the formal analysis goes, so do we. That said, here are some remarks on a more complete analysis.

Return to Theorem 2, in which the economics convention is successfully invaded by an institutionally approved mutant. What happens "after" that initial invasion? There are two sources of further dynamics.

First, once Ⓡ makes its initial appearance, the landscape of updates will begin to shift, not just for the newborn mutant, but also for the pre-existing choices $\alpha$ and $\rho$. For instance, the initial invasion occurs in the interval $(x_1^*, x_2^*)$, but after the mutant becomes significant in that range, the Bayes' update of $x$ following an observation of $\alpha$ will be different from the previous baseline. That in itself will keep the process moving, at least over some rounds. It satisfactorily "proves" that the economics convention is unstable, but it does not deliver us all the way to a random order convention. For that, *the default itself will need to evolve.*

Several choices are available to model evolving defaults. It is unclear that any of them is canonical, though it also appears that most reasonable choices will yield the same result — namely, a transition to the random-order convention. Here is one. Suppose that for each pair-up of authors, random order is taken to be the default with a probability that's increasing in the "going prevalence" of random order. Then there will be an ever-increasing incidence of random order, fed in turn by its rising proclivity to become a default. Meanwhile, $\alpha$ will lose salience as a default, and gain in importance as a powerful signal of the relative contribution of Archimedes, just as $\rho$ remains a powerful signal for Boethius's relative contribution. The newly-acquired, substantially asymmetric meaning of $\alpha$ must eventually eliminate it as a default outcome; that would be as absurd as Boethius insisting on $\rho$ as a default. The random order convention would now reign supreme.

6.3. **Other Conventions And Actions.** In a world where contributions $(x, 1-x)$ can be taken seriously along with their full cardinal import, there are alternative conventions that can achieve higher degrees of efficiency, and mutants built along those lines might even invade the random order convention.

Indeed, there is a formal mechanism that attains full efficiency, as follows. Suppose that, for each $x < 1/2$, $\rho$ is used and Ⓧ is attached to the names, whereas, if $x \geq 1/2$, $\alpha$ is used and Ⓧ is published. Neither agent then ever experiences any guilt, so that overall expected payoffs are $1 + \delta$, which is the upper bound. Moreover, this convention is stable: there is, of course, no deviation from it that both authors would consent to.

But once written this way, the absurdity of such a mechanism is evident. It pushes too hard the assumption that the agents have mutual knowledge of $x$. Realistically, such a convention would lead to endless bitter arguments about the value of $x$ that we now minimize with the convention $E$ and that we should continue to minimize.

6.4. **Randomizing Citations.** An alternative to our proposed mechanism would be to keep published papers with the authors names' listed alphabetically, but to randomize 50-50 each time a *citation* is made.[19]

[19]Leeat Yariv proposed this device, perhaps as a supplement to the mechanism here.

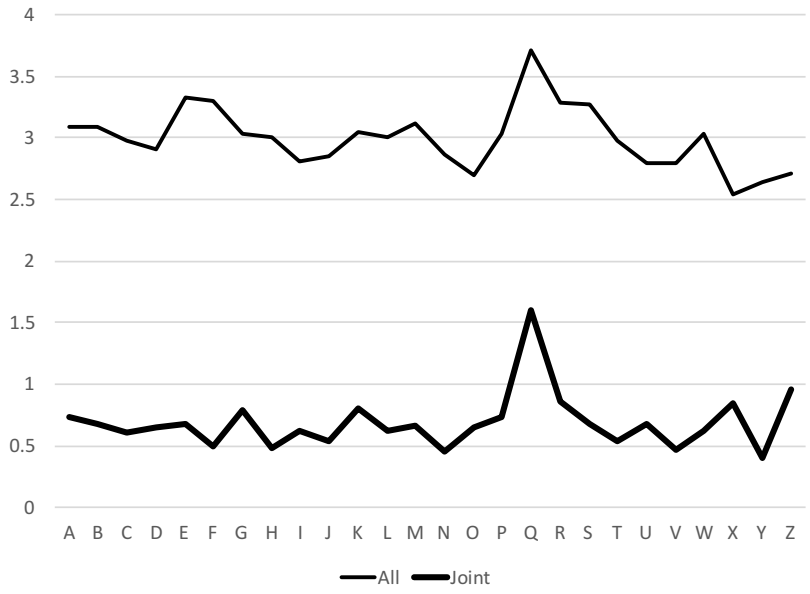



**Figure 1.** Papers Per Author Over All Letters of the Alphabet. Data Source: *EconLit*, 1969–2013.

We believe that the ® mechanism has advantages over this scheme. In the first place, it would be hard to ensure that all researchers citing the paper diligently randomize. This would be especially true in seminar presentations, where much of the harm is done anyway by the verbal use of "and coauthors" after a mention of the first author.

Furthermore, although it might be possible to cite the famous eponymous paper as Douglas-Cobb instead of Cobb-Douglas, for example, it seems it would be difficult to cite it sometimes as Cobb-Douglas and sometimes as Douglas-Cobb.[20]

Moreover, this alternative mechanism does not provide for the counterfactual possibility that Cobb-Douglas might have been Douglas-Cobb, if Douglas had indeed done the lion's share of the work. If that had happened, Douglas would presumably have strenuously objected to 50-50 randomization of first authorship in citations. A real-life example is the well-known Stolper-Samuelson theorem, where the possibly greater contributions of Stolper — presumably signaled with the gentlemanly consent of Samuelson — would subsequently be marred by randomization in citations. The ® symbol, inserted at source as it were, helps to keep track of the different interpretations.

6.5. **Strategic Authorship Decisions.** Authors may choose whom to co-author with, given the going convention. For instance, Einav and Yariv (2006) show that later authors are more reluctant to engage in projects with multiple co-authors, for fear of falling into the *et al* dungeon. One might also make the opposite argument: that in alphabetical order, early authors offer co-authorship with greater ease to late authors, knowing that this will have only a small effect on their payoffs, being listed first anyway.[21] Indeed, Archimedes might be excessively eager to offer co-authorship to Zeno, anticipating that "Archimedes and Boethius" would now be transformed into "Archimedes *et al.*" The random order convention would put an end to such acts of seeming generosity, depriving later authors of a more readily available source of research papers.

A full accounting of these and other strategic factors in the selection of co-authors demands a model; one of us has indeed written down a set of notes to this effect; Ray (2013). But that said, there seems to not be

[20]It would be useful to find some way of equalizing credit for *past* publications, where it is not possible to adopt the present suggestion.

[21]We are grateful to Sahar Parsa and Phil Reny, both lexicographically challenged and clearly on the lookout for unsuspected dangers, for this point.

much evidence that supports the hypothesis that late co-authors are invited with lexicographic considerations in mind. Figure 1 plots papers per author by first letter of last name, both for all papers and for all joint papers. Apart from from the fact that authors beginning with Q appear to be relatively prolific and an intriguing little bump-up in co-authored papers by Z-authors, there is really no trend here at all.

6.6. **Three Or More Authors.** The entire analysis in this paper has been for the case of two authors. While we foresee no eventual difficulty in extending the analysis to the case of three or more authors, there are enough conceptual issues in doing so that we have decided to omit a formal analysis. For instance, one would have to deal with hybrid issues of the form:

[Zeno, Boethius]Ⓡ and Archimedes

that are best left unexplored. Common sense suggests that if three or more authors choose to randomize, they should sort out such sub-coalitional issues first and then randomize the entire list; e.g.,

Zeno, Archimedes and BoethiusⓇ

leaving matters at that. The extension of the analysis in this paper to such conventions should be relatively straightforward.

6.7. **A Last Word on Institutions.** The role of institutional approval for this proposal should completely obvious. It is essential that leading journals or associations agree on the semantics of the Ⓡ symbol and that the the use of the symbol be made freely available to all authors.

However, it is assuredly *not* essential that institutions mandate the use of randomization. It should simply be offered as an additional instrument to the authors, and the rest will take care of itself. At least, that is what the theory here argues.

Journals can also offer to carry out the randomization for authors, thereby ensuring that there are no post-randomization squabbles or denials. After all, while authors might graciously agree to randomize *ex ante*, there is no guarantee that they will want to abide by the results *ex post*.

## 7. Conclusions

In this paper we describe a scheme — random order — for assigning credit to papers with two coauthors. We first characterize the current system of joint authorship as modified-alphabetical, where the author who is earlier in the alphabet can offer first authorship to the other, if the contributions are very unequal. This is motivated by a sense of guilt or fairness on the part of the earlier author.

The new scheme involves flipping a coin to determine first authorship and adding the notation Ⓡ to the list of the two authors when this has been done. In addition, we allow either author to offer first authorship to the other, without the Ⓡ notation, again motivated by a sense of guilt when the contributions have been extremely unequal.

We show that if such a convention arises as a mutant in the present system, it can enter on the basis of a "equilibrium deviation", a concept related to neologism-proof equilibrium. On the other hand, there is no such possibility of reverting to the old equilibrium once in the new equilibrium. In short, there is no issue of imposing such a system. We claim that if it is offered, it will be adopted "in equilibrium." Moreover, we show that the new equilibrium not only entails equal expected utilities for the the two players, but involves a higher sum of expected utilities than does the old. The new mechanism would then be strictly preferred on the basis of plausible social welfare criteria.

The beauty of the mechanism Ⓡ is that it does not demand any more of the agents than does the present economics convention $E$, despite being fairer and more efficient. The convention Ⓡ simply allows *either* player to concede first authorship, instead of allowing only the first author to have this option, as in the convention $E$. Although such an option can lead to arguments, it is indeed exercised on occasion in reality.[22]

[22]Flipping a coin and adding Ⓡ to the list of authors also entail costs, but these seem trivial.

It should be noted that the analysis in this paper only focuses on the "end points;" i.e., a situation in which alphabetical order prevails or random order prevails, and we have examined potential deviations from these end points. A more complete analysis would need to examine the movement of the full dynamical system in which both systems could conceivably co-exist. The difficulty lies in describing status quo choices for each other: if they disagree, what can they insist on? For instance, in the transition from an old to a new convention, the status quo would presumably switch at some point from the old convention to the new. It seems that this would speed up subsequent adoption of the new convention, though in this paper we have not studied a formal model to that effect (see, however, Section 6.2).

In summary, random order appears to maintain all the ethical niceties of alphabetical order, but in addition: (a) it distributes the psychological and perceptual weight given to first authorship evenly over the alphabet, (b) it allows *either* author to signal credit when contributions are extremely unequal, (c) it will be willingly adopted even in an environment where alphabetical order is dominant, (d) it is robust to deviations, (e) it dominates alphabetical order on the grounds of ex-ante efficiency, and (f) barring the addition of a simple symbol, it is no more complex than the old system, and brings perfect symmetry to joint authorship.

## References


Carney, D. R. and Banaji, M. R. (2012), "First is Best," *PLoS ONE* **7**(6), e35088. doi:10.1371/journal.pone.0035088

Chambers, R., Boath, E., and Chambers, S. (2001), "The A to Z of Authorship: Analysis of Influence of Initial Letter of Surname on Order of Authorship," *British Medical Journal* 323(22-29 Dec.), 1460–1461. doi:10.1136/bmj.323.7327.1460

Einav, L. and L. Yariv (2006), "What's in a Surname? The Effects of Surname Initials on Academic Success," *Journal of Economic Perspectives* **20**, 175–188.

Engemann, K. and H. Wall (2009), "A Journal Ranking for the Ambitious Economist," *Federal Reserve Bank of St. Louis Review* **91**, 127–39.

Engers, Maxim; Gans, Joshua S.; Grant, Simon; and King, Stephen P. (1999) "First-Author Contributions," *Journal of Political Economy* 107, 859-883.

Farrell, Joseph. (1993). "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior* **5**, 514-531.

Feenberg, Daniel R., Ganguli, Ina, Gaule, Patrick and Jonathan Gruber (2015) "It's Good to be First: Order Bias in Reading and Citing NBER Working Papers," NBER Working Paper No. 21141.

Haque, A., and Ginsparg, P. (2009), "Positional Effects on Citation and Readership in arXiv," *Journal of the American Society for Information Science and Technology* **60** (11), 2203–2218. doi:10.1002/asi.21166

Itzkowitz, J., Itzkowitz, J. and Rothbort, S. (2016), "ABCs of Trading: Behavioral Biases Affect Stock Turnover and Value," *Review of Finance*, forthcoming.

Jacobs, H. and Hillert, A. (2016), "Alphabetic Bias, Investor Recognition, and Trading Behavior," *Review of Finance* **20**, 693–723.

Ray, D. (2013), "All the Names: Some Strategic Consequences of Alphabetical Order in Joint Research," mimeo., New York University.

van Praag, C. M. and van Praag, B. M. S. (2008), "The Benefits of Being Economics professor A (Rather than Z)," *Economica* **75**, 782–796. doi:10.1111/j.1468-0335.2007.00653.

Weber, Matthias (2016) "The Effects of Listing Authors in Alphabetical Order: A Survey of the Empirical Evidence." SSRN: http://ssrn.com/abstract=2803164, or http://dx.doi.org/10.2139/ssrn.2803164

## 8. Appendix: Proofs

*Proof of Theorem 2.* Recall we have fixed the convention $E$ and its associated reversal threshold $\epsilon$, given by (4.4), and are supposing that society assigns a credit pair $(a, b) = (a, 1-a)$ to the observation of random order.

**Lemma 1.** *There exists $\bar{a} \in (\epsilon/2, [1+\epsilon]/2)$ with the following properties: on $[\epsilon/2, \bar{a}]$, there is a continuous function $x^\alpha(a)$ taking values in $(0,1)$, such that Archimedes prefers random order over $\alpha$ whenever realized contributions $(x, y)$ satisfy $x \in [0, x^\alpha(a)]$, and strictly prefers $\alpha$ to random order when $x > x^\alpha(a)$. Moreover,*

$$x^\alpha(\epsilon/2) > \epsilon, \tag{8.1}$$

*while away from this end-point,*

$$x^\alpha(a) > a \text{ for all } a \in [\epsilon/2, \bar{a}) \text{ and } x^\alpha(\bar{a}) = \bar{a}. \tag{8.2}$$

*Proof.* Random order at realization $(x, y)$ yields an expected payoff to Archimedes of

$$a + \frac{\delta}{2} - \Gamma(y - b) = a + \frac{\delta}{2} - \Gamma\left(y - [1-a]\right) \tag{8.3}$$

while alphabetical order generates a payoff of

$$\frac{1+\epsilon}{2} + \delta - \Gamma\left(y - \frac{1-\epsilon}{2}\right)$$

as described in (4.1). Therefore random order is weakly preferred to $\alpha$ if

$$\Gamma\left(y - \frac{1-\epsilon}{2}\right) - \Gamma\left(y - [1-a]\right) + a \geq \frac{1+\epsilon}{2} + \frac{\delta}{2}. \tag{8.4}$$

If this inequality holds for some $y$, then *equality* must hold for some $y^\alpha(a)$, because for $y$ small enough the inequality is always strictly reversed.[23] Let $x^\alpha(a) \equiv 1 - y^\alpha(a)$. (If (8.4) fails for all $y$, set $x^\alpha(a) = 0$.) Because $\Gamma(z)$ is strictly convex when $z > 0$, the LHS of (8.4) is *strictly* increasing in $y$ (and decreasing in $x$). That shows that Archimedes will indeed prefer random order to $\alpha$ when $x \in [0, x^\alpha(a)]$, and $\alpha$ to random order when $x > x^\alpha(a)$.

To establish (8.1), set $a = \epsilon/2$ and $x = \epsilon$, so that $y = 1 - \epsilon$. Then

$$\begin{aligned}
\Gamma\left(y - \frac{1-\epsilon}{2}\right) - \Gamma\left(y - [1-a]\right) + a &= \Gamma\left(\frac{1-\epsilon}{2}\right) - \Gamma\left(-\frac{\epsilon}{2}\right) + \frac{\epsilon}{2} \\
&= \frac{1}{2} + \delta + \frac{\epsilon}{2} \\
&> \frac{1+\epsilon}{2} + \frac{\delta}{2},
\end{aligned}$$

where the second inequality employs the definition of $\epsilon$ in (4.4).[24] That shows that (8.4) holds strictly when $x = \epsilon$. Because the left-hand side of (8.4) is decreasing in $x$, (8.1) is true.

Notice moreover that $x^\alpha(a)$ moves continuously in $a$ as long as it is strictly positive. At the same time, (8.4) must *strictly* fail for every $y$ when $a = (1+\epsilon)/2$. By the Intermediate Value Theorem, there exists a *first* $\bar{a} \in (\epsilon/2, [1+\epsilon]/2)$ such that $x^\alpha(\bar{a}) = \bar{a}$. It must be that $x^\alpha(a) > a$ for all $a \in [\epsilon/2, \bar{a})$, which establishes (8.2) and completes the proof. □

Our next lemma establishes a corresponding threshold for the comparison of random order and reverse-alphabetical order $\rho$. We will only need to work on the domain $[\epsilon/2, \bar{a}]$.

[23] Recall that $\Gamma(z) = 0$ for all $z \leq 0$, and continuous everywhere. This, combined with $a \leq (1+\epsilon)/2$, guarantees that (8.4) must fail for $y$ small enough.

[24] Intuitively, the credits are assigned are the same as in $\rho$, and Archimedes is indifferent between $\alpha$ and $\rho$ at $\epsilon$. So he will strictly prefer random order, which yields the same credit to Boethius but gives Archimedes the extra payoff of $\delta$ half the time.

**Lemma 2.** *There is a continuous function $x^\rho(a)$ on $[\epsilon/2, \bar{a}]$ such that Archimedes prefers random order to $\rho$ if $x \geq x^\rho(a)$, and strictly prefers $\rho$ to random order if $x < x^\rho(a)$. Moreover,*

$$x^\rho(\epsilon/2) = 0, \text{ and } x^\rho(a) < a \text{ for all } a \in [\epsilon/2, \bar{a}]. \tag{8.5}$$

*Proof.* Reverse order $\rho$ yields a payoff to Archimedes given by (4.2), which is

$$\frac{\epsilon}{2} - \Gamma\left(y - \left[1 - \frac{\epsilon}{2}\right]\right).$$

Combining with (8.3), we see that random order is weakly preferred to $\rho$ if

$$\Gamma\left(y - [1-a]\right) - \Gamma\left(y - \left[1 - \frac{\epsilon}{2}\right]\right) - a \leq \frac{\delta}{2} - \frac{\epsilon}{2}. \tag{8.6}$$

Again, because $\Gamma(z)$ is strictly convex when $z > 0$, the LHS of (8.6) is strictly increasing in $y$ (and so decreasing in $x$) thereby showing that (8.6) holds (if it holds at all) over some interval of the form $[x^\rho(a), 1]$. $x^\rho(a)$ is either zero (if (8.6) always holds), or is the equality solution $y = 1 - x^\rho(a)$ if (8.6) holds for some $y$. (The remaining possibility, that (8.6) never holds, is ruled out by a parallel argument to that in Footnote 23, this time using $a \geq \epsilon/2$.)

Finally, we establish (8.5). Note that when $a = \epsilon/2$, the same credits are associated to random order as with reversal. So guilt is the same whether Archimedes reverses or randomizes, but in the latter case he at least picks up the $\delta$ payoff half the time. So he will prefer random order. Formally, at $a = \epsilon/2$, (8.6) holds for all values of $(x, y)$, so $x^\rho(\epsilon/2) = 0$.

To establish the second part of (8.5), suppose that at $(a, b)$, the realized contributions are *also* $(a, b)$. Setting $y = b = 1 - a$, we see that

$$\Gamma\left(y - [1-a]\right) - \Gamma\left(y - \left[1 - \frac{\epsilon}{2}\right]\right) - a = \Gamma(0) - \Gamma\left(\left[\frac{\epsilon}{2} - a\right]\right) - a \leq -\frac{\epsilon}{2} < \frac{\delta}{2} - \frac{\epsilon}{2},$$

which shows that (8.6) is satisfied with strict inequality when $x = a$.[25] Therefore $x^\rho(a) < a$.

□

Let $\phi(a)$ denote the conditional expected contribution by Archimedes over all values of $x$ for which Archimedes prefers random order to $\alpha$ or $\rho$. This is just the expectation of $x$ conditional on $x$ lying in the interval $[x^\rho(a), x^\alpha(a)]$. Lemmas 1 and 2 together tell us that $x^\rho(a) < a \leq x^\alpha(a)$ whenever $a \in [\epsilon/2, \bar{a}]$, so $\phi$ is well-defined on the entire domain $[\epsilon/2, \bar{a}]$. Because $x^\alpha$ and $x^\rho$ are continuous, so is $\phi$. We know from (8.1) that $x^\alpha(\epsilon/2) > \epsilon$, and we know from (8.5) that $x^\rho(\epsilon/2) = 0$, so it follows that

$$\phi\left(\frac{\epsilon}{2}\right) > \frac{\epsilon}{2}. \tag{8.7}$$

We also know from (8.2) that $x^\alpha(\bar{a}) = \bar{a}$, so that

$$\phi(\bar{a}) \leq \bar{a}. \tag{8.8}$$

Combining (8.7) and (8.8) and invoking the continuity of $\phi$, we must conclude that there exists $a^* \in [\epsilon/2, \bar{a}]$ such that

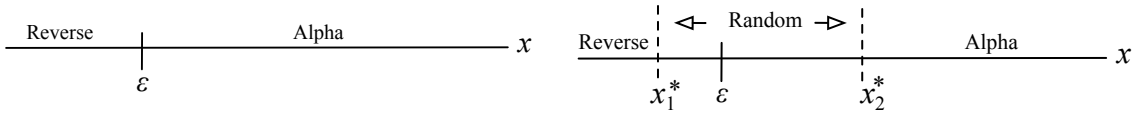
$$\phi(a^*) = a^*.$$

Define $x_1^* = x^\rho(a^*)$ and $x_2^* = x^\alpha(a^*)$. See Figure 2. Archimedes will willingly wish to announce random order whenever $x$ lies in the interval $(x_1^*, x_2^*)$, provided the public assigns credit of $(a^*, 1 - a^*)$. That is, he prefers Ⓡ to both $\alpha$ and $\rho$ in this zone.

**Lemma 3.** *Boethius also prefers Ⓡ to $\alpha$ in the region $(x_1^*, x_2^*)$.*[26]

[25] The weak inequality in the chain uses the fact that $a \geq \epsilon/2$ and the assumption that $\Gamma(z) = 0$ when $z \leq 0$.

[26] Boethius might continue to prefer $\rho$ over Ⓡ.




**(a)** Existing Regime **(b)** Deviation to Random Order

**Figure 2.** Incentives to Deviate to Random Order

*Proof.* Note that Boethius receives the payoff $(1-\epsilon)/2$ under $\alpha$,[27] while under random order his lowest possible payoff is $b^* + (\delta/2) - \Gamma(x_2^* - a^*) = (1-a^*) + (\delta/2) - \Gamma(x_2^* - a^*)$ (this corresponds to the highest contribution $x_2^*$ by Archimedes for which Archimedes prefers random order over $\alpha$). Comparing these, it will be sufficient to show that

$$\frac{\delta}{2} + \left(\frac{1+\epsilon}{2} - a^*\right) \geq \Gamma(x_2^* - a^*). \tag{8.9}$$

Recall that Archimedes himself is indifferent over the two options at $(x_2^*, y_2^*)$, so that

$$\delta + \frac{1+\epsilon}{2} - \Gamma\left(y_2^* - \frac{1-\epsilon}{2}\right) = \frac{\delta}{2} + a^* - \Gamma(y_2^* - b^*) = \frac{\delta}{2} + a^*,$$

(where the second equality uses the observation that $y_2^* \leq b^*$). Transposing terms and using $x^* = 1 - y^*$,

$$\frac{\delta}{2} + \left(\frac{1+\epsilon}{2} - a^*\right) = \Gamma\left(y^* - \frac{1-\epsilon}{2}\right) = \Gamma\left(\frac{1+\epsilon}{2} - x_2^*\right). \tag{8.10}$$

To establish (8.9), then, it will be equivalent to show that $x_2^* - a^* \leq (1+\epsilon)/2 - x_2^*$, or that

$$x_2^* \leq \frac{1}{2}\left[a^* + \frac{1+\epsilon}{2}\right],$$

but given (8.10), it means that it is also equivalent to show that

$$\frac{\delta}{2} + \left(\frac{1+\epsilon}{2} - a^*\right) = \Gamma\left(\frac{1+\epsilon}{2} - x_2^*\right) \geq \Gamma\left(\frac{1}{2}\left[\frac{1+\epsilon}{2} - a^*\right]\right) \tag{8.11}$$

Because $\Gamma$ is convex, it will suffice to show that (8.11) holds at the *largest* possible value of $\frac{1+\epsilon}{2} - a^*$, which is precisely $1/2$ (recall $a^* \geq \epsilon/2$). But setting this term equal to $1/2$, (8.10) reduces to the requirement that

$$\frac{\delta}{2} + \frac{1}{2} \geq \Gamma(1/4),$$

which is guaranteed by the second end-point condition in (3.3). □

We now complete the proof of the theorem by showing that in the zone $(x_1^*, x_2^*)$, Ⓡ is the unique SPE outcome. We show that Archimedes can guarantee Ⓡ by announcing the default $\alpha$. For Boethius, $\alpha$ cannot be a best response; by announcing Ⓡ (and having Archimedes subsequently agree — which he will do), Boethius can implement Ⓡ, which he prefers (Lemma 3). The only other option is for Boethius to announce $\rho$, in which case Archimedes can reject and counterpropose Ⓡ. By Lemma 3 again, Boethius will accept this counterproposal. So Archimedes can guarantee Ⓡ. Because Ⓡ is Archimedes's best option in the zone $(x_1^*, x_2^*)$, it must therefore be the unique equilibrium.

We have therefore established that there is a disruption of $E$ by an "equilibrium deviation" to random order, in which Archimedes and Boethius both willingly participate, completing the proof of Theorem 2. □

*Proof of Theorem 3.* Consider a random order convention with threshold $\epsilon$. If realized credits are $(x, 1-x)$, Archimedes's total payoff from Ⓡ at $x$ is $1/2 + \delta/2 - \Gamma(1 - x - 1/2) = 1/2 + \delta/2 - \Gamma(1/2 - x)$. Similarly,

[27] There is no guilt loss, because $x \leq (1+\epsilon)/2$ for every $x \in [x_1^*, x_2^*]$.

Archimedes's overall payoff from $\rho$ is $\epsilon/2 - \Gamma(1-x-(1-\epsilon/2)) = \epsilon/2 - \Gamma(\epsilon/2 - x)$, which simply equals $\epsilon/2$ if $x \geq \epsilon/2$. With these payoffs in mind, consider the function

$$\Delta(\epsilon) = \frac{1}{2} + \frac{\delta}{2} - \epsilon/2 - \Gamma\left(\frac{1}{2} - \epsilon\right).$$

By the end-point conditions in (3.3), $\Delta(0) = 1/2 + \delta/2 - \Gamma(1/2) < -\delta/2 < 0$ and $\Delta(1/2) = 1/2 + \delta/2 - \Gamma(0) - 1/4 > 0$. Moreover, $\Delta$ is concave because $\Gamma$ is convex. It follows that there exists a *unique* $\epsilon^* \in (0, 1/2)$ such that $\Delta(\epsilon^*) = 0$. That is,

$$\frac{1}{2} + \frac{\delta}{2} - \epsilon^*/2 = \Gamma\left(\frac{1}{2} - \epsilon^*\right), \tag{8.12}$$

Observe that the three ranges $[0, \epsilon^*], [\epsilon^*, 1-\epsilon^*]$ and $[1-\epsilon^*]$ are all non-empty.

Consider first the range $x < \epsilon^*$. We show that in this range, Archimedes strictly prefers $\rho$ to Ⓡ, while Boethius strictly prefers $\rho$ to either Ⓡ or $\alpha$.

To this end, observe that $\rho$ yields Archimedes $\epsilon^*/2 - \Gamma(1-x-(1-\epsilon^*/2)) = \epsilon^*/2 - \Gamma(\epsilon^*/2 - x)$. On the other hand, Ⓡ yields Archimedes $1/2 + \delta/2 - \Gamma(1/2 - x)$. If we define $\tilde{\Delta}(x) = \epsilon^*/2 - \Gamma(\epsilon^*/2 - x) - 1/2 - \delta/2 + \Gamma(1/2 - x)$, then the convexity of $\Gamma$ implies that $\tilde{\Delta}$ is a decreasing function of $x$. Hence $\tilde{\Delta}(x) > 0$ since $\tilde{\Delta}(\epsilon^*) = \epsilon^*/2 - 1/2 - \delta/2 + \Gamma(1/2 - \epsilon^*) = 0$, using $\Delta(\epsilon^*) = 0$. Hence Archimedes strictly prefers $\rho$ to Ⓡ when $x < \epsilon^*$.

Turning now to Boethius's preferences in the range $x < \epsilon^*$, we first show that Boethius strictly prefers $\rho$ to Ⓡ. Boethius's lowest payoff under $\rho$ occurs when $x = \epsilon^*$; it is $1 - (\epsilon^*/2) + \delta - \Gamma(x - \epsilon^*/2) = 1 - (\epsilon^*/2) + \delta - \Gamma(\epsilon^*/2)$. Under Ⓡ, it is $(1/2) + (\delta/2) - \Gamma(x - 1/2) = (1/2) + (\delta/2)$. Consequently we need to show that

$$\frac{1}{2} + \frac{\delta}{2} - \epsilon^*/2 > \Gamma(\epsilon^*/2). \tag{8.13}$$

Given (8.12) and $\epsilon^* < 1/2$, (8.13) follows as long as $\epsilon^* < 1/3$. Given the properties of the function $\Delta$ established above, it suffices to show that $\Delta(1/3) > 0$. But

$$\Delta(1/3) = \frac{1}{3} + \frac{\delta}{2} - \Gamma(1/6) \tag{8.14}$$

and this is indeed positive, given the convexity of $\Gamma$ and the second end-point condition in (3.3).[28] In short, $\epsilon^* \in (0, 1/3)$, so that (8.13) applies, and Boethius strictly prefers $\rho$ to Ⓡ whenever $x < \epsilon^*$.

To complete the proof of the claim, we now show that Boethius strictly prefers $\rho$ to $\alpha$ when $x < \epsilon^*$. For $\rho$ yields Boethius a payoff of $\delta + 1 - \epsilon^*/2 - \Gamma(x - \epsilon^*/2)$, whereas $\alpha$ yields just $\epsilon^*/2$. The desired result then follows if $\delta + 1 - \epsilon^* > \Gamma(x - \epsilon^*/2)$. This, in turn, follows from (8.12): for $x < \epsilon^*$, $2\Gamma(\epsilon^*/2) > \Gamma(\epsilon^*/2) > \Gamma(x - \epsilon^*/2)$, so that $\delta + 1 - \epsilon^* > \Gamma(x - \epsilon^*/2)$.

We now apply this claim to verify that $\rho$ is the unique SPE outcome in the range $x < \epsilon^*$. Suppose that Boethius chooses $d$ at Stage 2. If Archimedes also chooses $d$, then $d$ is the outcome. If Archimedes chooses $\alpha$ at Stage 2, Boethius rejects this at Stage 3, and counteroffers $\rho$. Archimedes accepts this and $\rho$ is the outcome. If Archimedes chooses $\rho$ at Stage 2, Boethius accepts this and $\rho$ is the outcome. Hence Archimedes has $\alpha$ or $\rho$ as best replies to Boethius choosing $d$, both of which choices yield $\rho$. Since $\rho$ is Boethius's best outcome, this is the unique SPE outcome.

Entirely analogous arguments apply to the range $x \in (1 - \epsilon^*, 1]$, where $\alpha$ is the unique SPE outcome.

Consider, finally, the range $x \in [\epsilon^*, 1 - \epsilon^*]$. Suppose at first that $x \in [\epsilon^*, 1/2]$. Boethius prefers Ⓡ to either $\alpha$ or $\rho$ (that is, $d$=Ⓡ is Boethius's favorite outcome).[29] To show this, note that in this range, Ⓡ yields Boethius a payoff of $\delta/2 + 1/2$; $\alpha$ yields $\epsilon^*/2$ and $\rho$ yields $\delta + 1 - \epsilon^*/2 - \Gamma(x - \epsilon^*/2)$. Hence Boethius strictly prefers Ⓡ to $\alpha$ simply because $\epsilon^* < 1$. To show that Boethius weakly prefers Ⓡ to $\rho$, it is enough to show that $\Gamma(x - \epsilon^*/2) \geq \delta/2 + 1/2 - \epsilon^*/2 = \Gamma(\epsilon^*/2)$, where the equality uses (8.12). But this follows since $x \geq \epsilon^*$.

[28] If $f$ is concave and $f(0)$ and $f(z)$ are both nonnegative for some $z > 0$, then so is $f(\lambda z)$ for any $\lambda \in (0, 1)$. The function $f(z) \equiv 2z + (\delta/2) - \Gamma(z)$ is concave, with $f(0)$ and $f(1/4)$ both nonnegative, the latter by the second end-point condition in (3.3). Setting $\lambda = 2/3$, it follows that $f(1/6) = (1/3) + (\delta/3) - \Gamma(1/6) \geq 0$. Therefore $\Delta(1/3) > f(1/6) \geq 0$.

[29] In this range, Archimedes prefers Ⓡ to $\rho$, but this observation is not needed here.

It follows that Boethius can guarantee $d$ as an SPE outcome by choosing $d$ at Stage 2, and then refusing any alternative offer made by Archimedes.

An analogous argument with the roles reversed applies in the range $x \in [1/2, 1-\epsilon^*]$, so that Archimedes can guarantee his favorite outcome $d$ by choosing $d$ at Stage 1. This completes the proof of Theorem 3. □

*Proof of Theorem 4.* Consider first the total expected payoff under $E$, with threshold $\epsilon$ as in Theorem 1. There are four relevant ranges for $x$:

If $x \in [0, \epsilon/2)$, then Archimedes's payoff is $\epsilon/2 - \Gamma(\epsilon/2 - x)$; whereas Boethius's payoff is $1 - \epsilon/2 + \delta$.

If $x \in [\epsilon/2, \epsilon)$, Archimedes's payoff is $\epsilon/2$; whereas Boethius's payoff is $1 - \epsilon/2 + \delta - \Gamma(x - \epsilon/2)$.

If $x \in [\epsilon, (1+\epsilon)/2)$, Archimedes's payoff is $(1+\epsilon)/2 + \delta - \Gamma((1+\epsilon)/2 - x)$; whereas Boethius's payoff is $(1-\epsilon)/2$.

If $x \in [(1+\epsilon)/2, 1]$, Archimedes's payoff is $(1+\epsilon)/2+\delta$; whereas Boethius's payoff is $(1-\epsilon)/2-\Gamma(x-(1+\epsilon)/2)$.

Hence the total expected payoff under $E$ is given by

$$\begin{aligned} W(E) &\equiv 1 + \delta - \int_0^{\epsilon/2} \Gamma(\epsilon/2 - x)dx - \int_{\epsilon/2}^{\epsilon} \Gamma(x - \epsilon/2)dx \\ &\quad - \int_{\epsilon}^{(1+\epsilon)/2} \Gamma((1+\epsilon)/2 - x)dx - \int_{(1+\epsilon)/2}^{1} \Gamma(x - (1+\epsilon)/2)dx \\ &= 1 + \delta - 2\int_0^{\epsilon/2} \Gamma(x)dx - 2\int_0^{(1-\epsilon)/2} \Gamma(x)dx, \end{aligned} \tag{8.15}$$

where the second equality follows from a suitable change in variables.

For the Ⓡ equilibrium, there are three signals, with each signal range divided into two halves. Again, overall expected utility depends on the integral of the guilt function over each of these ranges. An argument analogous to the one used to obtain (8.15) also applies to obtain the total expected payoff under Ⓡ:

$$W(\text{Ⓡ}) \equiv 1 + \delta - 4\int_0^{\epsilon^*/2} \Gamma(x)dx - 2\int_0^{1/2-\epsilon^*} \Gamma(x)dx. \tag{8.16}$$

We must compare $W(\text{Ⓡ})$ to $W(E)$. Begin by noting that $\epsilon^* > \epsilon/2$. Suppose, on the contrary, that $\epsilon \geq 2\epsilon^*$. From (3.3), we know that $\Gamma(1/2 - \epsilon^*) + \epsilon^*/2 = 1/2 + \delta/2$. It follows that $\Gamma((1-\epsilon)/2) \leq \Gamma(1/2 - \epsilon^*) = 1/2 + \delta/2 - \epsilon^*/2 < 1/2 + \delta$, which contradicts (4.4). Therefore $\epsilon^* > \epsilon/2$, as claimed, and with this in hand we complete the proof.

Define

$$D \equiv \left[\int_0^{\epsilon/2} \Gamma(x)dx + \int_0^{(1-\epsilon)/2} \Gamma(x)dx\right] - \left[2\int_0^{\epsilon^*/2} \Gamma(x)dx + \int_0^{1/2-\epsilon^*} \Gamma(x)dx\right].$$

Given (8.15) and (8.16), it suffices to show that

$$D > 0.$$

For clarity, we consider two cases. Suppose first that $\epsilon^* < \epsilon$. Then, using $\epsilon^* > \epsilon/2$, we see that

$$D = \int_{\epsilon^*/2}^{\epsilon/2} \Gamma(x)dx + \int_{1/2-\epsilon^*}^{1/2-\epsilon/2} \Gamma(x)dx - \int_0^{\epsilon^*/2} \Gamma(x)dx.$$

The total length of the intervals over which the positive integrals are taken is $\epsilon^*/2$, which is the length of the interval over which the negative integral is taken. In addition, because $\epsilon^* \leq 1/3$, the smallest value of $\Gamma(x)$ in the second integral — which is $\Gamma(1/2 - \epsilon^*)$ — is no smaller than the largest value of $\Gamma(x)$ from the negative integral, $\Gamma(\epsilon^*/2)$. It follows that $D > 0$ in this case.

Finally, suppose that $\epsilon^* \geq \epsilon$. In this case,

$$D = -\int_{\epsilon/2}^{\epsilon^*/2} \Gamma(x)dx + \int_{1/2-\epsilon^*}^{1/2-\epsilon/2} \Gamma(x)dx - \int_0^{\epsilon^*/2} \Gamma(x)dx,$$

again using $\epsilon^* > \epsilon/2$. The length of the interval over which the positive integral is taken is again equal to the combined length of the intervals over which the two negative integrals are taken. The smallest value of $\Gamma$ in the positive integral, $\Gamma(1/2 - \epsilon^*)$, is no less than the largest value of $\Gamma$ in either negative integral, which is $\Gamma(\epsilon^*/2)$, because $\epsilon^* \leq 1/3$. It follows that $D > 0$ yet again, completing the proof of the theorem. $\square$