Comments on Wolfgang Schwarz's "Lost Memories and Useless Coins: Revisiting the Absentminded Driver"

Mike Titelbaum
titelbaum@gmail.com
sites.google.com/site/michaeltitelbaum

University of Wisconsin-Madison

2012 Formal Epistemology Workshop

Outline

1. Coins in the Head

2. Coins in the Hand

The Hobgoblin Game

The game

There is a button in front of you. I will ask you to decide whether to push it, then make you forget your first decision, then ask you to decide again. If you push the button exactly once you win $1M; otherwise you get nothing.

[Compare (Richter 1986)]

Suppose perfect Uniformity:
Each time you choose, your credence that you Push the other time conditional on your Pushing this time is 1; similarly for no-Push.

(If you don't like memory erasure, we can play the game with you and your identical twin.)

On EDT, your expected (and certain!) payoff is $0.
(To dramatize, if I take pity and offer you $0.01 if you Push twice, EDT jumps on that option.)

Causal Hobgoblin

Key for CDT: One choice doesn't affect the other.

But how confident are you that you'll Push right now?
If your credence you'll Push is > 0.5, then your credence you Push the other time is > 0.5, so no-Push has higher expected utility.

But resolving to no-Push makes that more likely, then Push is preferred.
(Compare Death in Damascus, Matching Pennies game.)

Only stable credence in Push is 0.5.
Schwarz: You should be undecided between the two options.
Indecision comes in degrees.
In this case the degree of indecision should be 0.5—perfectly undecided.

(If I sweeten double-Pushing, indecision level moves a bit towards Push.)

Indecision as a Decision Process

Schwarz: Suppose the agent has an internal mechanism for choosing when undecided.

When indecision level is c, ideal agent's mechanism seems to him to be stochastic with chance c of choosing one option.

This is *not* a new act added to the decision problem.
It's an internal way of choosing among the acts already offered.

In Hobgoblin Game, stochastic choice with chance 0.5 of Push maximizes long-run average return.
In Absentminded Driver, stochastic choice with chance 2/3 of continuing maximizes long-run average return.

Question: This works by loosening the Uniformity assumption.
(Assumption is now that you choose with the same *chance* each time.)
Why not give EDT the same option? What would happen then?

What's Indecision?

Degrees of indecision might be like:

1. I'm weighing whether to see Batman or Spider-Man movie. Haven't finished tallying pros and cons, but at this preliminary stage I'm *leaning* towards Batman. (Is that just a prediction of my ultimate decision?)
2. I lean towards chocolate ice cream over vanilla. But I'm not totally decided on chocolate. Any time you give me a choice, the chance is c I'll pick chocolate.

Why bother with degrees of indecision at all?
Why not just say that an ideal agent will have credence c that he'll make a particular choice? (Compare Arntzenius 2008 and others.)

Outline

1. Coins in the Head

2. Coins in the Hand

What's "halfing"? What's "thirding"?

Given that the first coin flipped has objective chance c of tails (continue), what's the driver's credence in first-coin-tails when he reaches an exit?

	Halfing	Thirding
$P(\text{Tails}_1)$	c	$2c/(c+1)$

When $c = 1/2$ (Sleeping Beauty Problem):

	Halfing	Thirding
$P(\text{Tails}_1)$	$1/2$	$2/3$

Idea behind thirding: Conditional on being at the first exit, the first coin hasn't been flipped yet and credence $= c$ it'll come up tails. Conditional on being at the second exit, the first coin definitely came up tails! So calculate a weighted average of c and 1.

Schwarz: Expected utilities for the Absentminded Driver have always been calculated by assuming the thirder credence distribution.

Intersecting two questions

If you have to choose a fixed coin bias ahead of time, a choice of $c = 2/3$ will seem to yield the best long-run average payoff.

Schwarz demonstrates that this coin bias is considered optimal by the following combinations of positions:

- CDT and thirding
- EDT and halfing

Intriguingly, (Briggs 2010) argues on Dutch Book and accuracy grounds that "Causal decision theorists should favor the Thirder Rule, while evidential decision theorists should favor the Halfer Rule." (p. 32)

Schwarz also shows that for the CDT halfer, allowing randomization does not make the unstable decision problem stable—contrary to lore. Lore's argument assumes utility of randomized choice equals expected utility relative to chosen bias (i.e. objective chance). But in Absentminded Driver credences deviate from objective chances—even for the halfer!