# PARRONDO'S PARADOX AND EPISTEMOLOGY – WHEN BAD THINGS HAPPEN TO GOOD COGNIZERS (AND CONVERSELY)

Fredrik Stjernberg

Philosophy, Linköping University
frest@irk.liu.se

ABSTRACT: According to Parrondo's Paradox, there will be cases where a subject facing two probabilistically losing strategies has a probabilistically winning strategy by combining these losing strategies in a random way. Conversely, the subject can have two winning strategies, that when combined result in a probabilistically losing strategy. This unexpected result has found applications in economics, biology and electronic engineering, to name a few cases. It is argued that this result should have some applicability in epistemology as well.

## 1. Introduction

When navigating the world, we are forced to do as best we can, managing with tools that are less than perfect. Our environment is at times hostile, and it is easy to get things wrong. Probabilistically based methods will of course get things wrong every now and then, but in the long run, they will serve us better than making do without a method. But how are such tools and rules of thumb related to human knowledge, our attempts to know something about the world? One need not be a Cartesian absolutist to find the connections between knowledge and probability puzzling (though it might help): knowledge is held to entail truth, but every kind of probabilistically based assertion is consistent with the falsity of what is being asserted. If I have a lottery ticket, the chances of winning are extremely slim, say one in a million. So I have very good probabilistic support for saying that I won't win. But it seems that I can't say that I *know* I won't win, no matter how many tickets there are in the lottery, not until the results are announced.[1] Some have expressed a very dim view of the relations between

[1] For recent discussions of this lottery problem, see for instance Williamson (2000) ch. 11 and Hawthorne (2005, pp. 21ff).

Printed from: *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.* Ed. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson & D. Egonsson, 2007. www.fil.lu.se/hommageawlodek

probabilistic reasoning and knowledge, as Hume: ”But knowledge and probability are of such contrary and disagreeing natures, that they cannot well run insensibly into each other, and that because they will not divide, but must be either entirely present, or entirely absent.”[2] Others have been more optimistic, connecting probabilistically grounded methods directly with knowledge. Ramsey and various kinds of reliabilists would be examples of this approach.

As a view of knowledge, reliabilism has always been a bit underspecified. Saying that something is knowledge if it is arrived at by reliable means has much to recommend it, but there has been a curious silence on *how* reliable the method should be – how good does a method have to be? Buying a lottery ticket is a *very* reliable way to lose money, but I still don’t want to say that I know that I will lose. Setting a numerical value, and saying that above it we have knowledge, below only belief, is no simple task – no matter where we draw the line, it appears that we are just arbitrarily drawing a line, and that no interesting demarcation is being captured by our legislating.[3] One option would be to dispose of the notion of knowledge altogether, just resting with the degree to which a given belief is justified. Another option that doesn’t leave room for such doubts about arbitrariness is to hold that only *completely* reliable methods can give us knowledge, that we need the probability of being right to be 1. This handles the difficulty at hand, but goes against the gist of the reliabilist conception of knowledge, taking it right back to a kind of Platonist or Cartesian infallibilism. We might protest that this fiddling with numerical values is of little relevance to what the reliabilist means. Of course, reliabilists typically don’t come out and provide a numerical value for how reliable a certain method of obtaining information is supposed to be, and many suggested methods are in fact hard to give a precise numerical value. But it seems that without some kind of numerical value, talk of reliability threatens to be little more than a *façon de parler*. Even if a reliabilist might be hard pressed to here and now provide a numerical value,  indicating when there is sufficient reliability for knowledge, he shouldn’t hold that providing numerical values is in principle misguided.

So according to such a view, a state of the subject is held to be knowledge if it is arrived at by a reliable process – the state can be branded knowledge, if the process used is such that it has a certain high (very high) success rate. The other way to see things is to concentrate on to what extent some given belief is justified, without worrying too mcuh about whether it all adds up to knowledge. A recently discovered paradoxical result in game theory can turn out to create difficulties for both these views, however.

---

[2] Hume, *Treatise*, I.iv.I.

[3] This is in itself perhaps worrying, but it is not any kind of knock-down argument against the reliabilist. Some such problem will be around for virtually *all* accounts of knowledge (*how* justified does a belief have to be in

## 2. Parrondo's Paradox

According to Parrondo's Paradox, a combination of two losing strategies can sometimes have good (winning) results, and vice versa. Discussions of this unexpected game-theoretical result have pointed to possible applications in physics, populations genetics, economics, electronic engineering and sociology.[4] The result should have some relevance for epistemology as well.

Parrondo cases are as follows. A subject *A* is situated on a set of stairs, numbered –500, –499, ... , 0, 1, 2, ..., 500. *A* starts at 0, and the object of the game is to climb up the stairs. There are two different strategies available for *A* in playing the game, *S* and *C*, for *Simple* and *Complicated*, respectively.

Game *S*:

*A* flips a coin $c_1$, heads win and tails lose. But the coin is biased by $\varepsilon = 0.05$:

$Pc_1(\text{heads}) = 0.45$

$Pc_1(\text{tails}) = 0.55$

This is obviously a losing game, so *A* would inexorably walk down the steps by repeatedly using this strategy. Game *C* is more complicated.

Game *C*:

As before, we flip a coin, and heads wins, tails loses. This time, we have two coins, $c_2$ and $c_3$.

One coin, $c_2$, has: $Pc_2(\text{heads}) = 0.745$ (i.e., $0.75 - \varepsilon$)

$Pc_2(\text{tails}) = 0.255$ (i.e., $0.25 + \varepsilon$)

The other coin, $c_3$, has: $Pc_3(\text{heads}) = 0.095$ (i.e., $0.1 - \varepsilon$)

$Pc_3(\text{tails}) = 0.905$ (i.e., $0.9 + \varepsilon$)

Which coin is to be flipped? If you are situated on a step whose number is not a multiple of 3, you flip $c_2$, otherwise you flip $c_3$. This is also a losing game, even if it is harder to see this.

---

order to be knowledge?). Knowledge may well turn out to be a vague concept, and could perhaps be handled in different ways that are analogous to familiar positions in discussions of vagueness.

[4] See Harmer & Abbott [1999].

Two of the coins, $c_1$ and $c_3$, are bad for *A*, whereas c*2* gives *A* a good chance of winning. One might believe that game *C* is good for *A* in general, for the following reason: since coin $c_3$ is used 1/3 of the time, and $c_2$ is used 2/3 of the time, the probability of winning should be:

$$P_{win} = (2/3)\,(0.745) + (1/3)\,(0.095) \approx 0.53$$

which is better than 0.5, hence a good game. But this reasoning is not correct, because $c_3$ is *not* used one third of the time: it is used more often than that. It turns out that *A*, due to what has been called a "ratchet effect", has difficulties in getting above a step 3*n*, which is a multiple of 3, whereas he every now and then descends to a lower multiple, 3*n*–3. Then the pattern repeats, so *A* will descend the stairs. Assume that *A* starts on step 0. A flip of $c_3$ will probably take *A* to step –1, and then flipping $c_2$ will probably take *A* back to 0 again. *A* will probably be oscillating between 0 and –1 for a while. The probability of flipping the bad coin, $c_3$, is thus somewhat higher than 1/3, somewhere between 1/3 and 1/2. The net effect of this is that game *C* is also a losing game. It has been shown that *C* is a losing game for $\varepsilon > 0$.[5]

The steps that are multiples of 3 thus turn out to be a kind of ceiling for *A*, not completely impenetrable, but hard to break through. And then the whole process repeats itself, which each step that is a multiple of 3 being a new ceiling for *A*, while *A* goes down the steps. So *C* is also a losing game for *A*.

But surprisingly enough, *A* can *combine* the two games, and end up winning by alternating between the games. There are several ways to do this. One is to choose games randomly, another is to alternate between them systematically, by for instance playing *SSCCSSCC*, and so on. Simulations have shown that both these strategies, and yet others, enable *A* to climb the stairs.[6] Combining two losing strategies can thus lead to a winning strategy. This is of course strictly speaking not a paradox; it is highly counterintuitive, however.

Not only can two losing strategies be combined; since the assignment of values to heads and tails is arbitrary, we can show that it works the other way round as well: two winning strategies can be combined systematically yet still get a losing result.

---

5 See Harmer & Abbott [1999] for details.

6 There are several programs on the Internet, simulating Parrondo situations. See for instance Juan Parrondo's website for links, http://seneca.fis.ucm.es/parr/GAMES, which gives a few links to simulation programs as well. There is also a mathematical proof of the winning strategy, not just a simulation; see Harmer & Abbott [1999].

This kind of result has been shown to have applications in economics, for instance in portfolio management. These results should have an application in epistemology as well. We can think of *A*'s situation as that of a person trying to arrive at a correct theory of the world, using probabilistically reliable strategies for gathering knowledge. Intuitively, 0, *A*'s starting point, is some kind of case where *A* doesn't know anything about the world, and doesn't really have any theory at all about what the world is like, –500 is a more or less completely false theory about the world, and 500 is something like the state of an omniscient being.[7] So A's task is to climb the stairs, and acquire as much knowledge as possible. If *A* get things right – wins a toss of the coin – *A* can take a step up the stairs; if *A* loses, he has to take a step down.

As described above, both strategies *S* and *C* for acquiring knowledge are failures – there's a greater chance of arriving at falsehoods. We can think of *S* and *C* as somewhat unreliable groups of informants. Yet by randomly combining the strategies, *A* can still get better and better results – climb the stairs, arrive at knowledge, or increasing the justification of the beliefs, by using two losing methods. He will also be able to do the opposite: use two winning – epistemically good – strategies in a combination, yet get further and further away from knowledge the longer he does it.[8]

## 3. Discussion

Our usual understanding of knowledge seems to be at variance with these results. Both results – that bad strategies when combined randomly yield good results, and that good strategies when combined randomly yield bad results – are at odds with how we would like to think of the acquisition of knowledge. Having bad theories with a guarantee of arriving at a correct theory of the world is at the very least at variance with the letter of reliabilism, as ordinarily conceived.[9] On the usual understanding of reliabilism, something is knowledge if it is arrived at through a reliable method. Here we can arrive at full knowledge about the world through unreliable methods, or else use reliable

[7] The end-points are a bit controversial here. Davidson often presented arguments to show that completely false theories about the world are impossible, and Grim has presented an argument to show that omniscience is impossible (see Bibliography). These concerns are none of our business here, and the reader can let the end-points represent as much falsehood and omniscience as she can see it fit to countenance. It is enough that we can be better or worse off in our theories about the world.

[8] There is a problem in adapting Parrondo's Paradox to epistemology that admittedly is a bit troubling. How are we to determine which coin to use in game *C*? In the original Parrondo cases, the steps are simply capital owned by the subject, and there is no problem in giving that a numerical value. It appears to be difficult to find a clear-cut and realistic example of what epistemic capital would be like in real life, but I'll leave it as an exercise for the reader to come up with a good example. Even a clearly arbitrary metric will do the job required for the Parrondo Paradox, as long as we get an ordinal scale.

[9] It also appears to be in conflict with a "no false lemma" demand on knowledge, often invoked in discussions of Gettier cases. If something is to be knowledge, it cannot be based on falsehoods or false beliefs. In the case at hand, we appear to be basing our knowledge claims on false methods, methods probabilistically guaranteed to give you the wrong answer.

methods and arrive at a false theory of the world. The idea that such beliefs about the world – beliefs arrived at by using substandard methods in a random fashion – are justified is a bit worrying.

But isn't this just another example of something we knew all along – that we can get things right while using bad methods, or that we can get things wrong even if we are doing our best? Or that there are cases where it can pay to do something irrational? I think Parrondo cases point to something deeper and more worrying. These cases point to a systematic and pervasive way in which certain available strategies, when combined, lead to unexpected results. We might want to liken the cases with instances of epistemic luck, happening to get things right, but this situation is somewhat different. Parrondo paradox cases would not be cases of some kind of epistemic luck, but would be systematic, and it could even be known by the subject that the available strategies were of that kind. The subject could have a guarantee that she will get things right, even if each move in seeking knowledge is probabilistically guaranteed to be a bad choice. The large-scale and systematic nature of Parrondo cases points towards a situation we didn't really think we might find ourselves in: being at an epistemically good place by using epistemically bad methods, or, conversely, being at an epistemically bad place by using epistemically good methods.

The possibility of Parrondo cases seems to pull us in two directions. Consider the case where *A* has played the game a great number of times, and is near the +500 level. *A*'s theory of the world is very good, but is it knowledge? On the one hand, many want to say that it is *obviously* not a case of knowledge – *A*'s epistemic situation has not been the result of proceeding in a responsible way. The methods used for gathering knowledge have not been reliable, so the end result is not an instance of *knowing* that the world is such and such. Yet others may want to say that this demand is too strong; after all, the subject is *right* about the world, and the process, taken as a whole, is guaranteed to get to the point where she has a correct theory: what could it matter whether each single step was ill-advised? As long as the end result is guaranteed to be as successful as we could expect from probabilistic reasoning, why be concerned that the *parts* of the amalgamate "method" (changing randomly between *S* and *C*) are not reliable? The path that *A* takes through the knowledge stages is more like a kind of biased random walk; there is no systematic attempt at getting things right.

A very brief and minor survey among philosophically trained informants has not resulted in any real consensus.[10] It even seems that many harbour *both* the above intuitions, and that we simply without good reason assumed that there would never be a clash between them, since ordinarily, using combinations of good methods will lead to good results. Parrondo cases show that this assumption doesn't hold generally, and should force us to revise some of our thoughts about knowledge and probabilistic reasoning.

### 4. Making the assumptions a bit more realistic

There is a problematic amount of idealization and abstraction in the above case, and it is not entirely clear to me how one should react to it. The original Parrondo cases require a clear answer to the question where the agent is on a scale. It is important that we can give the agent's situation a numerical value, since the construction of the complex game *C* works only if we can tell which of the two subgames the agent is to play at a particular point. For the economic case, we can simply use the amount of capital the agent has at a given moment. Other extensions also require some suitable numerical value for the various stages in the game.

In adapting the Parrondo Paradox for the case of knowledge, I simply assumed without much argument that this requirement of numerical values is not problematic, and claimed that as long as we could arrange amounts of knowledge on an ordinal scale, given some kind of metric, we could get the ratchet effect going. But there are still a few problems in making this work. We can make good enough sense of an original state of no knowledge, and give this state the numerical value 0. This state would then not contain mistaken theories about the world, but be a kind of Lockean *tabula rasa*. We can think of this state as something an inquirer is in, before any steps in the enquiry have been taken. A person who is completely fooled by appearances, who is in the kind of situation the skeptic fears, would then be in a negative state. Such a person would have to *unlearn* a great many things, before arriving at knowledge about the world. So it at least makes sense to speak of states with negative numerical values, where the agent is worse of than an agent who doesn't know anything. And for the states with positive numerical values, it at least has *some* foundation in how we think of knowledge, and especially of the growth of knowledge.

We – *pace* stray Kuhnians in our midst – think that it does make sense to speak of a given agent knowing more than another, or an agent amassing knowledge over time. We think of this sequence as displaying at least transitivity: if *a* knows more than *b*, *b* knows more than *c*, then *a* knows more than *c*. But the relation is not connected. It at least seems to be possible for two agents to know different things, yet it might make very little sense to say that one of them *must* know more than the other, or else their bodies of knowledge are identical. Consider for instance two agents, who have a roughly average knowledge of geography, with exceptions. One agent knows virtually everything there is to

---

[10] Why should we only rely on philosophically trained informants – perhaps untrained informants give just as interesting and suggestive answers? This is quite possible, and Stich, Warfield and Nichols (2000) provide one

Printed from: *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.* Ed. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson & D. Egonsson, 2007. www.fil.lu.se/hommageawlodek

know about Liechtenstein, the other knows virtually everything there is to know about Laos. Apart from their special competences, they know more or less the same things. Perhaps we could *impose* a metric on their states of knowledge to settle which one of the agents that knew most, but such a decision seems unavoidably arbitrary. There is no clear sense of one of the agents knowing more than the other that is there to be captured by us. The task would be simple if it were possible to give numerical values to the agents' states of knowledge by counting: how many propositions does the agent know? But this seems pretty hopeless. If I know that Vaduz is in Liechtenstein, I know that Vaduz is in Liechtenstein or Oslo is north of Stockholm, Vaduz is in Liechtenstein or bachelors are married, and so on.

We can give a slightly different feel for what the paradox is like for epistemology by the following, ridiculously idealized, example. Imagine that there are 100 empirical propositions, $a_1$ to $a_{100}$, that can be arranged as follows. They can be put into an order of knowledge, from 1 to 100. If a subject knows anything at all, she knows $a_1$. If she knows $a_{100}$, she is omniscient (omniscient for all practical purposes, given how the end-points turned out to be problematic; see fn. 7 above). No subject knows $a_{100}$ without being omniscient. These empirical propositions are arranged so that if a subject knows $a_n$, she knows all the preceding propositions $a_{n-m}$. There are no shortcuts in the acquisition of knowledge; there is a regular progress in knowledge of these propositions. They have to be learned in a specific order, $a_1$, $a_2$, ..., $a_{100}$. A higher index will mean that the subject is closer to omniscience, more advanced in the acquisition of knowledge. Not many (not any?) speakers get as far as $a_{100}$, Two speakers that are on the same step in this order will be counted as knowing equally much, irrespective of how much else they happen to know. So now we can group all knowers with respect to their position on the stairs to omniscience. If we accept this picture, we can apply the above reasoning directly, and show that a subject might be on a given level $a_n$, reason randomly with an array of bad methods for acquiring knowledge, yet have a probabilistic guarantee of arriving at $a_{100}$ ("epistemic nirvana"). Since the propositons described are empirical propositions, the very use of probabilistic methods in getting to know them should not be seen as fundamentally wrong. Descending the stairs is harder to understand on this picture, since we don't normally *unknow* things we used to know. What we could have here is rather that the subject, being on step $a_n$ might find his beliefs being without support, and thus retracting that belief, which would mean descending to $a_{n-1}$.

The discussion above was conducted in terms of knowledge, but as I said above, there is no real need to frame the discussion in terms of knowledge. We can skip all talk about knowledge, and instead concentrate on the agent's degree of justification, without worrying about whether the agent knows something or not. The agent could start off with a group of beliefs that are justified to degree *n*, and

---

example of what asking the unschooled could yield.

use the above strategies to get new pieces of information. This simpler situation is also worrying for our view of what it could mean to have probabilistically justified beliefs. Again we get the result that the subject could be justified in her beliefs, use a combination of bad methods, still be guaranteed to arrive at ever better justified sets of beliefs, and we can again ask what effect the badness of the methods have for the justification of the beliefs.

## 5. Concluding words

So how important is this result for philosophy, and especially epistemology? Some things indicate that it is of minor interest. The idealization is difficult to get a clear grip on (have we idealized beyond what is warranted?), and I have said nothing about whether there are any life-like examples of the mixed probabilistic strategies needed to get Parrondo cases. Also, it is obviously nowhere close to some of the more well-known – or proper – paradoxes, such as Russell's paradox, which sent people back to their drawing-boards, revising a whole philosophical programme. It is also nothing like the Liar, which shows that our grip on a central philosophical concept was more problematic than we might have believed. But Parrondo's paradox seems to be of both interest and relevance for the epistemologist, however. The existence of Parrondo's paradox does indicate something surprising and counterintuitive for our chances of getting to know something about the world in a systematic, albeit probabilistic, way. How serious is the result? I will leave that for others to decide, but it seems to me that the very fact that we in an admittedly very idealized setting can get a probabilistic guarantee that we will do quite well epistemically, even if we (stupidly?) persist in using bad epistemic strategies in a random way, does show that our understanding of the acquisition of knowledge, or the justification of our beliefs, by means that are only probabilistically safe should be reexamined.[11] [12]

## Bibliography


Davidson, D., "A Coherence Theory of Truth and Knowledge" (1983), reprinted in Davidson (2001)

Davidson, D., *Subjective, Intersubjective, Objective*, Oxford 2001, pp. 137-158


[11] As before, there is also a converse result: if we go on using epistemically sound strategies, combined in a random manner, we well get into epistemic difficulties. Start off with some knowledge, reason soundly – probabilistically soundly – and we might still find ourselves in a situation where we are worse off than we used to be, before we learned the sound strategies. So the sound strategies are making us worse off. In that case, why persist in calling the strategies sound?


[12] Thanks to Jan Odelstad for valuable information concerning the niceties of measuring and of devising scales.

Grim, P., 1988, ”Truth, Omniscience, and the Knower”, *Philosophical Studies* 54, 9-41

Grim, P. and A. Plantinga, 1993, ”Truth, Omniscience and Cantorian Arguments” *Philosophical Studies* 71, 267-306

Harmer, G. P. and D. Abbot, “Parrondo’s Paradox”, *Statistical Science*, 14 (1999), 206-213

Hawthorne, J., *Knowledge and Lotteries*, Oxford 2005

Hume, D., *A Treatise of Human Nature*, ed. by L.A. Selby-Bigge, Oxford 1968

Parrondo, J., website at http://seneca.fis.ucm.es/parr/GAMES

Williamson, T., *Knowledge and its Limits*, Oxford 2000