Random Order:
Or, How to Co-Author If You Must\textsuperscript{1,2}

Debraj Ray and Arthur Robson\textsuperscript{®}
New York University and Simon Fraser University

June 2016

1. Introduction

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 the names is a signal that the names are in random order. The purpose of this paper is to argue two things:

1. Alphabetical order may be an acceptable system in that it avoids imputations of relative contributions, and so encourages joint research without excessive squabbles. But random order does better.

2. That random order, if offered as an option, will overturn alphabetical order as a convention, much as a mutant can invade a population.

2. Background

Many years ago, when Debraj worked at Boston University and Arthur visited there, we spent one of our many enjoyable lunches together railing against the indignities of alphabetical order, which is the dominant name-ordering convention for economics publications. A quick perusal of our last names will suffice to explain why we railed. To add insult to injury, one of us (Ray) had just been enthusiastically recommended a “wonderful paper” by Banerjee \textit{et al} (2001), on which he was a co-author.\textsuperscript{3}

\begin{footnote}{\textsuperscript{1}Ray thanks the National Science Foundation for research support under grant number SES-1261560. Robson thanks the Canada Research Chairs Program and the Social Sciences and Humanities Research Council of Canada. We both thank Ariel Rubinstein and Leeat Yariv, whose names are given alphabetically here, for helpful comments.

\textsuperscript{2}Relevant anecdotes were provided by Noah Williams and Stanley Zin, unsurprisingly, perhaps. Zin describes his delight at being about to break the curse of the perennial last listed author by writing with Irina Zviadadze of Stockholm, who is, indeed, the fourth to last listed Member of the Econometric Society. We commiserate with Yanos Zylberberg of Bristol who is listed dead last. It is no doubt a coincidence that he has sole-authored 3 out of the 4 working papers listed for him on RePEc. Zin also relates how Noah Williams recently lost a first-name tie-breaker to remain in his customary last position. We salute, in connection with ties, the facetious contribution of Goodman, Goodman, Goodman, and Goodman (2014), whose names warrant listing in full, regardless of the norm in effect, for obvious reasons. They note, for example, that a reference to Goodman \textit{et al} (2014) disparages the contribution of no author. Noah Williams, in turn, refers us to the astoundingly lexicographically blessed Dr. Aad, of Marseille, who has appeared as the lead author on 458 scientific papers, on which alphabetic order was adhered to. \textsuperscript{3}

\textsuperscript{3}Teetering delicately on the brink of empirical work, we decided on the following test. Find the median Fellow of the Econometric Society and compare his or her name to the median Member of the ES. Discrimination against those lower in the alphabet, that is, would mean the median Fellow was later in the alphabet than the median Member. However, the median Fellow, Ed Leamer, was also the median Member! Since then more careful and extensive examinations of the issue have...\end{footnote}
Alphabetical order is, all things considered, a good arrangement. Our colleagues from other disciplines express wonderment that such a self-centered subspecies — the academic economist — actually uses this civilized convention. Table 1 reports the prevalence of alphabetical order in economics. Around 85% of two-author papers are written in alphabetical order. That percentage falls with more authors, possibly capturing the fear of the notorious et al effect, or there could be research assistants involved.\(^4\) Compare this to the cutthroat nature of much of the 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.\(^5\)

And yet, there are features of alphabetical order that create significant and unwarranted advantages for names earlier that appear earlier in the alphabet:

1. Psychologically, names that appear first are more likely to be given extra credit given that society as a whole appears to be attuned to merit-based rankings. This is certainly on line with research on marketing: products presented earlier exhibit higher probabilities of selection, as the aptly ordered article by Carney and Banaji (2012) observes.

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. Another aptly ordered paper by Haque and Ginsparg (2009) note that article positioning in the ArXiv repository is correlated with citations of that article. Feenberg, Ganguli, Gaule, and Gruber (2015) demonstrate established a strong case that such discrimination exists, so we need no longer rely on our eager but limited empirical skills. A recent survey of the evidence is by Weber (2016), whose last name starts with W. Linda Welling, for whom this is also true, provided this lead.

\(^4\)The et al effect alone is possibly captured better by papers in which the first author is out of alphabetical order; this percentage is lower as Table 1 reveals.

\(^5\)That said, 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.

---

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).

<table>
<thead>
<tr>
<th>Number of Authors</th>
<th>Two</th>
<th>Three</th>
<th>Four</th>
<th>Five</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>53858</td>
<td>17857</td>
<td>1865</td>
<td>340</td>
</tr>
<tr>
<td>Alphabetical</td>
<td>45337</td>
<td>13124</td>
<td>1155</td>
<td>163</td>
</tr>
<tr>
<td>Non-Alphabetical</td>
<td>8521</td>
<td>4733</td>
<td>710</td>
<td>177</td>
</tr>
<tr>
<td>% Non-Alpha</td>
<td>15.82</td>
<td>26.51</td>
<td>38.07</td>
<td>52.06</td>
</tr>
<tr>
<td>First Author Non-Alpha</td>
<td>8521</td>
<td>2754</td>
<td>339</td>
<td>95</td>
</tr>
<tr>
<td>% First Author Non-Alpha</td>
<td>15.82</td>
<td>15.42</td>
<td>18.18</td>
<td>27.94</td>
</tr>
</tbody>
</table>
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.\(^6\)

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 an 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. 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 nevertheless damaging shortcut.

Is there any evidence that these considerations matter at all, or is this the resentful fantasy of two disgruntled economists with surnames far down the alphabet? There is, actually, quite a bit of evidence. In a paper published in the *Journal of Economic Perspectives*, Einat 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 Einat-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 Einat 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

\(^6\)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.”
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.”

A quite different theoretical model based on Nash bargaining is due to Engers et al (1999). They have no counterpart to the direct gain from first-authorship that we will assume. Indeed, under the Engers et al approach, a modified alphabetical system can disadvantage the author who is earlier in the alphabet, because a reversal can be used to signal a higher contribution by the other author, but there is no comparable signal for the earlier author.7

The observations that Einav and Yariv (2006) present, however, seem to require such a first-author premium. That is, although natural, 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.8

3. Why Change The Convention?

By the norms of Pareto-optimality, it is possible to argue that alphabetical order is indeed an efficient arrangement: earlier names are better off. So why care? Economists who eschew interpersonal comparisons of utility (and especially those with earlier surnames), are likely to oppose any change in alphabetical order, arguing that existing alternatives would lead to nasty fights, grumpy co-authors, and ultimately a disintegration of the happy system of joint research that we all admire.

There are at least three responses to this point of view. First, one might argue that the game in question isn’t zero-sum. After all, people 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 take here. (It would add to our arguments, but in a fairly standard way.) It is unclear that moral hazard in research effort is first-order; given co-authorship, some people naturally work hard, and others don’t, and the shirkers are strewn all over the alphabet.

Second, the fear of being relegated to second or third author, or even the dreaded et al dungeon, might discourage authors from writing papers with those more fortunately placed in the alphabet. Einav and Yariv (2004) provide some evidence that individuals further down the alphabet are more averse to writing with multiple co-authors. Again, this is an efficiency loss.

Third, when ordering is alphabetical but relative contributions are not, 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 always a fall-back option, it is unclear that the “disappointment-minimizing” choice of name order is

---

7 On the other hand, Engers et al derive the contributions of the authors from endogenous effort choices on their part, and this aspect of their model underlies their results. In particular, they show the the alphabetic system will be the equilibrium, despite the greater efficiency of a system that rewards the author whose contribution is greater with first-authorship.

8 For example, if authors are always listed alphabetically, which is a limiting possibility for Engers et al., the credit assigned will be equal.
invariably made.\textsuperscript{9} These ex-post considerations can then create inefficiencies ex-ante, as authors weigh the various payoffs conditional on name-order. Below, we address this last efficiency question.

But all that said, fairness is good enough for us. We see no reason why people further down the alphabet should be treated badly!

4. Conventions as Equilibria, and Institutional Intervention

Then there is the question of implementing a regime change. After all, social conventions are typically equilibria that are immune to deviations by individuals or even by sub-segments of the population. Suppose, for instance, that Archimedes and Zeno, working together,\textsuperscript{10} gallantly agree to break the convention by 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, “Zeno and Archimedes” would be a statement that Zeno has done most of the research for that paper, whereas “Archimedes and Zeno” 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 Zeno gains a lot when his does. Zeno will agree to the \textit{ex-ante} randomization, but Archimedes will not. That is one reason it is hard to “invade” an alphabetical society with a mutant scheme.

But institutions can help. Here is a simple variant of the randomization scheme — a mutant — which will set it apart from pure randomization. We propose that any randomized name order be presented with the symbol $\mathcal{R}$ immediately following it. We propose, moreover, that any written citation of such a paper be provided with the $\mathcal{R}$-symbol; e.g., Ray and Robson$^{\mathcal{R}}$ (2016) is the appropriate reference for this paper. And we propose that an august body such as the \textit{American Economic Association} make an announcement to this effect. \textit{We only ask that 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.

\textsuperscript{9}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.

\textsuperscript{10}This perhaps stretches realism a bit, given that Zeno was a bit more than 200 years older than Archimedes, but we’re not off by much more than the average assumption in a theory paper.
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 “fairness.” But fairness 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 institutionalized random order as just described, but not private randomization. In short, in the presence of institutionalized random order, alphabetical order is “unstable.”

Now for the details.

5. 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 Zeno. Their contributions are $x$ and $y$ respectively, where $x + y = 1$. Ex ante, $x$ is uniformly distributed on $[0, 1]$. Ex post, $(x, y)$ is observed by the authors but not by the public.

Let $n$ be a name order: e.g., alphabetical order ($n = \alpha$), reverse-alphabetical order ($n = \rho$), or random order ($n = r$). 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 accrues over and above the “direct credit” that the public will estimate. There are also payoffs stemming from fairness that we develop in detail below.

5.1. Social Conventions. The use of a particular name order sends signals to the public about the contributions $(x, y)$, but these signals 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 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 $\epsilon \in (0, 1/2)$. \(^{12}\)

Let $a(n, C)$ and $b(n, C)$ denote the credits to Archimedes and Zeno respectively when convention $C$ is followed and the name-ordering $n$ is observed. (These are not overall payoffs, which will need to include both the $\delta$ and fairness concerns.) For the modified alphabetical convention $E$ used in Economics with threshold $\epsilon$, the use of $n = \alpha$ yields a credit of

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

\(^{11}\)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.

\(^{12}\)We adopt the simplification that $\epsilon$ is the same for all author pairs. This can be relaxed in the analysis below by positing players with heterogeneous standards of fairness.
to Archimedes and Zeno respectively, while, if the reverse order $\rho$ is observed, the corresponding credits are

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

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

5.2. **Fairness.** A final ingredient 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 $U(z)$: the “unfairness function.” It has its domain the difference between the true credit due to an author, and the inferred credit that he obtains from the announced name order and the going convention.

For example, consider the convention $E$ described above. Suppose that $x > \epsilon$ is the true contribution of $A$, whereas the inferred contribution under $E$ is $a(\alpha, E) = 1 + \frac{\epsilon}{2}$. Then, from $Z$’s perspective, $z = x - 1 + \frac{\epsilon}{2}$, and the unfairness experienced by $Z$ is $U(x - 1 + \frac{\epsilon}{2})$.

Focus for instance on the first author Archimedes. Given the realization $(x, y)$, Archimedes’s overall utility from a choice of name order $n$ is assumed to be the difference of two components:

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

(ii) the unfairness $U$ generated by $n$, as described above.

We impose the following restrictions on $U$. First, we presume that $U(z) = 0$ when $z \leq 0$; that is, Archimedes isn’t interested in altruistically giving Zeno more credit than he deserves. Second, $U$ is continuous everywhere, and increasing and strictly convex for $z \geq 0$. Finally, we impose two end-point conditions that 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:

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

The first end-point condition can be interpreted as follows. Consider a situation in which Zeno has done all the work, and Archimedes none, so that $(x, y) = (0, 1)$. By insisting on alphabetical order, Archimedes “steals” approximately 1/2 a unit of credit from Zeno. In the extreme case that $\epsilon$ is zero, this generates a sense of unfairness (in Archimedes’s mind) equal to $U(1/2)$. The first inequality in (5.3) 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 an entirely 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. In such a system, assume that if Archimedes and Zeno have contributed exactly equally, Archimedes would prefer his name to go first, rather than reverse names. Then it follows that $U(1/4) \leq 1/2$.  


5.3. Payoffs, Deviations, Equilibrium. Behavior under a convention can be formalized using game theory, though the game 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. Specifically:

In Stage 1, contributions are revealed. In Stage 2, both Archimedes and Zeno 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.

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 (5.1) and (5.2), 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 $\mathbb{R}$-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 $\mathbb{R}$ 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).

We are proposing then a particular 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 $E$ could be supported as an equilibrium, as could a wide variety of other outcomes. This particular refinement is intuitively attractive.

We make the following simplifying assumption—that the fraction of the population that are $\mathbb{R}$ mutants is “small”. The useful implication of this is that the inferred contributions after observing $\alpha = AZ$ or $\rho = ZA$ are not affected by the presence of the mutant.

A convention, then, is assessed within the context of which actions are allowable. Given that allowable set, (a) the specified actions of the convention must constitute a subgame perfect equilibrium (henceforth, SPE) of the contextual game, and

---

13 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 $s$. 
(b) there must be no unused action from the allowable set which can serve as an equilibrium deviation.

6. The Economics Convention $E$ as an SPE

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 (5.1), Archimedes’s utility from $\alpha$ is

$$a(\alpha, E) + \delta - U(y - b(\alpha, E)) = \frac{1 + \epsilon}{2} + \delta - U\left(y - \frac{1 - \epsilon}{2}\right),$$

while if $\rho$ is implemented, then using (5.2), his utility is

$$a(\rho, E) - U(y - b(\rho, E)) = \frac{\epsilon}{2} - U\left(y - \left[1 - \frac{\epsilon}{2}\right]\right).$$

6.1. Baseline. In the baseline setting without $\bar{R}$, 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 Zeno contributes a lot) and then use a fixed point argument to make sure it coincides with society’s anticipated threshold. To this end, using (6.1) and (6.2), and substituting in the values for credits, observe that Archimedes will reverse when

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

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 Zeno. 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 (6.3), we see that that $\epsilon$ must solve

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

Condition (5.3), and the assumption that $U$ is increasing, guarantees that there is a unique value of $\epsilon \in (0, 1/2)$ that solves (6.4).

Notice how the strict convexity of $U$ and the end-point conditions (5.3) are necessary to get what we see in the data. For instance, if $U$ is linear, then Archimedes simply trades off units of his credit for units of Zeno’s, and if he places a higher weight on own credits, he will never reverse. If he places a higher weight on Zeno’s credits, he will always reverse. Presumably, we see neither, which suggests that the “marginal unfairness” climbs as Zeno’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. Zeno might want to reverse for even lower values of his contribution $y$, but Archimedes will have none of it: his sense of unfairness is not piqued enough for that to happen. It should also be noted that Zeno 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 Zeno receives the overall payoff $\delta + [1 - (\epsilon/2)] - U(\epsilon/2)$ on reversal. In the status quo, he gets $(1 - \epsilon)/2$. Consequently, Zeno will always be happy with reversal provided that $\delta + [1 - (\epsilon/2)] - U(\epsilon/2) \geq (1 - \epsilon)/2$, which yields

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

Comparing this requirement with (6.4) and noting that $\epsilon < 1/2$, we see that (6.5) will always be satisfied.

What other options do Archimedes and Zeno 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 the active support of $\mathbb{R}$.

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 easily seen to be the unique equilibrium. We have therefore established:

**Theorem 6.1.** In the baseline setting, there is a unique value of $\epsilon \in (0, 1/2)$, given by the solution to (6.4), for which the economics convention $E$ with threshold $\epsilon$ is an SPE.

### 6.2. Disrupting $E$ With Random Order $\mathbb{R}$

Now we introduce the possibility of random order (with $\mathbb{R}$, where not explicitly mentioned). This must be an option that is institutionally provided, say by a consortium of the leading journals, so that the meaning of the $\mathbb{R}$ 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 (5.1) and (5.2), where $\epsilon$ is pinned down by (6.3). Should Archimedes and Zeno 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 unfairness 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.

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 Zeno both agree to adopt random order.

**Theorem 6.2.** The convention $E$ is disrupted by an equilibrium deviation to random order.

**Proof.** The Appendix provides a complete proof of this key result. This proof is moderately involved, but it can be outlined as follows. Fix the convention $E$ and its associated reversal threshold $\epsilon$, given by (6.4). 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 $x^\alpha(a)$, such that Archimedes prefers random order to alphabetical order whenever his realized credit falls below $x^\alpha(a)$, and another, called $x^\rho(a)$, 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 $a^*$ for which the conditional expected value of Archimedes’s credit over this zone exactly equals $a^*$. We then verify that Zeno prefers random order in this zone over alphabetical order.

In the zone $(x^*_1, x^*_2)$, it can then be shown that $\mathbb{R}$ is the unique SPE outcome. This is because $\mathbb{R}$ is Archimedes favorite outcome in this range and Archimedes can ensure $\mathbb{R}$ as the SPE outcome. Hence a small group of $\mathbb{R}$ mutants can certainly invade, in particular, as an “equilibrium deviation”, so completing the proof of the Theorem.

**7. Analysis of the Random Order Convention**

We turn now to the analysis of the random order convention, one that uses $\mathbb{R}$. This is a hypothetical convention, for now at least, but we can envisage desiderata for it that run parallel to the economics convention. First, $\mathbb{R}$ now becomes the default choice for either author, so that both authors would need to be happy with any change. For instance, $\mathbb{R}$ can be overridden by Archimedes in favor of $\rho = ZA$, if Zeno agrees, but Zeno cannot insist on this. A similar possibility arises with the roles reversed. As for alphabetical order, $\alpha = AZ$, that too is available as a choice with no default status, so that, in the context of the new convention, neither author can unilaterally insist on it.\textsuperscript{15}

We use the abbreviation $\mathbb{R}$ for the default option where the two players randomize 50-50 between $\alpha\mathbb{R}$ and $\rho\mathbb{R}$. We follow the rules that we described in Section 5.3, where the set of available actions is now $\{d, \alpha, \rho\}$, and where the default is now $d = \mathbb{R}$.

More precisely, in Stage 1, contributions are revealed. In Stage 2, both Archimedes and Zeno independently choose from $\{d, \alpha, \rho\}$. In Stage 3, if the same actions are chosen, this is implemented. If both parties take different non-default actions, then

\textsuperscript{15}The game we propose gives a more detailed and precise expression to this intuitive description.
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.

7.1. Subgame Perfect Equilibrium. Consider a pattern of outcomes of the fol-
lowing form. There exists a $\epsilon \in (0, 1/2)$, such that—(i) if $x < \epsilon$ then the outcome
is $\rho$, that is, $ZA$, (ii) if $x > 1 - \epsilon$ then $\alpha$, that is, $AZ$, is the outcome, and, (iii)
if $x \in [\epsilon, 1 - \epsilon]$ then $\mathcal{R}$ is the outcome. We describe this succinctly as a “random
order convention with threshold $\epsilon$”.

The key result here is then—

**Theorem 7.1.** There is a unique value of $\epsilon^* \in (0, 1/3)$ for which the random-order
convention $\mathcal{R}$ with threshold $\epsilon^*$ is supported as an SPE outcome.

**Proof.** The Appendix provides a detailed proof.

However, an outline of the proof is as follows. First it is shown that there exists
an $\epsilon^*$ such that $A$ prefers $\rho$ to $\mathcal{R}$ if $x < \epsilon^*$ but prefers $\mathcal{R}$ to $\rho$ if $x > \epsilon^*$. Since $\rho$
can be shown to be $Z$’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 $Z$’s favorite outcome, so $Z$ can exploit $d$’s
status as the default to guarantee $d$—that is, $d$ is the only possible SPE outcome. Analogously,
$A$ must obtain his favorite outcome $d$, in any SPE, in the range $x \in [1/2, 1 - \epsilon^*].$

7.2. Efficiency. In this section, we show the total expected payoff under the ran-
don order convention $\mathcal{R}$ strictly exceeds that under the convention $E$. Since the
convention $\mathcal{R}$ also entails equal payoffs for both agents, it follows that any sym-
metric quasiconcave welfare criterion would strictly prefer $\mathcal{R}$ to $E$.

This is not surprising since the number of different signals under $\mathcal{R}$ is three; whereas
the number of different signals under $E$ is only two. However, it is possible, on the
face of it, for the three ranges under $\mathcal{R}$ 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. First, 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 $\epsilon^* = 1/2$.
For simplicity, we adopt here the notation $\epsilon^*$ for the equilibrium value of $\epsilon$
under the convention $E$, as in Eq (6.4). However, Eq (5.3) implies that $\epsilon^* < 1/2$.

Secondly, the convention $\mathcal{R}$ is also inefficient in the class of three signal mechanisms.
Again, efficiency requires that the ranges for the three signals be equal. It we adopt
the notation $\epsilon$ to refer to the solution of Eq (11.12), then efficiency requires that
$\epsilon = 1/3$. However, Eq (11.14) shows that $\epsilon \leq 1/3$.

---

16Eq (6.4) becomes $LHS = U \left(1+\frac{1-\epsilon^*}{2}\right) = \frac{1}{2} + \delta = RHS$. If, however, $\epsilon^* = 1/2$, then Eq (5.3)
implies that $LHS = U(1/4) \leq 1/2 + \delta/2 < 1/2 + \delta = RHS$, so that it must be that $\epsilon^* < 1/2$. 

12
Under either mechanism, there is an inadequate incentive for an agent to concede first authorship, leading to inefficiency.

There is a formal mechanism that attains full efficiency, as follows. Suppose that, for each \( x < 1/2 \), the notation \( px \) is published; whereas, if \( x \geq 1/2 \), \( \alpha x \) is published. Neither agent ever then experiences any unfairness, so that total expected payoffs are \( 1 + \delta \), which is the upper bound. There is no deviation from this that is Pareto-improving, of course. But this mechanism pushes the assumption that the agents have mutual knowledge of \( x \) too hard. 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.

It is an important consideration that the convention \( R \) is fairer than is the convention \( E \) since it yields equal payoffs for the two agents, whereas \( E \) favors \( A \). A welfare criterion that was sufficiently averse to inequality, sufficiently quasiconcave, that is, might then prefer \( R \) to \( E \). However, the case for \( R \) does not rest on that. That is, the sum of the two agents’ expected utilities is higher under \( R \) than under \( E \), so that Bentham would strictly prefer the former. Indeed, any symmetric quasiconcave welfare criterion would strictly prefer \( R \) to \( E \).\(^{17}\) The remainder of this subsection is devoted to proving this claim.

Consider first the total expected payoff under \( E \). There are four relevant ranges for \( x \). Recall that we denote the equilibrium value of \( \epsilon \) for \( E \), as in Eq (6.4), as \( \epsilon' \).

If \( x \in [0, \epsilon'/2] \), then \( A \)'s payoff is \( \epsilon'/2 - U(\epsilon'/2 - x) \); whereas \( Z \)'s is \( 1 - \epsilon'/2 + \delta \).

If \( x \in [\epsilon'/2, \epsilon] \), \( A \)'s payoff is \( \epsilon'/2 \); whereas \( Z \)'s payoff is \( 1 - \epsilon'/2 + \delta - U(x - \epsilon'/2) \).

If \( x \in [\epsilon, (1 + \epsilon)/2] \), \( A \)'s payoff is \( (1 + \epsilon)/2 + \delta - U((1 + \epsilon)/2 - x) \); whereas \( Z \)'s payoff is \( (1 - \epsilon)/2 \).

If \( x \in [(1 + \epsilon)/2, 1] \), \( A \)'s payoff is \( (1 + \epsilon)/2 + \delta \); whereas \( Z \)'s payoff is \( (1 - \epsilon)/2 - U(x - (1 + \epsilon)/2) \).

Hence total expected payoff for both agents is \( W' = 1 + \delta - \int_{\epsilon'/2}^{\epsilon} U(\epsilon'/2 - x)dx - \int_{\epsilon'/2}^{\epsilon} U(\epsilon'/2 - x)dx - \int_{(1 + \epsilon)/2}^{(1 + \epsilon)/2} U((1 + \epsilon)/2 - x)dx - \int_{(1 + \epsilon)/2}^{(1 + \epsilon)/2} U((1 + \epsilon)/2 - x)dx \).

Using suitable changes in variable, this simplifies to \( W' = 1 + \delta - 2 \int_{0}^{\epsilon'/2} U(x)dx - 2 \int_{0}^{(1 - \epsilon)/2} U(x)dx \).

For the \( R \) equilibrium, there are three signals, with each signal range divided into two halves. Again, overall expected utility depends on an integral of the unfairness function. Again, the six ranges can be collapsed into just two types. That is, an analogous argument applies to obtain the total expected payoff under \( R \) as \( W = 1 + \delta - 4 \int_{0}^{\epsilon'/2} U(x)dx - 4 \int_{0}^{1/2 - \epsilon} U(x)dx \), with \( \epsilon \) now taken as the solution to Eq (11.12) for the \( R \) convention.

**Theorem 7.2.** The \( R \) convention is more efficient than the \( E \) convention, so that \( W > W' \).

\(^{17}\)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 \( x \).
**Proof.** See the Appendix. This proof matches up as much as possible the integrals involved in $W$ and $W'$. Essentially because $W$ entails three different signals, but $W$ entails only two, the residual integrals that cannot be matched are then shown to involve higher values of $U$ for $W'$ than for $W$.

8. Transitions

We've shown so far that the alphabetical order convention is not immune to an invasion by (institutionalized) 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 6.2, in which the economics convention is successfully invaded by a mutant, hopefully to be created by an august body such as the American Economic Association. 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 newly born mutant, but also for pre-existing choices such as $α$ and $ρ$. For instance, we showed that the initial invasion will occur in the interval $(x_1^*, x_2^*)$, so, after the mutant becomes significant in that range, the Bayes update of $x$ following an observation of $α$ 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 by no means does it deliver us all the way to a random order convention. For that, the default itself will need to evolve.

That brings us to the second source of dynamics, which entails a satisfactory model of changing defaults. Several choices are available here. 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, $α$ will lose salience as a default, and gain in importance as a powerful signal of the relative contribution of Archimedes, just as $ρ$ remains a powerful signal for Zeno’s relative contribution. The newly-acquired, substantially asymmetric meaning of $α$ must eventually eliminate it as a default outcome; that would be as absurd (or paradoxical?) as Zeno insisting on $ρ$ as a default. The random order convention would now reign supreme.

9. An Alternative Mechanism

A simpler alternative to our mechanism would be to keep published papers with the authors names’ listed alphabetically, but to randomize 50-50 each time a citation is made. We believe that our $®$ mechanism has advantages over this scheme.

---

We thank Leeat Yariv for this suggestion.
In the first place, it would be hard to ensure that all users of the paper diligently randomize. Further, it seems there would be strong social pressure to refer to the classic production function paper, for example, only as Cobb-Douglas, or settle instead only on Douglas-Cobb, avoiding, in any case, referring to it sometimes as Cobb-Douglas and sometimes as Douglas-Cobb.

Finally, this alternative mechanism does not provide for the counterfactual possibility that Cobb-Douglas might have been Douglas-Cobb, if Douglas had done the lion’s share of the work (despite the burden of being a US Senator). If that had happened, Douglas would presumably have objected to 50-50 randomization of first authorship in citations. The present system, with the possibility of name reversal, generates two signals with different interpretations—Cobb-Douglas and Douglas-Cobb. On the other hand, randomized citations does not exploit name reversal in the same fashion, since it imputes the same meaning to the two signals. Although randomizing citations is clearly fairer than the current system, then, it might well then be less efficient.

As discussed above, the three signals available with our convention®, each with a different interpretation, helps ensure it is more efficient than the current system, as well as fairer.

10. Conclusions

In this paper we advocate a new system for assigning credit to papers with two coauthors. We first characterize the present system as modified alphabetical, where the author who is earlier in the alphabet can offer first authorship to other, if the contributions are very unequal. This is motivated by a sense of fairness on the part of the earlier author.

Our 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 fairness when the contributions have been extremely unequal.

For the sake of expositional clarity, we have only studied two-author papers. But nothing prevents us from extending the same idea to three or more authors. The extended notion of randomization is obvious, at least when all authors wish to randomize. Moreover, there would be departures from this system only when all authors agree.

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 $\mathcal{R}$ is that it does not require any more of the agents than does the present convention $E$, despite being fairer and more efficient. The convention $\mathcal{R}$ simply allows either player to concede first authorship, instead of allowing only $A$ to have this option, as in the convention $E$. Although such an option can lead to arguments, it is exercised on occasion in reality, so the transactions cost of this we are willing to pay.\footnote{Flipping a coin and adding $\mathcal{R}$ 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.

In summary, 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 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


### 11. Appendix: Proofs

**Proof of Theorem 6.2**

Recall we have fixed the convention $E$ and its associated reversal threshold $\epsilon$, given by (6.4), and are supposing that society assigns a credit pair $(a,b) = (a,1-a)$ to the observation of random order. We now have—

**Lemma 11.1.** 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,

(11.1) \[ x^\alpha(\epsilon/2) > \epsilon, \]

while away from this end-point,

(11.2) \[ x^\alpha(a) > a \text{ for all } a \in [\epsilon/2, \bar{a}) \text{ and } x^\alpha(\bar{a}) = \bar{a}. \]
Proof. Random order at realization \((x,y)\) yields an expected payoff to Archimedes of
\[
a + \frac{\delta}{2} - U(y - b) = a + \frac{\delta}{2} - U(y - [1 - a])
\]
while alphabetical order generates a payoff of
\[
\frac{1 + \epsilon}{2} + \delta - U \left( y - \frac{1 - \epsilon}{2} \right)
\]
as described in (6.1). Therefore random order is weakly preferred to \(\alpha\) if
\[
U \left( y - \frac{1 - \epsilon}{2} \right) - U \left( y - [1 - a] \right) + a \geq \frac{1 + \epsilon}{2} + \frac{\delta}{2}.
\]
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.\(^{20}\) Let \(x^{\alpha}(a) \equiv 1 - y^{\alpha}(a)\). (If (11.4) fails for all \(y\), set \(x^{\alpha}(a) = 0\).) Because \(U(z)\) is strictly convex when \(z > 0\), the LHS of (11.4) 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 (11.1), set \(a = \epsilon / 2\) and \(x = \epsilon\), so that \(y = 1 - \epsilon\). Then
\[
U \left( y - \frac{1 - \epsilon}{2} \right) - U \left( y - [1 - a] \right) + a = U \left( \frac{1 - \epsilon}{2} \right) - U \left( -\frac{\epsilon}{2} \right) + \frac{\epsilon}{2}
\]
\[
= \frac{1}{2} + \frac{\delta}{2} + \frac{\epsilon}{2}
\]
\[
\geq \frac{1 + \epsilon}{2} + \frac{\delta}{2},
\]
where the second inequality employs the definition of \(\epsilon\) in (6.4).\(^{21}\) That shows that (11.4) holds strictly when \(x = \epsilon\). Because the left-hand side of (11.4) is decreasing in \(x\), (11.1) is true.

Notice moreover that \(x^{\alpha}(a)\) moves continuously in \(a\) as long as it is strictly positive. At the same time, (11.4) 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 (11.2) and completes the proof. \(\square\)

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 \([\epsilon / 2, \bar{a}]\).

**Lemma 11.2.** There exists 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,
\[
x^{\rho}(\epsilon / 2) = 0, \text{ and } x^{\rho}(a) < a \text{ for all } a \in [\epsilon / 2, \bar{a}].
\]
\(^{20}\)Recall that \(U(z) = 0\) for all \(z \leq 0\), and continuous everywhere. This, combined with \(a \leq (1 + \epsilon) / 2\), guarantees that (11.4) must fail for \(y\) small enough.
\(^{21}\)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 Zeno but gives Archimedes the extra payoff of \(\delta\) half the time.
Proof. Reverse order \( \rho \) yields a payoff to Archimedes given by (6.2), which is
\[
\frac{\epsilon}{2} - U \left( y - \left[ 1 - \frac{\epsilon}{2} \right] \right).
\]
Combining with (11.3), we see that random order is weakly preferred to \( \rho \) if
\[
U (y - [1 - a]) - U \left( y - \left[ 1 - \frac{\epsilon}{2} \right] \right) - a \leq \frac{\delta}{2} - \frac{\epsilon}{2}.
\]
Again, because \( U(z) \) is strictly convex when \( z > 0 \), the LHS of (11.6) is strictly increasing in \( y \) (and so decreasing in \( x \)) thereby showing that (11.6) holds (if it holds at all) over some interval of the form \([x_\rho(a), 1]\). \( x_\rho(a) \) is either zero (if (11.6) always holds), or is the equality solution \( y = 1 - x_\rho(a) \) if (11.6) holds for some \( y \). (The remaining possibility, that (11.6) never holds, is ruled out by a parallel argument to that in Footnote 20, this time using \( a \geq \epsilon/2 \).)

Finally, we establish (11.5). Note that when \( a = \epsilon/2 \), the same credits are associated to random order as with reversal. So unfairness 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 \( a = \epsilon/2 \), (11.6) holds for all values of \((x, y)\), so \( x_\rho(\epsilon/2) = 0 \).

To establish the second part of (11.5), suppose that at \((a, b)\), the realized contributions are also \((a, b)\). Setting \( y = b = 1 - a \), we see that
\[
U (y - [1 - a]) - U \left( y - \left[ 1 - \frac{\epsilon}{2} \right] \right) - a = U (0) - U \left( \left[ \frac{\epsilon}{2} - a \right] \right) - a \leq -\frac{\epsilon}{2} < \frac{\delta}{2} - \frac{\epsilon}{2},
\]
which shows that (11.6) is satisfied with strict inequality when \( x = a \).\(^{22}\) Therefore \( x_\rho(a) < a \).

\[\square\]

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 11.1 and 11.2 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 (11.1) that \( x_\alpha(\epsilon/2) > \epsilon \), and we know from (11.5) that \( x_\rho(\epsilon/2) = 0 \), so it follows that
\[
\phi \left( \frac{\epsilon}{2} \right) > \frac{\epsilon}{2}.
\]
We also know from (11.2) that \( x_\alpha(\bar{a}) = \bar{a} \), so that
\[
\phi(\bar{a}) \leq \bar{a}.
\]
Combining (11.7) and (11.8) and invoking the continuity of \( \phi \), we must conclude that there exists \( a^* \in [\epsilon/2, \bar{a}] \) such that
\[
\phi(a^*) = a^*.
\]
\(^{22}\)The weak inequality in the chain uses the fact that \( a \geq \epsilon/2 \) and the assumption that \( U(z) = 0 \) when \( z \leq 0 \).
Define \( x_1^* = x^\rho(a^*) \) and \( x_2^* = x^\alpha(a^*) \). 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 \( \Diamond \) to both \( \alpha \) and \( \rho \) in this zone.

**Lemma 11.3.** Zeno also prefers \( \Diamond \) to \( \alpha \) in the region \( (x_1^*, x_2^*) \).\(^{23}\)

**Proof.** Note that Zeno receives the payoff \((1 - \epsilon)/2\) under \( \alpha \),\(^{24}\) while under random order his lowest possible payoff is \( b^* + (\delta/2) - U(x_2^* - a^*) = (1 - a^*) + (\delta/2) - U(x_2^* - a^*) \) (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

\[
\frac{\delta}{2} + \left( \frac{1 + \epsilon}{2} - a^* \right) \geq U(x_2^* - a^*).
\]

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

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

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

\[
\frac{\delta}{2} + \left( \frac{1 + \epsilon}{2} - a^* \right) = U\left(y^* - \frac{1 - \epsilon}{2}\right) = U\left(\frac{1 + \epsilon}{2} - x_2^*\right).
\]

To establish (11.9), 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 (11.10), it means that it is also equivalent to show that

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

Because \( U \) is convex, it will suffice to show that (11.11) 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 \( 1/2 \), (11.10) reduces to the requirement that

\[
\frac{\delta}{2} + \frac{1}{2} \geq U(1/4),
\]

which is guaranteed by the second end-point condition in (5.3). \( \square \)

We now complete the proof of the theorem by showing that in the zone \( (x_1^*, x_2^*) \), \( \Diamond \) is the unique SPE outcome. We show that Archimedes can guarantee \( \Diamond \) by announcing the default \( \alpha \). For Zeno, \( \alpha \) cannot be a best response; by announcing \( \Diamond \) (and having Archimedes subsequently agree — which he will do), Zeno can implement \( \Diamond \), which he prefers (Lemma 11.3). The only other option is for Zeno to announce \( \rho \), in which case Archimedes can reject and counterpropose \( \Diamond \). By Lemma 11.3 again, Zeno will accept this counterproposal. So Archimedes can guarantee \( \Diamond \). Because \( \Diamond \) is Archimedes’s best option in the zone \( (x_1^*, x_2^*) \), it must therefore be the unique equilibrium.

\(^{23}\) Zeno might continue to prefer \( \rho \) over \( \Diamond \).

\(^{24}\) There is no unfairness loss, because \( x \leq (1 + \epsilon)/2 \) for every \( x \in [x_1^*, x_2^*] \).
We have therefore established that there is a disruption of $E$ by an “equilibrium deviation” to random order, in which Archimedes and Zeno both willingly participate, completing the proof of Theorem 6.2.

**Proof of Theorem 7.1.**

Hypothesize then that a “random order convention with threshold $\epsilon$” has arisen in an SPE. Including fairness, if contributions are $(x, 1-x)$, Archimedes’s total payoff from $\mathbb{R}$ at $x$ is $1/2 + \delta/2 - U(1 - x - 1/2) = 1/2 + \delta/2 - U(1/2 - x)$. Similarly, Archimedes’s overall payoff from $\rho$ is $\epsilon/2 - U(1 - x - (1 - \epsilon/2)) = \epsilon/2 - U(\epsilon/2 - x)$, which simply equals $\epsilon/2$ if $x \geq \epsilon/2$. With these payoffs in mind, consider the function

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

By the end-point conditions in (5.3), $\Delta(0) = 1/2 + \delta/2 - U(1/2) < -\delta/2 < 0$ and $\Delta(1/2) = 1/2 + \delta/2 - U(0) - 1/4 > 0$. Moreover, $\Delta$ is concave because $U$ is convex. It follows that there exists a unique $\epsilon^* \in (0, 1/2)$ such that $\Delta(\epsilon^*) = 0$. That is,

$$
(11.12) \quad \frac{1}{2} + \frac{\delta}{2} - \epsilon^*/2 = U\left(\frac{1}{2} - \epsilon^*\right),
$$

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

Consider first the range $x < \epsilon^*$. In this range, we now show that Archimedes strictly prefers $\rho = ZA$ to $\mathbb{R}$. Zeno strictly prefers $\rho = ZA$ to either $\mathbb{R}$ or $\alpha = AZ$.

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

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

$$
(11.13) \quad \frac{1}{2} + \frac{\delta}{2} - \epsilon^*/2 > U(\epsilon^*/2).
$$

We have $\epsilon^*$ solves $\Delta(\epsilon^*) = 0$, as in Eq (11.12), implying that Eq (11.13) follows as long as $\epsilon^* < 1/3$. Given the properties of the function $\Delta$ just established, it suffices to show that $\Delta(1/3) > 0$. But

$$
(11.14) \quad \Delta(1/3) = \frac{1}{3} + \frac{\delta}{3} - U(1/6)
$$

and this is indeed positive, given the convexity of $U$ and the second end-point condition in (5.3).\footnote{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 $f(z) \equiv 2z + (\delta/2) - U(z)$ is concave, with $f(0)$ and $f(1/4)$ both nonnegative, the latter by the second end-point condition in Eq (5.3). Setting $\lambda = 2/3$, it follows that $f(1/6) = (1/3) + (\delta/3) - U(1/6) \geq 0$. Therefore $\Delta(1/3) > f(1/6) \geq 0.$} In short, $\epsilon^* \in (0, 1/3)$, so that Eq (11.13) applies, and Zeno strictly prefers $\rho$ to $\mathbb{R}$ whenever $x < \epsilon^*$.
Finally, for the case $x < \epsilon^*$, we show that Zeno strictly prefers $\rho = ZA$ to $\alpha = AZ$.

In this range, $\rho$ yields Zeno a payoff of $\delta + 1 - \epsilon^*/2 - U(x - \epsilon^*/2)$, whereas $\alpha$ yields just $\epsilon^*/2$. The desired result then follows if $\delta + 1 - \epsilon^* > U(x - \epsilon^*/2)$. Using Eq 11.12, this, in turn, follows since, for $x < \epsilon^*$, $2U(\epsilon^*/2) > U(\epsilon^*) > U(x - \epsilon^*/2)$ so that $\delta + 1 - \epsilon > U(x - \epsilon^*/2)$.

We now apply these results concerning the two players’ preferences to verify that $\rho$ is an SPE outcome in the range $x < \epsilon^*$. To see this, suppose that Zeno chooses $d$ at Stage 2, given that it is revealed that $x < \epsilon$. That is, $d$ is Zeno’s favorite outcome. This observation is sufficient for the present purpose. To show this, note that, in this range, $\rho$ yields Zeno a payoff of $\delta/2 + 1/2 + 1/2 - \epsilon^*/2 = U(\epsilon^*/2)$, whereas $\alpha$ yields $\delta + 1 - \epsilon^*/2 - U(x - \epsilon^*/2)$. Hence Zeno strictly prefers $\alpha$ to $\rho$ simply because $\epsilon^* < 1$.

To show that Zeno weakly prefers $\rho$ to $\rho$ it is enough to show that $U(x - \epsilon^*/2) \geq \delta/2 + 1/2 - \epsilon^*/2 = U(\epsilon^*/2)$, using Eq (11.12). But this follows since $x \geq \epsilon^*$.

It follows that Zeno can now 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 then completes the proof of Theorem 7.1.

**Proof of Theorem 7.2.**

We start by noting that $\epsilon > \epsilon'/2$. Indeed, recall from Eq (5.3) that $\epsilon$ satisfies $U(1/2 - \epsilon) + \epsilon/2 = 1/2 + \delta/2$. Suppose now that $\epsilon' = 2\epsilon$. It follows that $U((1 - \epsilon')/2) = U(1/2 - \epsilon) = 1/2 + \delta/2 - \epsilon/2 < 1/2 + \delta$. Since $\epsilon'$ should satisfy $U((1 - \epsilon')/2) = 1/2 + \delta$, as in Eq (6.4), it follows that $\epsilon' < 2\epsilon$.

To prove Theorem 7.2 it suffices to show that $LHS = 2 \int_0^{\epsilon'/2} U(x)dx + \int_0^{1/2 - \epsilon} U(x)dx < \int_0^{\epsilon'/2} U(x)dx + 2 \int_0^{1/2 - \epsilon'} U(x)dx = RHS$ For clarity, we consider two cases. Suppose first that $\epsilon < \epsilon'$. Now $RHS - LHS = \int_{\epsilon'/2}^{\epsilon'/2} U(x)dx + \int_{1/2 - \epsilon}^{1/2 - \epsilon'} U(x)dx - \int_0^{\epsilon'/2} U(x)dx$, using $\epsilon > \epsilon'/2$.

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 the smallest value of $U(x)$ in the second integral is $U(1/2 - \epsilon)$ which is not less than the largest value of $U(x)$ from the negative integral, $U(\epsilon/2)$, since $\epsilon \leq 1/3$, from Eq (11.14). Hence $RHS - LHS > 0$ and $W > W'$. 

22
Now consider the case that $\epsilon \geq \epsilon'$. Now

$$RHS - LHS = -\int_{\epsilon'/2}^{\epsilon/2} U(x) dx + \int_{1/2-\epsilon}^{1/2-\epsilon'/2} U(x) dx - \int_{0}^{\epsilon/2} U(x) dx,$$

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 $U$ in the positive integral, $U(1/2 - \epsilon)$, is no less than the largest value of $U$ in either negative integral, $U(\epsilon/2)$, because $\epsilon \leq 1/3$, from Eq (11.14). It follows that $RHS > LHS$ and hence $W > W'$, completing the proof of Theorem 7.2.