This is a repository copy of *A Simpler, More Compelling Money Pump with Foresight*.

Article:

A Simpler, More Compelling
Money Pump with Foresight*†

Johan E. Gustafsson and Wlodek Rabinowicz

ABSTRACT: One might think that money pumps directed at agents
with cyclic preferences can be avoided by foresight. This view was chal-
lenged two decades ago by the discovery of a money pump with fore-
sight, which works against agents who use backward induction. But
backward induction implausibly assumes that the agent would act ra-
tionally and retain her trust in her future rationality even at choice
nodes that could only be reached if she were to act irrationally. This
worry, however, does not apply to BI-terminating decision problems,
where at each choice node backward induction prescribes a move that
terminates further action. For BI-terminating decision problems, it is
enough to assume that rationality and trust in rationality are retained
at choice nodes reachable by rational moves. The old money pump
with foresight was not BI-terminating. In this paper, we present a new
money pump with foresight—one that is both BI-terminating and con-
siderably simpler.

Suppose you prefer *A* to *B*, *B* to *C*, and *C* to *A*. These preferences over
options *A*, *B*, and *C* are cyclic. Are such preferences rationally permissi-
ble? The standard argument against the rational permissibility of cyclic
preferences of this kind is based on an exploitative money-pump set-up,
where you start off with *A* and then, following your preferences, you are
led to accept a series of trades such that you pay to end up with *A*. That is,
you end up paying for the same option you had at the beginning, which
you could have kept for free. This seems irrational.

Yet it may be objected that you can avoid the exploitation if you have
foresight and see in advance where the trades would lead. An agent who
knows that she is being taken for a ride can avoid exploitation if she
uses backward induction. Or so it might seem. This common view was

challenged two decades ago by the discovery of a money pump with foresight, which works against agents who use backward induction in their deliberation.

But is backward induction a satisfactory deliberation method? A standard objection to backward induction is that it implausibly assumes that the agent would act rationally and retain her trust in her future rationality even at choice nodes that could only be reached if she were to act irrationally. This objection, however, does not apply to the use of backward induction in BI-terminating choice problems, that is, choice problems where backward induction at each choice node prescribes a move that terminates further action. The assumptions needed to defend backward-induction solutions in BI-terminating choice problems are weaker and more plausible than in other cases. One only needs to assume that rationality and trust in rationality are retained at choice nodes reachable by rational moves. The old money pump with foresight was not BI-terminating, however. In this paper, we present a new money pump with foresight—one that is both BI-terminating and considerably simpler. Hence we have an exploitation scheme against agents with cyclic preferences which only assumes a very plausible form of backward-induction reasoning.

<p style="text-align:center">∗ ∗ ∗</p>

As mentioned, we suppose that you prefer A to B, B to C, and C to A. Letting '$X > Y$' denote that option X is strictly preferred to option Y, we can state your preference cycle as follows: $A > B > C > A$. In addition, suppose that you would still prefer A to B even if you had to pay a small amount for A. Let A^- be A with a small payment. Hence you prefer A^- to B. Needless to say, you prefer A to A^-. And, just as you prefer C to A, you also prefer C to A^-. Accordingly, you have the following preferences: $C > A^- > B > C > A > A^-$. These preferences are assumed to be stable, so that, for example, if you were to exchange A for C, you would still prefer B to C. Your preferences are retained when you make your choices.

Now, consider the standard version of the money pump:[1]

[1] Amos Tversky, "Intransitivity of Preferences," *Psychological Review*, LXXVI, 2 (January 1969): 31–48: at p. 45; based on an example in Donald Davidson, J. C. C. McKinsey, and Patrick Suppes, "Outlines of a Formal Theory of Value, I," *Philosophy of Science*, XXII, 2 (April 1955): 140–60, pp. 145–46. This version is a slight simplification of the original set-up in that we only consider one trade payment. The other payments are superfluous.

The Standard Money Pump
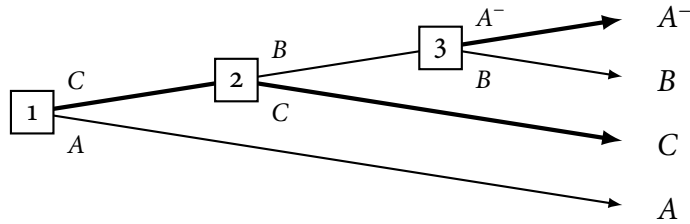
$$C > A^- > B > C > A > A^-.$$

In this case, you start off with A, then an exploiter will offer you three consecutive trades: first, a trade from A to C; second, if you accept the first trade, a trade from C to B; and, third, if you also accept the second trade, a trade from B to A for a small payment (that is, a trade from B to A^-). In the above diagram, the three choice nodes for these potential trades are represented by numbered squares. At these nodes, you have a choice whether to accept the trade (by going up) or to turn it down (by going down). The two options noted to the right of each choice node correspond to the options you have in your possession after accepting the trade (the upper option) or turning it down (the lower option).

Since you prefer C to A, it seems that you should accept the first trade, from A to C. Next, since you prefer B to C, it seems that you should accept the second trade, from C to B. And, finally, since you prefer A^- to B, it seems that you should accept the third trade, from B to A^-. But, if you accept all of these trades, you end up with A^- (that is, you pay for A), even though you could have kept A for free by turning down the first trade offer.

At this point, it may be objected that you should have had some foresight and seen in advance where the trades would lead.[2] To be at all plausible, the money-pump argument must rely on a sequence of trade offers which is known to the agent in advance. The exploiter must not possess any information that the agent lacks, because being exploited by someone who possesses more information is not a sign of irrationality. But, if you know the potential trades in advance, you can avoid being exploited in the Standard Money Pump by being prudent. (In what follows, we refer to the combination of foreknowledge with prudence as "foresight.") More precisely, you can avoid exploitation in the Standard Money Pump if you

---

[2] Frederic Schick, "Dutch Bookies and Money Pumps," this JOURNAL LXXXIII, 2 (February 1986): 112–19, at pp. 117–18.

use *backward induction*, that is, if you predict what would be chosen at later choice nodes and take these predictions into account in your decisions at earlier choice nodes.[3]

To use backward induction in the Standard Money Pump, consider first node 3. Since you prefer A^- to B, you would choose A^- at that node. Next, consider node 2. Taking into account your prediction that A^- would be chosen at node 3, the choice at node 2 is effectively between A^- and C. And, since you prefer C to A^-, you would choose C at node 2. Finally, consider node 1. Taking into account your prediction that C would be chosen at node 2, the choice at node 1 is effectively between A and C. Since you prefer C to A, you choose C at node 1. (The choices you would make if you follow backward induction are represented by the thicker lines in the diagram.) Hence, if you use backward induction, you trade only once and end up with C. Thus you avoid being money pumped: you do not pay money to get what you had from the beginning and could have kept for free.[4]

Nevertheless, there is now a well-known exploitation scheme that works against agents with cyclic preferences who use backward induction. Consider the following money pump with foresight:[5]

---

[3] Brian Skyrms, "A Mistake in Dynamic Coherence Arguments?," LX, 2 (June 1993): 320–28 and Robert W. Rosenthal, "Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox," *Journal of Economic Theory*, XXV, 1 (August 1981): 92–100, at pp. 94–95.

[4] Edward F. McClennen, *Rationality and Dynamic Choice: Foundational Explorations* (New York: Cambridge, 1990), at pp. 163–67 and Wlodek Rabinowicz "To Have One's Cake and Eat It, Too: Sequential Choice and Expected-Utility Violations," this JOURNAL, XCII, 11 (November 1995): 586–620, at pp. 592–93.

[5] Wlodek Rabinowicz "Money Pump with Foresight," in Michael J. Almeida ed., *Imperceptible Harms and Benefits* (Dordrecht: Kluwer, 2000), pp. 123–54, at p. 141. Just like the Standard Money Pump, this version is a slight simplification of the original set-up in that we only consider one payment.
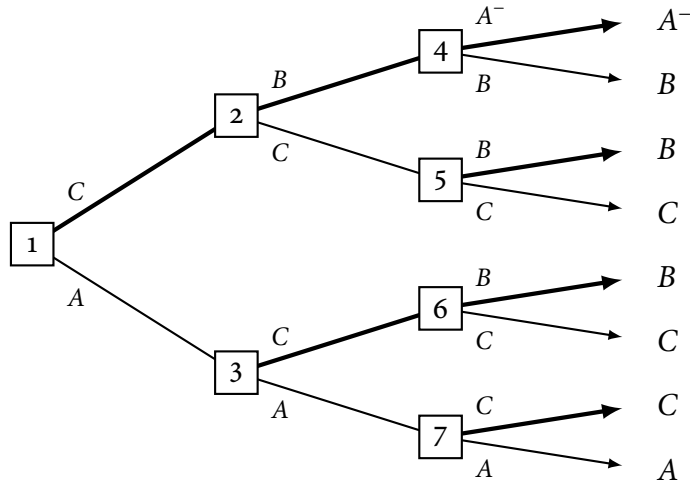
$$A^- > B > C > A > A^-.$$

In this case, the exploiter is assumed to be persistent: If you refuse a trade offer, he will repeat the same offer at the next stage. If you accept an offer, you will be given a new offer at the next stage. The offers are the same as in the Standard Money Pump: a trade from A to C, a trade from C to B, and a trade from B to A^-. There are three stages at which offers are going to be made.

How would you choose in the Money Pump with Repeated Offers if you use backward induction? First, consider the final choice nodes, that is, nodes 4, 5, 6, and 7. At each of these nodes, you would accept the trade, since it offers what you prefer. So you would go up in each of the final nodes. Next, consider node 2. Taking into account your prediction that you would trade B for A^- at node 4 and that you would trade C for B at node 5, the choice at node 2 is effectively between A^- and B. Since you prefer A^- to B, you would choose to go up at node 2 and thus trade C for B. Now, consider node 3. Taking into account your prediction that you would trade C for B at node 6 and that you would trade A for C at node 7, the choice at node 3 is effectively between B and C. Since you prefer B to C, you would choose to go up at node 3 and thus trade A for C. Finally, consider node 1. Taking into account your prediction that you would trade C for B at node 2 and then B for A^- at node 4 and that you would trade A for C at node 3 and then C for B at node 6, the choice at node 1 is effectively between A^- and B. Since you prefer A^- to B, you choose to go up at node 1 and thus trade C for A. So you make all three

trades and end up with A^-, even though you could have kept A for free. Hence you get exploited even though you use backward induction.

Nevertheless, one might question whether backward induction is a satisfactory decision method. The standard objection to backward induction is that it requires the agent to retain trust in the future rationality of the players (including her own rationality) at choice nodes that can only be reached by irrational play.[6]

This raises a worry about the use of backward induction in the Money Pump with Repeated Offers. Consider node 3. Since this node cannot be reached by a rational agent if backward induction codifies rational behaviour, it seems that you would have reason to distrust your own rationality if you were to find yourself at this node. If you at that node were to doubt your rationality, it seems that you might plausibly also doubt that you would choose B at node 6 or C at node 7, as rationality requires. But then the backward-induction reasoning unravels. You might, at choice node 3, doubt whether you should go up, since you might doubt whether it would give you a better final outcome than if you were to go down at that node.

Indeed, it is not only questionable whether an agent would retain trust in her rationality at nodes that she can only reach by irrational moves. It might also be questioned whether she or we, who analyse her decision problem, can expect that she would act rationally at such nodes. After all, irrational actions might corrupt the agent; they might become a bad habit.[7]

These worries, however, can be ignored in BI-terminating decision problems. A decision problem is *BI-terminating* if and only if, at each choice node, the move prescribed by backward induction is final: it is not followed by any opportunities for further moves.[8] The objection that backward induction requires the agent to stay rational and retain trust in her future rationality even at choice nodes that can only be reached by

---

[6] Ken Binmore "Modeling Rational Players: Part I," *Economics and Philosophy*, III, 2 (October 1987): 179–214, at pp. 196–200, Christina Bicchieri "Backward Induction without Common Knowledge," *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (1988): 329–43, at pp. 334–35, and Philip Pettit and Robert Sugden "The Backward Induction Paradox," LXXXVI, 4 (April 1989): 169–82, at pp. 171–74.

[7] John Cantwell and Tom Dougherty have both put forward money pumps that do not rely on backward induction. As we explain in the appendix, we do not find their approaches quite satisfactory.

[8] Wlodek Rabinowicz "Grappling with the Centipede: Defence of Backward Induction for BI-Terminating Games," *xiv*, 1 (April 1998): 95–126, at p. 101.

irrational choices can be sidestepped in BI-terminating decision problems. To defend the backward-induction solution for these decision problems, it is enough to assume that, at the choice nodes reachable by *rational* moves, but not necessarily at other choice nodes, the agent will remain rational and retain trust in her future rationality at the choice nodes that can be reached by rational moves.[9] We shall see how it works in a moment.

The Money Pump with Repeated Offers is *not* a BI-terminating decision problem. But the following money pump is:

The Upfront Money Pump
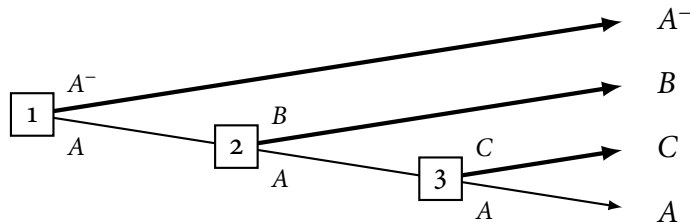
$$A^- > B > C > A > A^-.$$

In this case, you start off with A; then the exploiter offers you three consecutive trades: first, a trade from A to A^- (that is, an offer to freely give money away); second, if you turn down the first trade, a trade from A to B; and, third, if you also turn down the second trade, a trade from A to C. Note that the consecutive options in this scheme—A^-, B, C—are offered in the inverted order as compared with the Standard Money Pump, where the order is C, B, A^-. Another difference is that in the Standard Money Pump, refusals to trade are terminating moves, while in the Upfront Money Pump it is trade acceptances that terminate.

Can you avoid exploitation in this case if you use backward induction? First, consider node 3. Since you prefer C to A, you would trade A for C at node 3. Next, consider node 2. Taking into account your prediction

[9] Rabinowicz "Grappling with the Centipede," *op. cit.*, pp. 118–21. It should be noted that backward induction can be justified using such weak assumptions even in some decision problems that are not BI-terminating. For example, it may be that the move prescribed by backward induction at a certain decision node reachable by rational play is not terminating. Nevertheless, if we assume for reductio that some alternative move at that node is rational (and thus will be followed by rational moves), we can prove that the move prescribed by backward induction would be preferred no matter what moves it would be followed by. This applies, for example, to Brian Skyrms's ("A Mistake in Dynamic Coherence Arguments?," *Philosophy of Science*, LX, 2 (June 1993): 320–28, at pp. 323–24) Dutch book against violators of the principle of conditionalization.

that you would trade A for C at node 3, the choice at node 2 is effectively between B and C. Thus, since you prefer B to C, you would choose B at node 2 (that is, you would trade A for B). Finally, consider node 1. Taking into account your prediction that B would be chosen at node 2, the choice at node 1 is effectively between A^- and B. Since you prefer A^- to B, you choose A^- at node 1 (that is, you trade A for A^-). Hence, if you use backward induction, you end up paying money to retain what you had from the beginning and could have kept for free.[10]

Note that we can extend the Upfront Money Pump to exploit preferences cycles involving more than three options. Suppose that $X_1 > X_2 > \ldots > X_n > X_1$ and that the agent starts off with X_1. First, offer the agent the opportunity to pay a small amount to keep X_1 with no more trades. If the agent were to turn down the offer, offer a trade from X_1 to X_2. Then, if the agent were to turn down that offer too, offer a trade from X_1 to X_3, and so on until the agent is offered a final trade from X_1 to X_n.

The Upfront Money Pump is a BI-terminating decision problem: at each node, the move prescribed by backward induction is final. As we have noted, the assumptions needed to defend backward-induction solutions in BI-terminating decision problems are significantly weaker and more

---

[10] There is an interesting relationship between the Upfront Money Pump and the set-up described in Wlodek Rabinowicz "A Centipede for Intransitive Preferrers," *Studia Logica*, LXVII, 2 (March 2001): 167–78, at p. 170, in which an agent with cyclic preferences forgoes a sure benefit:

$$A^+ > A > B > C > A^+.$$
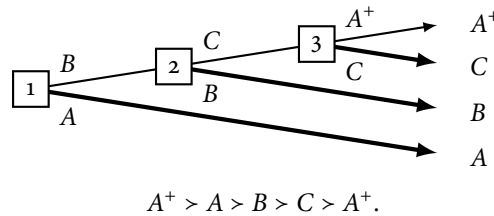
Here, A^+ is A with some extra money such that C is still preferred to A^+. In this BI-terminating set-up, the agent who uses backward induction ends up with A, forgoing a sure monetary benefit, that is, abstaining from A^+. While forgoing sure benefits may be problematic, it does not make one vulnerable to exploitation. Rabinowicz (*ibid.*, p. 174) still maintained that an exploitation scheme against agents with cyclic preferences would require a set up that is not BI-terminating. Much the same remarks are made in Wlodek Rabinowicz "Safeguards of a Disunified Mind," *Inquiry*, LIII, 3 (2014): 356–83, at p. 372. But, as the Upfront Money Pump now shows, this is not so. The Upfront Money Pump inverts the forgone benefit set-up: the benefit from refusing the last trade is replaced by the cost of accepting the first trade. Also, while in the forgone-benefit set-up it is refusals to trade that are terminating moves, in the Upfront Money Pump it is trade acceptances that terminate.

plausible than those needed to defend backward induction in general. We are now going to show how this defense works in the Upfront Money Pump.[11]
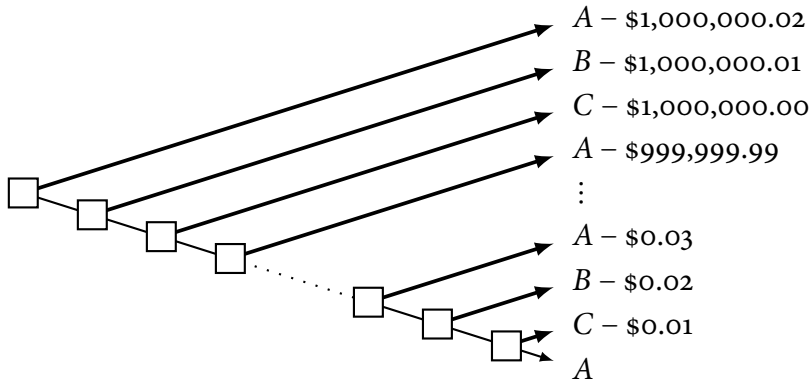
Suppose, for *reductio* that (i) it is rational to turn down the trade at node 1. Since we assume that trust in rationality at future nodes that can be reached by rational choices is retained at nodes reachable by rational choices, it follows that this trust would be retained at node 2. Suppose, again for *reductio*, that (ii) the move from node 2 to node 3 would be rational. But, if so, the agent would at node 2 expect to act rationally at node 3, that is, to trade A for C. Given this prediction, however, the move from node 2 to node 3 cannot be rational, contrary to our supposition (ii): The agent prefers B to C. The rational move at node 2 is thus to trade A for B. Consequently, given supposition (i) and our assumption that the agent continues to act rationally at nodes that are reachable by rational choices, it follows that the agent at node 2 would trade A for B. But then, contrary to supposition (i), moving to node 2 cannot be rational for the agent at node 1, since she prefers A^- to B. We thus get a contradiction. Refusing to trade at node 1 is not rational. So it must be rationally required to accept the trade at node 1, which is what is recommended by backward induction.[12]

We can extend the Upfront Money Pump to pump you for an arbitrarily large sum of money. At least we can do so if your cyclic preferences, $A > B > C > A$, are robust when it comes to small differences in money. In particular, suppose that for any two adjacent options X and Y in the preference cycle such that $X > Y$, however much these options might cost you, you would still prefer X to Y as long as it only costs 1 cent more than Y. If this is the case, we can extract a million dollars from you with the following set-up:

---

[11] This argument is adapted, more or less, from the one in John Broome and Wlodek Rabinowicz "Backwards Induction in the Centipede Game," *Analysis*, LIX, 4 (October 1999): 237–42, at pp. 240–41.

[12] It may be objected that this is too hasty. We have shown that refusing to trade at node 1 is irrational. But maybe trading would also be irrational? Maybe the agent would act irrationally whatever she would do at node 1? To dispel this worry, we note that it is perfectly consistent to suppose that trading at node 1 is rational. Such a move is obviously rational if the agent expects (as she well might) that she would accept the second trade and end up with B if she were to reject the first trade. Given this expectation, it is rational for her to trade A for A^- at node 1, since she prefers A^- to B.

The Ruinous Upfront Money Pump

$A - \$1,000,000.02$
$B - \$1,000,000.01$
$C - \$1,000,000.00$
$A - \$999,999.99$
$\vdots$
$A - \$0.03$
$B - \$0.02$
$C - \$0.01$
A

The moves prescribed by backward induction in the Ruinous Upfront Money Pump could be defended a by a *reductio* similar to the one that we have used to defend the backward-induction solution for the Upfront Money Pump. But, due to the size of the decision tree in the Ruinous Upfront Money Pump, the moves prescribed by backward induction can more conveniently be defended by induction on the length of the decision problem.[13]

We have seen how the backward-induction solution can be defended for the Upfront Money Pump. Is there a similar argument for the choices recommended by backward induction in the Money Pump with Repeated Offers? There is not. To see this, we similarly try to derive a contradiction in that pump from the assumption that (i) moving down to node 3 is rational. On this assumption, it follows that the agent at node 3 will act rationally and have trust in her future rationality at nodes she can reach by rational moves. Given this trust and the fact that she prefers *C* to *A*, she will expect to trade *A* for *C* at node 7 if we assume, again for reductio, that (ii) moving down to node 7 is rational. What about the move to node 6? This move cannot be irrational, since it guarantees the agent at least C and quite possibly might lead to *B*, which she prefers to *C*. So the move to node 6 must be rational. But then, upon this move, the agent would trade *C* for *B*, which in its turn implies that the move to node 7 cannot be rational, as its rationality would imply that it will lead to *C* which the agent prefers less than *B*. We therefore conclude that assumption (ii) leads to contradiction. This implies that moving up to node 6 is uniquely rational

[13] See Rabinowicz "Grappling with the Centipede," *op. cit.*, pp. 108–09. See also Robert J. Aumann "On the Centipede Game," *Games and Economic Behavior*, XXIII, 1 (April 1998): 97–105 and Broome and Rabinowicz "Backwards Induction in the Centipede Game", *op. cit.*, pp. 241–42.

in node 3. She will therefore make this move in node 3, whereupon she will also act rationally at node 6 and thus trade C for B in that node. Hence B will be her final holding. Now, to derive a contradiction from assumption (i), we would need to show that, if she (instead of moving to node 3) moved up to node 2, she would eventually arrive at A^- (which she prefers to B) as her final holding. But we cannot show this as long it is possible that moving to node 2 would be irrational. For then, upon making such an irrational move, the agent might not retain her trust in her future rationality or even act rationally any more. Consequently, it could not be excluded that, if she moved to node 2, she would eventually end up with C as her final holding. If she at node 1 expects C to be her final holding if she were to move to node 2, then the hypothesis that moving to node 3 is rational does not lead to any contradiction: The agent prefers B to C and, as we have seen, if moving to node 3 is rational, then B would be her final holding if she made that move.

So, being BI-terminating, the Upfront Money Pump has a clear advantage in comparison with the Money Pump with Repeated Offers. And, on top of that, it is considerably simpler.[14]

Appendix: Dougherty and Cantwell

Tom Dougherty suggests another money-pump set-up to get around the problems with backward induction.[15] His set-up assumes that the options A, B, and C come in fungible units. You start off with one unit of each of A, B, and C. Then you are offered three trades: first, you may trade one unit of A for one unit of C; second, you may trade one unit of C for one unit of B; and, third, you may trade one unit of B for one unit of A provided you pay a small sum of money. If you accept all these trades, you end up paying for the same amounts of A, B, and C you started with. The set-up can be diagrammed as follows:

---

[14] In this paper, we abstain from discussing whether vulnerability to exploitation does establish that there is some irrationality in the agent's psychological make-up and, in particular, whether vulnerability to money pumps does show that it is irrational to have cyclic preferences. A proper discussion of this issue would require another paper.

[15] Tom Dougherty "A Deluxe Money Pump," *Thought*, III, 1 (March 2014): 21–29, at pp. 25–28
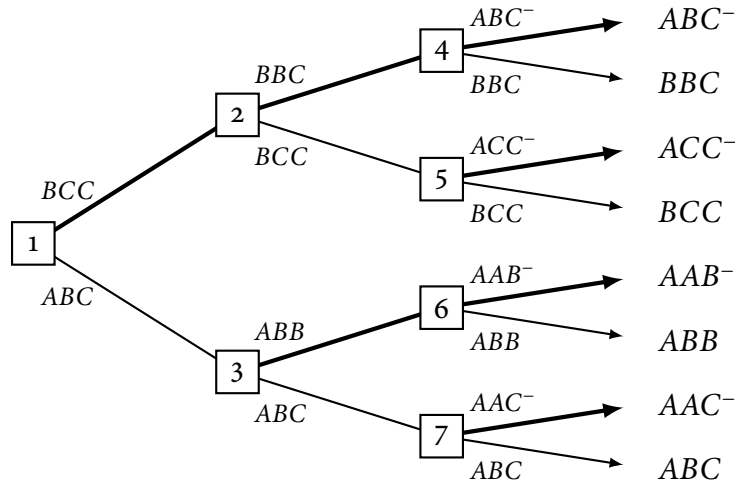
$$A^- > B > C > A > A^-.$$

In this case, the argument for exploitation does not need to rely on backward induction; we can instead appeal to the principle of statewise dominance: For each trade, the agent prefers to accept it whatever other choices she has made or is going to make at later nodes. But Dougherty himself admits that this money pump is significantly less general than the Money Pump with Repeated Offers, since it requires the objects of trades to come in quantifiable amounts and in addition assumes that the agent's preferences for extra units of A, B, and C do not depend on how many units of these goods she already has. Furthermore, the principle of dominance applies only if what the agent does at a choice node does not causally affect her moves at later choice nodes. Assuming this causal independence is problematic. Dougherty attempts to establish causal independence by the assumption that (i) your choices do not change your later preferences and (ii) you make each choice solely on the basis of your preferences.[16] But not even these questionable assumptions seem to be sufficient. Your current choices can negatively affect you future rationality. And, if you are irrational when you face the later trades you may perversely turn them down just because you prefer to accept them.

Another set-up that does not rely on backward induction has been put forward by John Cantwell:[17]

---

[16] Dougherty, *ibid.*, p. 28.
[17] John Cantwell "The Pragmatic Stance, Whither Dutch Books and Money Pumps?,"

The Three-Way Money Pump
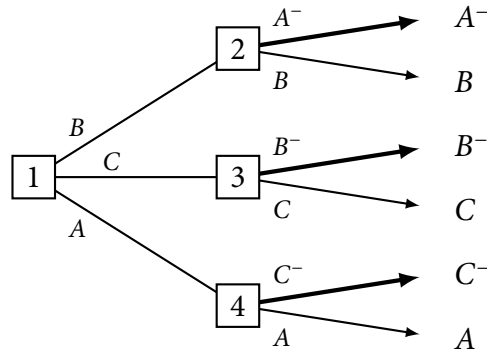
$$A > A^- > B > B^- > C > C^- > A.$$

In this set-up, you end up paying for an option you could have had for free, no matter what you choose at node 1. The option you end up paying for, however, is not necessarily the option you had at the beginning. This makes the set-up unsuitable as an exploitation scheme.[18] There is no clear sense in which the agent here gets exploited; indeed, she might become better off than she was before. And it is not clear whether there is an exploiter who profits—that is, whether anyone becomes better off at the agent's expense. The Upfront Money Pump does not suffer from these problems.

[18] See Arif Ahmed "Exploiting Cyclic Preference," *Mind*, CXXVI, 504 (October 2017): 975–1022, at pp. 1009–11.