

JEFFREY A. BARRETT and FRANK ARNTZENIUS

WHY THE INFINITE DECISION PUZZLE IS PUZZLING

ABSTRACT. Pulier (2000, *Theory and Decision* 49: 291) and Machina (2000, *Theory and Decision* 49: 293) seek to dissolve the Barrett–Arntzenius infinite decision puzzle (1999, *Theory and Decision* 46: 101). The proposed dissolutions, however, are based on misunderstandings concerning how the puzzle works and the nature of supertasks more generally. We will describe the puzzle in a simplified form, address the recent misunderstandings, and describe possible morals for decision theory.

KEY WORDS: Dutch book, Supertask, Puzzle

In two recent papers Pulier (2000) and Machina (2000) sought to dissolve the Barrett–Arntzenius infinite decision puzzle (1999). Pulier argues that the infinite decision puzzle relies on the assumption that a particular infinite sum is well defined when it is not. And Machina argues that supertasks like the one described by Barrett and Arntzenius do not lead to well defined states. Both of these arguments, however, are based on misunderstandings of how the puzzle works. We will first describe the infinite decision puzzle in a simplified form. Then we will address the misunderstandings. Finally, we will describe possible morals for decision theory.

II

The bank has an infinite stack of dollar bills with serial numbers from the top of the stack down: #1, #2, #3, ... The bank offers an agent a choice between two options:

- A. Receive the top three bills from the bank's stack, then return to the bank that bill from the set of bills the agent currently holds that has the least serial number. Once the agent returns a bill to



Theory and Decision 52: 139–147, 2002.

© 2002 Kluwer Academic Publishers. Printed in the Netherlands

the bank, the bank keeps it, and the agent will never be given that particular bill again.

Or

B. Receive the top bill from the bank's stack.

Option A nets the agent \$2 each time he chooses it, and option B nets \$1, so the agent should presumably always choose option A. But if the bank offers the agent this choice at $t = 1/2$, $t = 3/4$, $t = 7/8$, $t = 15/16$, ..., etc. (where t is the time in minutes, say, from the start of the game), an agent who always opts for A would have no money after one minute since for every serial number k , there is a time $t = (2^k - 1)/2^k$ before $t = 1$ when the agent would have to return the bill with serial number k to the bank. On the other hand, the agent who always opts for B would have all of the bank's money after one minute. Thus, the agent who acts in a stepwise rational way at every step ends up worse off than the agent who acts in a stepwise irrational way at every step. And this is the infinite decision puzzle.

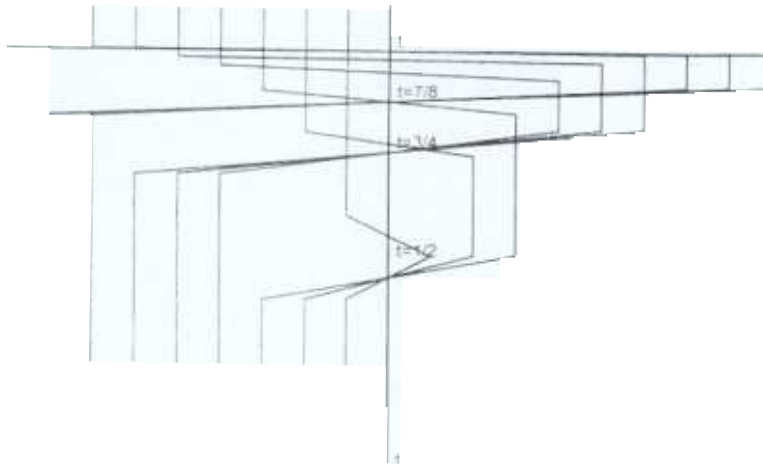
III

The infinite puzzle has nothing to do with the *sum* of the wealth of the agents; rather, it concerns the *specific bills* that they possess. This is an important distinction because the sum of the wealth of the agent who always opts for A

$$(\$3 - \$1) + (\$3 - \$1) + (\$3 - \$1) +$$

fails to be well defined (since one can get any final wealth one wants by suitably ordering the terms in the sum). But which bill the agent has at each time is perfectly well defined. That is, the infinite decision puzzle concerns keeping track of each bill in an infinite stack of bills, and this poses no special problems. One can keep track of each bill before, during, and after the minute that contains all of the agent's transactions with the bank. Indeed, if one wishes, one can stipulate a perfectly determinate and continuous physical trajectory for each bill at all times.

If the agent always opts for A, the space-time trajectories of the first nine bills might look something like Figure 1. Note that each bill in the figure follows a well-defined trajectory at all times and



BANK

AGENT

ends up in the bank after 1 min. Further, it does not take much imagination to see how one could similarly provide a perfectly determinate, continuous trajectory for every bill in the bank's stack. The point here is just that one can tell a perfectly coherent physical story where there is a determinate, well-defined physical state at all times for an agent who always opts for A, and on this story, the bank has all of the bills after 1 min. One can, of course, also tell a physical story that is no less determinate where an agent who always opts for B ends up with all of the bills after 1 min.

There is nothing mysterious about the fact that the infinite sum of wealth is undefined for an agent who always opts for A while the physical trajectories of the individual bills are perfectly well defined at all times. The physical trajectories provide a more faithful representation of the events than the infinite sum; they contain mathematical structure that the sum lacks. Since the bill trajectories are well defined when the infinite sum is not, the physical trajectories of the bills can determine the final wealth of an agent in situations where the infinite sum fails to determine the final wealth of the agent.

There is another way to think about this infinite-sum business. One might at first think that one could that an agent who always opts for A would have an infinite amount of money after 1 min.

by taking an infinite sum. For instance one might argue that for all finite n , after n stages one has $2n$ dollars, and that hence, it follows that after 1 min., when there have been an infinite number of stages, one will have an infinite number of dollars. However, this argument presupposes continuity of the wealth function at $t = 1$, and it is not continuous. Alternatively one might try writing the agent's wealth as the infinite sum above, then claim that this sum is infinite and that the agent will consequently have an infinite amount of money after 1 min. But this argument also fails since, as Pulier points out, such sums fail to be well defined. One might make such a sum well defined by stipulating a particular ordering of the operations (by taking the limit of a particular sequence of partial sums, for example), but then, at best, one would be back at the previous bad argument where one must assume that the wealth function is continuous at $t = 1$. But, of course, none of this means that the agent does not have a perfectly well-defined wealth after 1 min.; rather, what it means is that the standard ways of calculating the agent's wealth can fail to work if the wealth function is discontinuous. We know that the wealth function is discontinuous here because we know from the physical model that the agent will return every bill to the bank before one minute. The physical model that shows this is counterintuitive, but it is also perfectly coherent.

It is the existence of coherent model of the transactions, not a well-defined infinite sum of wealth, that the infinite decision puzzle requires for its coherence. Since the puzzle itself has nothing whatsoever to do with infinite sums of wealth, Pulier's (2000) argument that the puzzle is flawed since it relies on a mathematical mistake is itself mistaken. Indeed, Pulier seems to misunderstand the most important feature of the puzzle: what makes the infinite decision puzzle puzzling is that one's intuitions concerning the stepwise accumulation of wealth and how this is faithfully represented by the arithmetic of wealth are radically mistaken when one considers the possibility of an infinite number of transactions involving an infinite supply of individuated goods.

v

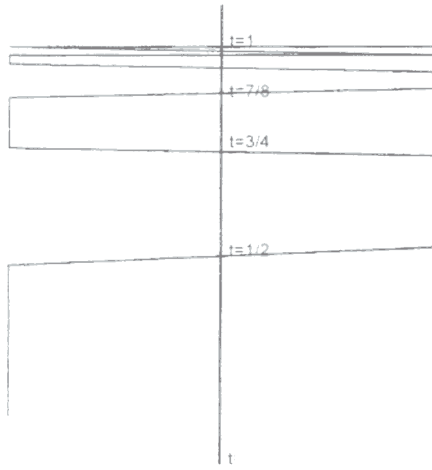
Machina's (2000) complaint is related but somewhat different. He seems to have understood that the infinite puzzle requires the determinateness of physical states rather than the determinateness of infinite sums of wealth. But he argues that the infinite decision puzzle involves indeterminate sequences of physical states. More specifically, Machina argues that the final state is not determinate. The main argument is an argument by analogy.

A supertask is a task that requires an infinite sequence of steps to complete but is completed in a finite time. Machina considers two supertasks that fail to yield determinate final states, then argues that the supertask involved in the infinite decision puzzle when the agent always opts for A similarly fails to yield a determinate final state. The problem with this argument is that while some supertasks do indeed fail to yield determinate final states, others yield perfectly determinate final states, and the supertask that involves the agent who always opts for A is one of this the latter sort: it is a supertask that yields a perfectly determinate final state.

One supertask that yields an indeterminate final state is the famous Thomson Lamp Puzzle. On this story, the Thomson lamp gets turned on and off and on faster and faster (off at $1/2$ min., on at $3/4$ min., off and $7/8$ min., etc.), then one asks what the state of the lamp is at the end of one minute. Machina notes that there can be no determinate final state for the lamp, then concludes that there can thus be no determinate final state in the infinite decision puzzle when an agent always opts for A.

In order to see clearly the disanalogy between this and the infinite decision puzzle, one might consider what we will call the Thomson Bill Puzzle. On this story there is just one dollar bill that gets passed back and forth between the bank and the agent. The bank gives the bill to the agent at $t = 1/2$ the agent returns the bill to the bank at $t = 3/4$, the bank gives the bill back to the agent at $t = 7/8$, etc. Who then has possession of the bill after one minute? (Figure 2).

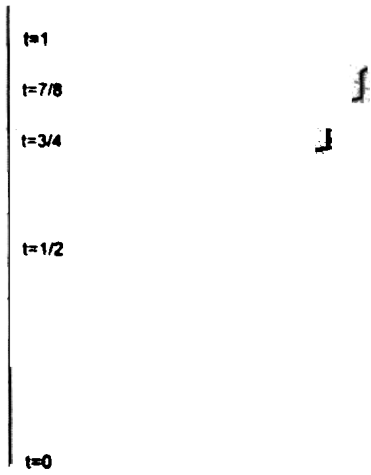
The Thomson Bill Puzzle *is* appropriately dissolved by noting that there is no well defined physical state at 1 min. or beyond. While one can easily provide a perfectly determinate and continuous physical trajectory for the bill for times before 1 min., there can be no continuous trajectory that satisfies the constraints of the story



BANK

AGENT

Figure 2.



BANK

AGENT

Figure 3.

and continues *beyond* 1 min. But while Machina is certainly right to conclude that *some* supertasks fail to yield determinate final states, one should not conclude that *all* supertasks fail to yield final states.

Consider what one might call the Achilles Bill Puzzle. Here there is again a single dollar bill, but this time the bank decides to deliver it to the agent in steps. The bank pushes the bill half of the distance to the agent at $t = 1/2$ half of the remaining distance at $t = 3/4$, half the remaining distance at $t = 7/8$, etc. Does the agent get the bill? (Figure 3).

Here, unlike the Thomson bill, one can provide the Achilles bill with a perfectly determinate, continuous trajectory for all times (before, during, and *after* 1 min. that satisfies the constraints of the story. And it turns out, of course, that the agent has possession of the bill from 1 min. on.

Some supertasks then do yield determinate final states. But it is not this that entails that an agent always opting for A in the infinite decision puzzle yields a determinate final state. Rather, it is that each of the bills in the infinite decision puzzle, just as with the Achilles bill, can be given a perfectly determinate, continuous trajectory for all times that satisfies the constraints of the story. The upshot is that while Machina is right to believe that some supertasks fail to yield determinate final states, this fact is ultimately irrelevant to the coherence of the infinite decision puzzle.¹

Given that the infinite decision puzzle emerges from recent criticisms unscathed, what does this mean for decision theory? One might get some insight into what morals to draw by considering the conditions for telling the puzzling story in the first place.

As Machina correctly argues, the infinite decision puzzle does not require that there be infinite potential wealth. More specifically, one can certainly tell a similarly puzzling story with finitely bounded total utility. Suppose, for example, that while each additional bill has positive marginal utility for the agent, each additional bill has less marginal utility so that having all of the bills has finite total utility. There is still a puzzle for our notion of stepwise rational action here.

But there is a deeper point concerning the role played by wealth. As we have seen, the infinite puzzle has nothing whatsoever to do with wealth in the abstract sense; rather, it has to do with the transaction of identifiable goods. What the puzzle apparently requires

is that there be a potentially infinite supply of such goods and a potentially infinite number of possible transactions. If this is right, then one can avoid the puzzle by only requiring that one's account of rational decision cover situations where there are only finite number of identifiable goods or situations where an agent knows that he will never be presented with an infinite sequence of decisions.

More generally, telling a story like ours requires one to assume things about the structure of the physical world and the nature of agents. If there were an upper bound to the total energy of any physical system, then one would still be able to tell an infinite decision story that works, but it must be considerably more subtle than the story told above (one might, for example, stipulate that each bill the bank gives the agent is left closer to the boundary between the bank and the agent so that less energy is required to return each bill and the total energy to return all of the bills remains finite). But if there were also some positive lower bound to the energy required for an agent to deliberate at each step, then it very well might be impossible to tell a consistent story that supports the puzzle.

One could then avoid the infinite decision problem by stipulating that our world is such that supertasks of the sort needed to support the infinite decision puzzle are in fact impossible. Or, somewhat weaker, one might stipulate that while super tasks of this sort might be in principle possible, our world is such that agents will in fact never face such a decision-making puzzle. The problem, of course, is that it remains an interesting question as to whether supertasks of the sort needed to support the infinite decision puzzle can occur in our world and whether there is any sense in which an agent might in fact be faced with such a puzzle. The larger moral is that whether or not a particular account of decision deserves to be called an account of rational decision, whether it will in fact lead to successful action in the long run, ultimately depends on specific details concerning the nature of agents and the physical world they inhabit.

One way to think about supertask puzzles in the context of the study of our physical theories is that they provide situations where local conservation laws are satisfied but the corresponding global conservation laws fail. Our puzzle extends this by providing a situation where local rationality holds but global rationality fails. The very possibility of there being a coherent account of both local and

global rationality may require one to assume that the relevant sort of supertask situations can never in fact occur in our world.

NOTES

For the record, the Thomson Lamp does not yield a determinate final state nor does Machina's story where one writes the natural numbers on the blackboard and keeps moving the list down so that the largest number is always at the top. But Machina also tells a story where a guest at the Hilbert Hotel is moved from room to room in numerical order faster and faster and concludes that this supertask fails to yield a well-defined final state. Curiously, a version of this supertask can be given that yields a perfectly determinate final physical state. John Earman and John Norton (1996) discuss such situations. For a brief introduction to the vast literature concerning supertasks and their physical possibility see Jon Perez Laraudogoitia's survey article (2000).

REFERENCES

- Barrett J. A. and tzenius, F. (1999). An infinite decision puzzle. *Theory and Decision* 46(3): 101–103.
- Earman, J. and Norton, J.D. (1996). Infinite pains: The trouble with supertasks, in A. Morton and S. Stich (eds.), *Benacerraf and His Critics*, pp. 231–261. Oxford: Blackwell.
- Laraudogoitia, J.P. (2000). Spacetime: Supertasks. in E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/spacetime-supertasks>
- Machina, M. J. (2000). Barrett and Arntzenius's, infinite decision puzzle. *Theory and Decision* 49(3): 293–297.
- Pulier, M. L. (2000). A flawed infinite decision puzzle. *Theory and Decision* 49(3): 291–292.

Address for correspondence: Professor J.A. Barrett, Department of Logic and Philosophy of Science, University of California-Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA.