Components of "Everyday Kantianism": Symmetry and causal illusions*

Jeffrey Goldberg[†] Lívia Markóczy[‡] Lawrence Zahn[§]

SJDM Poster Version 1.3
November 13, 2001

Abstract

The existence of cooperation in the face of temptation to free ride requires explanation. We discuss two psychological phenomena, "symmetry" and "the illusion of control", which we believe underly the "what if everyone acted that way" type of reasoning used in some types of cooperation. We provide a simple model of how these lead to cooperation. We also show how some bizarre causal beliefs, such as effect preceding cause, can follow from these phenomena. We also look at some existing evidence for these phenomena and present a simple model which illustrates how they can lead to cooperation. We report on our studies which support the model.

*Poster for SJDM meeting November 2001. Full version of draft can be found at `http://www.goldmark.org/jeff/papers/symmetry/symmetry.pdf`

[†]`Jeffrey@goldmark.org`, `http://www.goldmark.org/jeff/`

[‡]AGSM, University of California, Riverside. `Livia.Markoczy@ucr.edu`, `http://www.goldmark.org/livia/`

[§]AGSM, University of California, Riverside. `Lawrence.Zahn@ucr.edu`

Illustration of the phenomenon

D: Why are you reviewing? I'm sure that lots of people are reviewing for the conference and one person isn't going to make a difference.

C: True, one person won't make a difference, but lot's of "one persons" will.

D: Yes, but you are only *one* "one person". Let the others do it.

C: But suppose everybody thinks that way, then nobody would review.

D: But they don't all think like that. And if they did you would be left reviewing alone, so it still doesn't make sense to review.

C: Well if I decide not to review, other people who think like me may also decide not to.

D: That's crazy! Most of them have probably already made their decision and nobody is watching to see what *you* do.

C: Well it might be crazy, but that is the way that I feel, and I am glad that many people feel the same way.

The road less travelled

Approaches to transforming the PD into things with cooperative solutions can be divided into three categories

1. Allow for decisions to cooperate or defect to have consequences beyond the game itself. If a reputation for defecting will leave you excluded from future games or get you beaten up, then it is worthwhile to avoid such a reputation.

2. Emotions or other devices change the pay-offs, so that the problem is no longer a PD. If guilt feelings would make Alice feel so bad for defecting, then it effectively subtracts from the gain she would get for defecting, and so she would not actually be confronted with a true PD.

3. Cognitive quirks of reasoning, like the ones we propose in this paper, may transform the PD into something in which a cooperative solution is viable. Symmetry will convert a PD into something known as Newcomb's Problem (Lewis, 1985; Nozick, 1985). The illusion of control will also lead to a cooperative solution to Newcomb's problem.

In this paper we focus only on something that is in the third category, not because we think it is the most important – we don't think that – but because it is the most often overlooked by those examining the collective action problem.

Symmetry

Symmetry, in its most general form, is uncontroversial and indisputable: People predict how others would behave by imagining or remembering their own behavior.

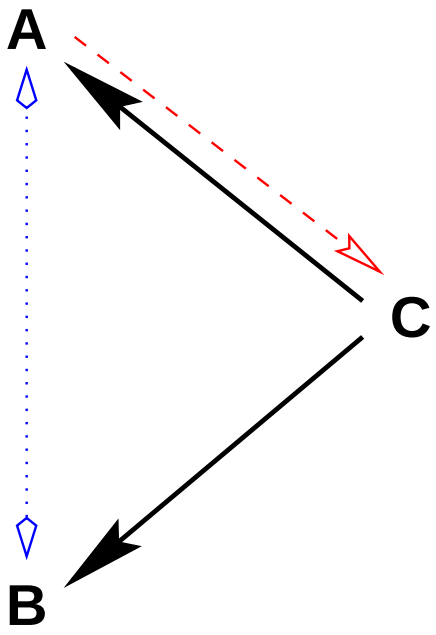
Symmetry, however, is not enough to explain the choice to cooperate. It may explain a prediction that the other will cooperate as well, but to get from "my actions are correlated with Bob's actions" to "I will cooperate so that Bob will cooperate" requires the illusion of control as well as symmetry.

The illusion of control

We borrow the phrase "illusion of control" from Langer (1975), but use it far more generally. It has been called other things by other people in different contexts. It is the belief that if your choice is correlated with something else then your choice must be the cause.

The Calvinist doctrine of predestination – to use an example from Quattrone and Tversky (1984) and Elster (1989) – exhibits the illusion of control to a very high degree: (1) those who are among the "elect" are elected by the deity before birth; (2) those who are elected will live virtuously; (3) therefore, if one wants to be among the elect, one should decide to live virtuously.

The confusion between diagnostic correlation and causal relationship is ever present in the human mind. In each of these cases, people are manipulating the consequence of something in order to effect its cause.



C is the cause of both A and B as indicated by the solid arrows. A has some illusory control over C as indicated by the dashed arrow. The perceived correlation between A and B is indicated by the dotted arrow.

Figure 1: Causal and non-causal relations

Examples of illusion of control

Consider the diagram in Figure 1 on the preceding page. In the case of the Calvinist, we can have A be "virtuous living", B be "going to heaven", and C be "being among the elect". According to the theology A and B are caused by C with no other causal relations involved. Yet a Calvinist will make efforts to live virtuously, presumably in the belief that A can sometimes cause C.

Imagine some inherited gene that has two effects: It decreases one's chance of heart disease and it increases one's tolerance for cold. In Figure 1 on the page before we can assign "having the gene" to C, "tolerance for cold" to A, and "non-tendency for heart disease" to B. People aware of the existence of such a trait might, if they fell victim to the illusion of control, try to display a higher tolerance for cold than those unaware of such a gene. Quattrone and Tversky (1984) performed exactly such an experiment and found exactly the increased tolerance for cold that the illusion of control would predict.

Another example is the much studies voter's illusion, characterized by Quattrone and Tversky (1984) on the voter's illusion discussin why people take the trouble to vote.

> [I]f one votes, then one's politically like-minded peers, who think and act like oneself, will also vote. Conversely, if one abstains, then one's like-minded peers will also abstain. Because the preferred candidates could defeat the opposition only if the like-minded citizens vote in larger number than do the unlike-minded citizens, the individual may conclude that he or she had better vote. [p. 244]

Again looking at Figure 1 on the preceding page, A can be "me voting", B can be "others like me voting" and C can be "whatever properties of socialization, norms and mindset lead me and others like me to vote".

	B does same as A	B does not same as A
A cooperates	A gets: 3	A gets: 0
A defects	A gets: 1	A gets: 5
	$p = .90$	$(1 - p) = .10$

Figure 2: Prisoner's Dilemma recast as Newcomb's Problem

The Expected Utility of Symmetry

Suppose that Alice suspects that there is a very high chance (say 90%) that Bob will come to the same decision in a PD as she does. Following Lewis (1985), we can recast the PD into what we see in Figure 2.

The probability that B decides the same way that A does the symmetry probability or p_s. Figure 3 lists various results for different p_s values.

	B does as A	B does not as A	Expected Utilities for $p_s =$ 1	.90	$\frac{5}{7}$	.60	.50
A coop.	3	0	3.00	2.70	2.14	1.80	1.50
A defects	1	5	1.00	1.40	2.14	2.60	3.00

Figure 3: Expected Utilities of strategies given various p_s values

Our studies

We have argued that there is actually a large body of experimental work which support our view on symmetry and the illusion of control, but that that work was not conducted or evaluated in the light of they theory we have presented here. For example, we believe that Shafir and Tversky (1992) have already demonstrated that the illusion of control plays a role in cooperation in PDs, although they came to a different conclusion.

Timing effects and "the White Queen's Principle"

If people see their choices as occurring before (in some unspecified sense of "before") the other person's choice they should be more inclined in invoke the illusion of control. We are aware of a number of unpublished[1] studies which have attempt to do this by setting up problems where subjects were told either that the others had already made their choice, or others had yet to make their choice we should expect that the illusion of control would play a larger role where the subject believes that they are going first. This is because it is harder to believe two impossible things (backwards causality and causality backwards in time) then just one impossible thing (backwards causality). However with very few exceptions those (unpublished) studies failed to show clear results. One exception is [Cite michael morris here] who did find weak results as we would predict, but attributed those results to other factors.

[1] It is a serious problem for the field that negative results are very difficult to publish.

Why most timing effect experiments have failed

We believe that those timing effect experiments failed for three reasons:

1. The effect that we are going after is relatively weak. Thus anything which substantially complicates the task of the subjects is likely to overwhelm the effect we are trying to isolate.

2. By having players play in different orders, the symmetry of the situation is broken. Subjects may no longer see that the other player is being faced with the same decision problem. This will also weaken the effect.

3. People who accept backwards causality are inclined to accept causality backwards in time with little extra effort. That is, once you believe the first impossible thing, believing the second is easy. At least that is the case for these two impossible things.

We used simple PDs that got at "timing" effects without breaking symmetry. Normal textual description of PDs presented with the trees in figures 5 on page 13 and 6 on page 14.

Hypothesis 1 *Our hypothesis is that those presented with the problem presented with figure 5 will be more like to cooperate than those presented with figure 6*

To deal with the third problem we identified those who are inclined to take on both backwards causation backwards in time, by presenting them with a non-supernatural version of Newcomb's problem. Those who take the one box must accept both backwards causation and causation backwards in time together.

Hypothesis 2 *The effect described in hypothesis 1 will be less pronounced among those who select one box in Newcomb's problem.*

Another hypothesis, which perplexingly is *not* fully supported by our data is

Hypothesis 3 *There should be a correlation between cooperation on a prisoners dilemma and taking one box in Newcomb's problem.*

	Coop	**Defect**	**Total**
Subject first	58	36	94
Other first	32	39	71
Total	90	76	165

Table 1: Cooperation and PD presentation

The 2001 study

Subjects had two tasks. First a Newcomb's problem task as in figure 4 and then with one of four varieties of a prisoners dilemma.

The two variants of the diagram are shown in figures 5 and 6 respectively.

Results 2001 study

We found that those presented with the "you first" version of the trees cooperated significantly more than those presented with the "other first" versions. The data are summarized in Table 1 which yield a $\chi^2 = 4.512$ and a one-tailed $p = 0.025$ by Fisher's exact test.[2] This confirms hypothesis 1.

We looked at how strong this effect was among those who took one box in Newcomb's problem and those who took two (table 2). For one-boxers in Newcomb's problem we see no timing effect at all ($N = 82$, $\chi^2 = .743$ *ns*) while for those who took both boxes, there is a strong effect ($N = 83$, $\chi^2 = 3.717$, one tailed $p = 0.044$). This supports hypothesis 2.

To our surprise we did not find a relation between taking one box in Newcomb's problem and cooperating in the PD as seen in table 3.

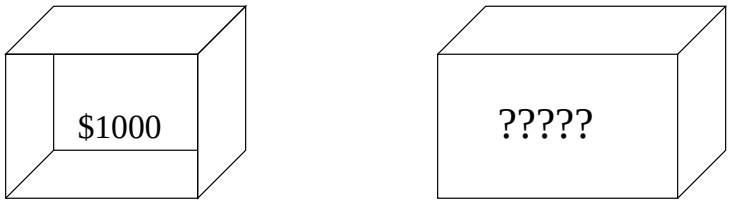
[2]We report the χ^2 for information, but calculate p using Fisher's exact test throughout.

In the figure are two boxes. One is open and you can see that it contains $1000; the other box is closed and you cannot see into it. The closed box contains either $100,000 or nothing. Your choice will be between

(a) taking only the closed box, *or*

(b) taking *both* boxes.

If the problem were as simple as this, it would be obvious that the best choice would be to (b) ("take both boxes"). But the problem is not that simple.



Imagine that there is a super-intelligent space alien which is an expert on human psychology and can, after a brief examination, predict individual human behavior extremely well. It has examined you some time in the past, and has either put $100,000 into the closed box or left it empty. If it thought that you would take both boxes, it left the closed one empty. If it thought that you would take only the closed box, it put the $100,000 in it.

You know that this alien psychologist is extremely accurate at this kind of prediction, and has done this with hundreds of people before you and has never (yet) made an error. But you also know that the procedure is carefully audited, and the money is already placed (or not placed) in the closed box before you are presented with this choice.

Based on this, please answer whether you take (**A**) only the closed box, or (**B**) both boxes.

Please circle the strategy you choose.

A **B**

Figure 4: Newcomb's problem as presented

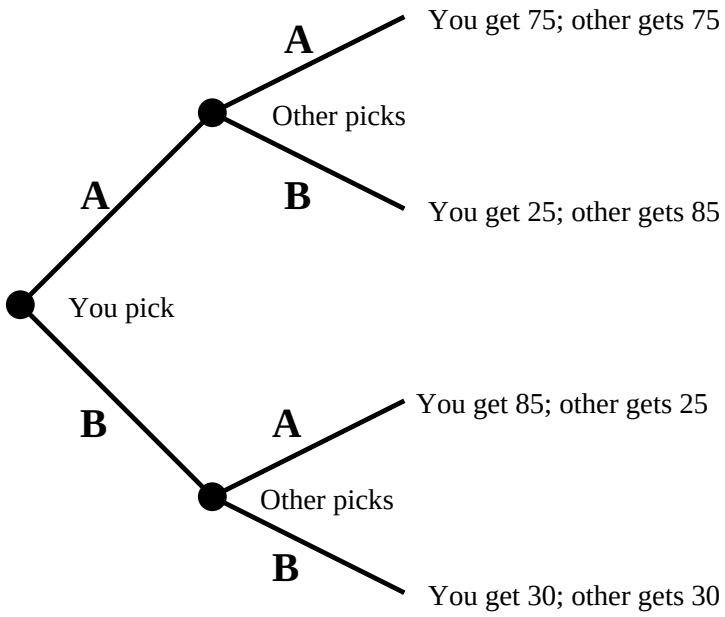


Figure 5: The "you first" presentation of the decision tree

One box

	Coop	**Defect**	**Total**
Subject first	35	23	58
Other first	12	12	24
Total	47	35	82

Two box

	Coop	**Defect**	**Total**
Subject first	23	13	36
Other first	20	27	47
Total	43	40	83

Table 2: Cooperation and timing by Newcomb's response

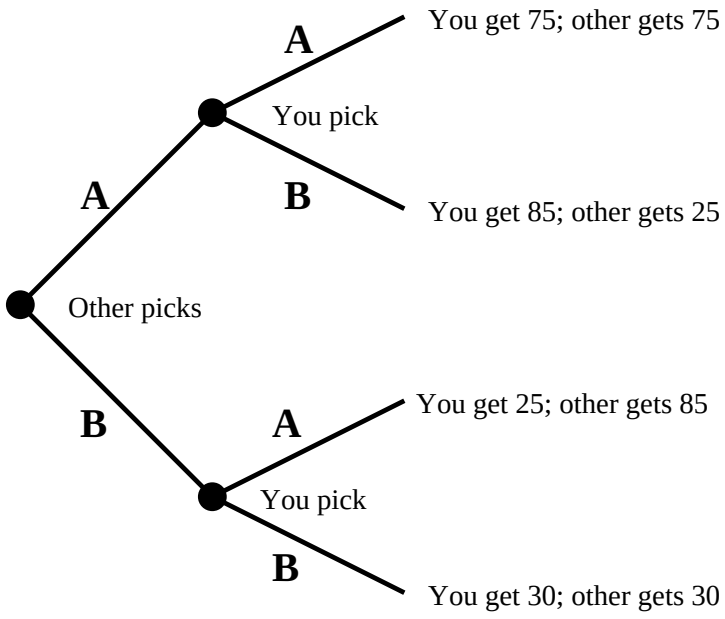


Figure 6: The "other first" presentation of the decision tree

	Coop	**Defect**	**Total**
One box	66	57	123
Two boxes	65	67	132
Total	131	124	255

Table 3: Newcomb's and cooperation in 2001 study

	Coop	**Defect**	**Total**
One box	43	9	52
Two box	25	30	55
Total	68	39	107

Table 4: Newcomb's and PD for 1998 study

The 1998 Study

Before we considered hypothesis 2 attempted a series of studies involving PDs in which subjects were told of the person their response will be paired with (a) has already made their choice; (b) is making their choice as you are; or (c) will make their choice some time in the future. The results of such attempts were erratic for reasons we suggest above.

In one of the studies involving full-time MBA students ($N = 107$) at Cranfield University in the UK we also queried Newcomb's problem along with the PD. The wording of Newcomb's problem was similar to that presented in 4 except that the values were £1000 and £100,000 for what may be in the boxes. Also the PD was first and there were several intermediate tasks before the Newcomb's problem task.

Here we found an overwhelmingly strong relation between cooperation on the PD and taking one box in Newcomb's problem ($N = 107$, $\chi^2 = 14.4$, one tailed $p < 0.0001$). See table 4.

Closing thoughts

1. While we have made progress in experimentally isolating symmetry and the illusion of control, there remain some still unexplained confounding factors that need to be isolated with more experiments.

2. Even though the evidence which fully isolates the illustion of control and symmetry as contributing to cooperation in various games can be hard to get reliably, there is so much out there by others (plus what we've added) the components of Everyday Kantianism are well supported by the literature as a whole.

3. How big of a role this plays in cooperation as compared to other explainations for cooperation is an entirely open question. As are individual differences in Everyday Kantianism.

4. Speculation about where symmetry and the illusion of control come from can be found in the full paper.

References

Elster, Jon (1989). *The Cement of Society: A study of social order*. Studies in Rationality and Social Change. Cambridge: Cambridge University Press. Cited on: 4

Langer, Ellen J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32(2): 311–328. Cited on: 4

Lewis, David (1985). Prisoners' dilemma is a Newcomb problem. In *Paradoxes of Rationality and Cooperaton: Prisoner's Dilemma and Newcomb's Problem* (eds. Richmond Campbell and Lanning Sowden), chapter 14, pp. 251–255. Vancouver: University of Vancouver Press. (Originally published in *Philosopy and Public Affairs*, Volume 8, no 3, 1979). Cited on: 3, 7

Nozick, Robert (1985). Newcomb's Problem and two principles of choice. In *Paradoxes of Rationality and Cooperaton: Prisoner's Dilemma and Newcomb's Problem* (eds. Richmond Campbell and Lanning Sowden), chapter 6, pp. 107–133. Vancouver: University of Vancouver Press. (Originally published in *Essays in Honor of Carl G. Hempel*, 1969, D. Reidel, Dordrecht). Cited on: 3

Quattrone, George A. and Amos Tversky (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology*, 46(2): 237–248. Cited on: 4, 6

Shafir, Eldar and Amos Tversky (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, 24: 449–474. Cited on: 8