MARK J. MACHINA

BARRETT AND ARNTZENIUS'S INFINITE DECISION PUZZLE

ABSTRACT. The Barrett and Arntzenius (1999) decision paradox involves unbounded wealth, the relationship between period-wise and sequence-wise dominance, and an infinite-period split-minute setting. A version of their paradox involving bounded (in fact, constant) wealth decisions is presented, along with a version involving no decisions at all. The common source of paradox in Barrett–Arntzenius and these other examples is the indeterminacy of their infinite-period split-minute setting.

KEY WORDS: Decision theory, Paradoxes of infinity

Barrett and Arntzenius (1999) pose an infinite-period repeated-choice problem, in which one option dominates another in each *individual* period, yet choosing the latter option in *every* period seems to dominate choosing the former in every period. It can be phrased as follows:

Say a bank has a countably infinite stack of dollar bills, with serial numbers #1, #2, #3, ... In each period $n = 1, 2, 3, \ldots$, it offers you the following choice:

A. Receive the next dollar bill off the top of stack, in addition to any you may already have

B. Receive the next 2^{n+1} bills off the top of the stack, after which you must destroy exactly one of your dollar bills, namely the one with the lowest serial number

Here, the 'periods' 1, 2, 3, ... are not equal-length time intervals, but rather, have successive lengths of $\frac{1}{2}$-minute, $\frac{1}{4}$-minute, $\frac{1}{8}$-minute ..., so that they are all contained within a single one-minute time interval.

In any individual period, choosing B obviously dominates choosing A. However, Barrett and Arntzenius observe that 'after one minute', someone who had always chosen A would now own each bill #1, #2, #3, ... and hence have *infinite* wealth, whereas someone

Theory and Decision **49:** 293–297, 2000.

who had always chosen B would have destroyed each bill #1, #2, #3, ... , and hence have *zero* wealth.

Powers of two and unbounded levels of wealth have been part of decision theory ever since 1728, when Cramer offered the puzzle that spurred Daniel Bernoulli to found our field. But while these features play a key role in the St. Petersburg Paradox, they are *not* required for the Barrett–Arntzenius puzzle, and to some extent, obscure what is happening there. The purpose of this note is to offer a simpler version of the puzzle, which retains its essence but better highlights the nature (and perhaps familiarity) of the paradox it invokes.

In this simpler version, wealth is not unbounded – in fact, it stays the same in every period:

Say the bank has a countably infinite stack of dollar bills with serial numbers #1, #2, #3, ... , and initially gives you bills #1 through #100. In each period $n = 1, 2, 3, \ldots$, it offers you the following choice:

A′. Return all bills currently held, then receive (or receive back) bills #1 through #100

B′. Return all bills currently held, then receive the new bills #100n + 1 through #100n + 100

Again, the periods 1, 2, 3, ... have successive lengths of $\frac{1}{2}$-minute, $\frac{1}{4}$-minute, $\frac{1}{8}$-minute, ... so that they are all contained within a one-minute interval.

Here, wealth will be $100 in each period, no matter which option is chosen, or has been previously chosen. After one minute, an agent who chose A′ in each period will own bills #1 through #100, and hence end up with $100 in wealth. But according to Barrett and Arntzenius's argument, if the agent had chosen B′ in each period, then each bill #i would have at some point been permanently returned to the bank, so as before, the agent would end up with zero wealth.

The above version of the Barrett–Arntzenius puzzle shows that while it is indeed a paradox of infinity, it is not inherently a paradox of *unbounded wealth*. For a more familiar example of the essentially the same paradox, consider the following two infinite hotels, both members of the world famous *Hotel Cantor* group (the division of the minute into periods is the same as before):

Hotel A always has the same 100 guests, who stay in Rooms #1– #100 each period

Hotel B always has the same 100 guests, who stay in Rooms #1–#100 for the first period, in Rooms #101–#200 for the second period, in Rooms #201–#300 for the third period ...

Once period 1 starts, no one ever checks into or out of either hotel. During each period, Hotel A has 100 guests and so does Hotel B. After one minute, Hotel A still has guests in its first 100 rooms, while each room in Hotel B is empty. Where did Hotel B's guests go?

To establish the formal connection between the classic Hotel Paradox and the above dollar-bill puzzle, assume that, for security, the bank retains physical possession of all the dollar bills, laying them out in order along an infinite tray, and keeps its records by placing a special removable sticker on each bill that is currently owned by the agent. An agent who always chose A' would see their 100 stickers remain on bills #1–#100 the whole time. An agent who always chose B' would always see their 100 stickers on *some* 100 bills, but the bank shifts them over by 100 bills each period. At the end of the minute, no individual bill can be said to have a sticker on it. Where did the stickers go? In the formal connection with Barrett and Arntzenius's original puzzle, an agent who always chose B would receive *additional* stickers each period, and only the sticker on the lowest-numbered bill will actually be shifted during a given period. But the stickers will still migrate out along the tray, and no individual bill will have a sticker at the end of the minute.

The Hotel Paradox is not a paradox because it has *logically incontrovertible but absurd implications*, but because it has *indeterminate implications*, which (as usual) allow for contradictory claims: For Hotel B, we get one final occupancy level if we take the initial registration of 100 and subtract the zero checkouts, but a different final occupancy level if we count up the rooms that are occupied at the end of the minute.

The same is true for the Barrett and Arntzenius puzzle: 'Always choosing B' is *not* uncontrovertibly dominated by 'always choosing A'. Rather, the key to the puzzle is that 'always choosing B' has *indeterminate implications*, which allow for claims in both directions. For example, while it is certainly correct to say that 'each dollar bill was destroyed', it is just as correct to also say that 'each dollar bill was *more than replaced* when destroyed'.

For a more formal counter-implication, say the bank maintains *public records* of the bills as they get destroyed, by posting their serial numbers on a lined blackboard (infinite, of course). When the first bill is destroyed, its serial number is posted on the top line of the blackboard. When the next bill is destroyed, its number gets posted on the top line, and the original serial number is moved down to line two. As each bill is destroyed, its serial number is posted at the top, and all existing serial numbers moved down a line. Obviously, after one minute, it would be impossible for any line on the blackboard to contain serial number #*i*. We have now proven that at the end of the minute: (1) *every* numbered bill has been destroyed, and (2) *no* numbered bill has been destroyed!

The infinite split-minute setting used by Barrett and Arntzenius, which allows them to speak of reaching (even surpassing) infinity, is a classic source of indeterminate implications. The simplest example of this is the 'Thomson Lamp' (Thomson 1954, Gardner 1971):

> Let the periods 1, 2, 3 ... be successive time lengths of $\frac{1}{2}$-minute, $\frac{1}{4}$-minute, $\frac{1}{8}$-minute ... , so they are all contained within a single one-minute time interval. An agent with a push-button lamp turns the lamp *on* at the start of period 1, *off* at the start of period 2, *on* at the start of period 3, etc. After one minute, is the lamp on or off ?

If you still think the Barrett and Arntzenius puzzle is revealing a problem in decision theory rather than simply the *ex post* indeterminacy of their infinite split-minute setting, think about this lamp example for a minute – and *then* decide if you find it illuminating.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 9870894. Pulier (2000, this issue) provides an additional criticism of Barrett and Arntzenius, based on their use of nonconvergent sums.

REFERENCES

Barrett, J. and Arntzenius, F. (1999). An Infinite Decision Puzzle, *Theory and Decision* 46(1): 101–103.

Bernoulli, D. (1738), Specimen Theoriae Novae de Mensura Sortis, *Commentarii Academiae Scientiarum Imperialis Petropolitanae* (Papers of the Imperial Academy of Sciences in Petersburg) V, 175–192. English translation: Exposition of a New Theory on the Measurement of Risk, *Econometrica* 22 (1954): 23–36.

Cramer, G. (1728), Letter to Nicholas Bernoulli. Included in Bernoulli (1738/1954).

Gardner, M. (1971), The Orders of Infinity, The Topological Nature of Dimension and 'Supertasks', *Scientific American* 224: 106–109.

Pulier, M.L. (2000), A Flawed Infinite Decision Puzzle, *Theory and Decision* 49: 291–292.

Thomson, J. (1954). Tasks and Super-Tasks, *Analysis* 15 (1): 1–13. Reprinted in W. Salmon (ed.), *Zeno's Paradoxes*. Indianapolis: Bobbs-Merrill Co. 1970.

Address for correspondence: Dr Mark J. Machina, University of California, San Diego, Department of Economics, 9500 Gilman Drive, La Jolla CA 92093, USA. Phone: (858) 534-2391; Fax: (858) 534-7040; E-mail: mmachina@ucsd.edu