JEFFREY A. BARRETT and FRANK ARNTZENIUS

AN INFINITE DECISION PUZZLE

ABSTRACT. We tell a story where an agent who chooses in such a way as to make the greatest possible profit on each of an infinite series of transactions ends up worse off than an agent who chooses in such a way as to make the least possible profit on each transaction. That is, contrary to what one might suppose, it is not necessarily rational always to choose the option that yields the greatest possible profit on each transaction.

KEY WORDS: Decision theory, Dutch book, Puzzles

One might expect to do better acting rationally than acting irrationally. It is possible, however, to tell a story where one is guaranteed to do better than a seemingly rational agent by always acting in what appears to be an entirely irrational way.[1]

Suppose one has an infinite stack of dollar bills with consecutive serial numbers: 1, 2, 3, etc. An agent, who starts with no money, is then offered the following choice, where n is equal to the total number of times that the choice has been offered so far:

1. Get one dollar bill off the top of the stack.
2. Get 2^{n+1} dollar bills off the top of the stack, but you must then return the bill with the smallest serial number that you currently have. The returned bill is immediately destroyed.

Suppose that the agent does not know how many times he will be offered this choice.

The first time the agent is offered the choice ($n = 1$) he accepts option 2 because the expected return of option 1 is \$1 while the expected return of option 2 is \$3. The second time ($n = 2$) the agent also accepts option 2 because the expected return of option 1 is \$1 and the expected return of option 2 is \$7. Indeed, since, for all positive n, $2^{n+1} - 1 > 1$, the agent chooses option 2 every time he is offered the choice. So his profit for each choice is \$3, \$7, \$15, etc. He takes all of his choices to be rational–after all, he always acts to maximize his profit at each step.

Theory and Decision **46**: 101–103, 1999.

Now suppose the agent is offered the choice at 1/2 minute, at 3/4 minute, at 7/8 minute, at 15/16 minute, etc. How much money will he have after one minute? There is a straightforward argument that by always choosing option 2 the agent ends up with nothing.[2]

For any bill k that the agent receives there is a time $1 - 1/2^k$ before one minute when he must return it. So after one minute the agent must have returned every bill that he received, and he is left with nothing. On the other hand, an agent who always chooses option 1, which would presumably be considered irrational at every step, would make an infinite profit. One might worry that irrationality pays so well.[3]

There is nothing inherently exotic about any of the choices that the agent is offered in the story when considered by itself. One can easily imagine being offered a choice between, say, option 1 ($1) and option 2 ($4 but return $1). One would naturally choose option 2, and, if one thought about it at all, one would most likely justify this choice by pointing out that it yields the greatest profit. The above story, however, illustrates that this fact does not by itself justify making the choice. In the story, the agent who makes the greatest possible profit on each choice ends up worse off in the long run than the agent who makes the least possible profit on each choice. Where does the first agent go wrong? Not knowing ahead of time what choices he will be offered, what strategy should the agent follow at each step?[4]

ACKNOWLEDGMENTS

We would like to thank Terry Parsons, Brian Skyrms, Wayne Aitken, and John Norton for valuable discussions of this puzzle (and others). We would also like to thank the two referees for their comments.

NOTES

1. See Earman and Norton (1996) for a discussion of the consistency and physical possibility of such supertasks and a review of the literature. In our story, each dollar bill can be taken to follow a perfectly ordinary trajectory in classical spacetime.

2. One can tell a similar story where the agent would with probability one end up with nothing if he were required to return a bill at random. See Ross (1988: 68–70).
3. One might construct an infinite Dutch-book story along similar lines. Suppose the agent must pay $1 of his own money to play. If he had the money, then he would certainly accept at each step since he would still always make a profit each time he chose option 2. Again, option 2 dominates option 1 at each step, but here, the agent would ultimately take an infinite loss by paying $1 at each step for the opportunity of choosing option 2. Unlike other infinite Dutch-book stories, the result here is certain. That a Dutch-book can be made against the agent who always chooses option 1 tells us that his choices are irrational. But what would a rational agent do in his shoes?
4. An obvious reaction to this story would be to deny that any agent will ever find himself in a situation where he must make an infinite number of decisions. But this may be too fast (so to speak). We presumably want our decision theory to be as strong as possible, which means that we presumably want it to work even for agents who inhabit worlds where one can make an infinite number of decisions in a finite time (indeed, perhaps something akin to this could happen in our world – see Hogarth (1992) and Earman and Norton (1996) for recent discussions of such possibilities). Further, if our decision theory cannot handle infinite sequences of decisions, then we undermine the standard practice of using infinite models in our analysis and justification of finite decisions (in the use of infinite Dutch-book stories to justify the axioms of probability theory for example – see Earman (1992: 33–44) for a short discussion of such arguments).

REFERENCES

Earman, J. (1992), *Bayes or Bust*. Cambridge, MA: MIT.

Earman, J. and Norton, J. (1996), Infinite pains: the trouble with supertasks in A. Morton and S. Stich (eds.), *Benacerraf and His Critics*. Cambridge: Blackwell.

Hogarth, M. (1992), Does general relativity allow an observer to view an eternity in a finite time? *Foundations of Physics Letters* 5(2): 173–181.

Ross, S. (1988), *A First Course on Probability*, 3rd edn. New York: Macmillan.

Address for correspondence: Jeffrey A. Barrett, Department of Philosophy, University of California, Irvine, CA 92717-4555, USA.

Frank Arntzenius, Department of Philosophy, University of Southern California