

expectation
 Supertasks
 Ross
 Two
 Utility
 Airtight
 binding
 decisions
 Altadena
 Guy
 paradox
 backwards
 Book
 antinomy
 Petersburg
 infinite
 Wine
 Pasadena
 theory
 Bayes's
 puzzles
 cooperation
 Surprise
 decision
 betting
 St
 Envelope
 rationality
 strategic
 probability
 Supererogation
 Choice
 rational
 induction
 Gamble
 Contradiction
 Series
 Choice
 Exam
 odds
 Sleeping
 indifference
 blindspots
 Theorem
 Game
 paradoxical
 Cable
 inference
 Ever
 Doomsday
 Beauty
 additivity
 Paradox
 Better
 Probabilism
 Centipede

50032 - Paradoxes of Rationality

Jake Chandler

Institute of Philosophy, KU Leuven
Center for Mathematical Philosophy, University of Munich
✉ jacob.chandler@hiw.kuleuven.be



University of Bayreuth
Winter 2010-2011

Overview

Decades of academic investigation in fields such as probability theory, decision theory and game theory, built on solid commonsense intuition, have provided us with a firm grasp of some of the basic principles that underpin rational choice and inference. Or have they? This course will present a selection of fascinating puzzles and paradoxes that point to various tensions and lacunae in the received view of rationality.

It will be assumed that the student will have some prior familiarity with the basic principles of (i) decision / game theory and (ii) probability theory. A refresher on infinite sets, which play a role in many of the puzzles discussed, will be given in the third session.

Admin

The course will consist of 14 weekly 2-hour discussion sessions, every Wednesday from 16:00 to 18:00, in S 35 (NW I). Attendance is compulsory. The working language of the seminar is English.

We shall be following a format already established in some of the modules at Bayreuth. Each session will be kickstarted by one or more beamer-based *student presentations*, based on a short essay (see below). Presentations will be expected to be 20 min long and will each be followed by a 10 min Q&A session. The essay on which the presentation is based is to be emailed to me by 09:00 on the morning of the day before the seminar and will be returned to the student, with marks and comments, at the end of the session.

The remainder of the session will be devoted to a *class discussion* of the topic. To help ensure that this is a productive exercise, you will be expected to have prepared a short list of issues that you think are worthy of general attention. Pertinent issues include points of unclarity, perceived weaknesses in the arguments or suggestions for further work on the topic.

Method of assessment will vary according to the amount of credits taken. For all credit options:

- (i) Short essay of 2 000 \pm 15% words + associated presentation (pass/fail).

In addition, for the 2 *credit* option:

(ii) (a) Short essay of 2 000 \pm 15% words (pass/fail). Deadline for receipt, via email: 11 February 2011.

For the 6 *credit* option:

(ii) (b) Extended essay of 4 500 \pm 15% words. Deadline for receipt, via email: 25 March 2011.

For the 8/10 *credit* option:

(ii) (c) Extended essay of 7 000 \pm 15% words. Deadline for receipt, via email: 25 March 2011.

Regarding the short essay(s), you should be simply be aiming for a concise, well-structured critical overview of one of the session topics. Allocation of topics for item (i) above will be made during the first session. For the extended essay, you will be expected to provide something more substantial, going beyond a mere literature review. For this assignment, you may chose any topic connected with the course, subject to my approval. Note that I am happy to make suggestions if you are stuck. Please book an appointment with me, via email, for the first week following the Christmas break to discuss your proposed choice. You will then be granted two weeks after the meeting to mail me a 750 word provisional abstract.

Essays must be word-processed (no handwritten contributions please!), double-spaced and properly referenced. Note that the deadlines are firm and are renegociable only under exceptional circumstances.

Schedule & Reading

I will make available pdfs of various relevant articles. Please do not hesitate to contact me if you require even further reading. The schedule will be the following (compulsory reading is marked with a ★):

1. *Introduction to the course*

2. *The nature of paradox* We take a look at the philosophical literature on the the nature of paradox, as well as some interesting recent work calling into question the standard assumption that no contradictions can be rationally believed.

★ Lycan, B. [ms]: ‘What, exactly, is a paradox?’

Priest, G. [1998]: ‘What’s so Bad about Contradictions?’, *Journal of Philosophy* 95, pp. 410-426.

Quine, W.V. [1966]: ‘The Ways of Paradox’, in his *The Ways of Paradox and Other Essays*, New York: Random House, pp. 3-20.

Sainsbury, R.M. [2009]: *Paradoxes, 3rd Edition*, Cambridge: Cambridge University Press. Ch. 7 ‘Are Any Contradictions Acceptable?’.

3. *Getting to grips with infinity* A number of the paradoxes discussed in this course crucially involve the infinite in some way, shape or form. In this session, we review some essential concepts, notation and results which you may or may not be familiar with.

Oppy, G. [2006]: *Philosophical Perspectives on Infinity*, Cambridge: Cambridge University Press. Ch. 2 ‘Mathematical Preliminaries’.

★ Suber, P. [1998]: ‘A Crash Course in the Mathematics of Infinite Sets’, *St. John’s Review*, XLIV(2), pp. 1-59. <http://www.earlham.edu/peters/writing/infapp.htm>

4. *Doomsday* ‘One might at first expect the human race to survive, no doubt in evolutionary much modified form, for millions or even billions of years. . . Contemplating the entire history of the race—future as well as past history—I should in that case see myself as a very unusually early human. I might well be among the first 0.00001 per cent to live their lives. But what if the race is instead about to die out? I am then a fairly typical human. Recent population growth has been so rapid that, of all human lives lived to far, anything up to about 30 per cent. . . are lives which are being lived at this very moment. Now, whenever lacking evidence to the contrary one should prefer to think of one’s own position as fairly typical rather than highly untypical. To promote the reasonable aim of making it quite ordinary that I exist where I do in human history, let me therefore assume that the human race will rapidly die out.’¹ Shocking conclusion, but where, if anywhere, does the reasoning go wrong?

Bartha, P. & C. Hitchcock [1999]: ‘No One Knows the Date or the Hour: An Unorthodox Application of Rev. Bayes’s Theorem’, *Philosophy of Science* 66, pp. 339-353.

Bostrom, N. [2002]: *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, NY: Routledge. Ch. 6 ‘The Doomsday Argument’ and Ch. 7 ‘Invalid Objections Against the Doomsday Argument’.

★ Leslie, J. [1996]: *The End of the World*, New York, Routledge. Ch. 5 ‘The Doomsday Argument’, pp. 187-236. Read up until p. 205 plus a few sections from ‘Objections to the Argument’ (whichever ones you feel like).

Korb, K.B. & J.J. Oliver [1998]: ‘A Refutation of the Doomsday Argument’, *Mind* 107(426), pp. 403-10.

5. *Sleeping Beauty* ‘Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you to back to sleep with a drug that makes you forget that waking. When you are first awakened, to what degree ought you believe that the outcome of the coin toss is Heads?’² This question turns out to be far from straightforward to answer.

¹Leslie, J. (ed.) [1990]: *Physical Cosmology and Philosophy*, Macmillan Publishing Company.

²Elga [2000]

Arntzenius, F. [2001]: 'Reflections on Sleeping Beauty', *Analysis* 62, pp. 53-62

Bradley, D. & H. Leitgeb [2006]: 'When Betting Odds and Credences Come Apart: More Worries for Dutch Book Arguments', *Analysis* 66, pp. 119-127.

★ Elga, A. [2000]: 'Self-Locating Belief and the Sleeping Beauty Problem', *Analysis* 60, pp. 143-147.

★ Lewis, D. [2001]: 'Sleeping Beauty: Reply to Elga', *Analysis* 61(3), pp. 171-176.

Hitchcock, C. [2004]: 'Beauty and the Bets', *Synthese* 139(3), pp. 405-420.

6. *The Cable Guy* 'You are certain that a cable guy will visit you tomorrow between 8 a.m. and 4 p.m. but you have no further information. . . [Y]ou agree to a bet on whether he will come in the morning interval (8, 12] or in the afternoon interval (12, 4). At first, you have no reason to prefer one possibility rather than the other. But you soon realise that there will definitely be a future time at which you will (rationally) assign higher probability to an afternoon arrival than a morning one, due to time elapsing. You are also sure there may not be a future time at which you will (rationally) assign a higher probability to a morning arrival than an afternoon one. It would therefore appear that you ought to bet on an afternoon arrival.'³ Was your afterthought a good one? Or should you have stuck with your first intuition?

★ Hájek, A. [2005]: 'The Cable Guy paradox', *Analysis* 65, pp. 112-119.

Kierland, B., B. Monton & S. Ruhmkorff [2008]: 'Avoiding certain frustration, reflection, and the cable guy paradox', *Philosophical Studies* 138, pp. 317-333.

Rowbottom, D. & P. Baumann [2009]: 'To Thine Own Self Be Untrue: A Diagnosis of the Cable Guy Paradox'. *Logique et Analyse* 51 (204), pp. 355-364.

★ Weintraub, R. [2009]: 'A Solution to the Cable Guy Paradox', *Erkenntnis* 71, pp. 355-359

7. *Ever Better Wine!* 'Before you gleams a bottle of Ever Better wine. . . The wine slowly improves with age. . . More good news: You are immortal. Consequently, you are indifferent as to when you consume a particular good. When should you drink the wine? Not now. The wine will be better later. Not later. For at any given time it will be true that the wine will be even better if you waited longer. But if you do not drink the wine now and do not drink it later, then you will not drink it at all! What went wrong?'⁴

★ Pollock, J.L. [1983]: 'How Do You Maximize Expectation Value?', *Nous* 17, pp. 409-422.

³Rowbottom & Baumann [2009]

⁴Sorensen, R. [2004]: 'Paradoxes of Rationality', in A.R. Mele & P. Rawling (eds.) *The Oxford Handbook of Rationality*, Oxford: Oxford University Press.

Slote, M. [1989]: *Beyond Optimising: A Study of Rational Choice*, Cambridge: Harvard University Press. Ch. 5 ‘On Rational Dilemmas And Rational Supererogation’.

Mintoff, J. [1997]: ‘Slote On Rational Dilemmas And Rational Supererogation’, *Erkenntnis* 46, pp. 111-126.

- ★ Sorensen, R. [1994]: ‘Infinite decision theory’, in J. Jordan (ed.) *Gambling on God*, Lanham MD: Rowman & Littlefield. Section II ‘Infinite Options’ A-D.

Sorensen, R. [2006]: ‘Originless Sin: Rational Dilemmas for Satisficers’, *The Philosophical Quarterly* 56(223), pp. 213-223

8. *The St. Petersburg Game* A fair coin is tossed until it lands heads. If it first lands heads on toss n , the player receives 2^n euros. Since there are infinitely many possible outcomes, the expected monetary value cannot be calculated as per usual. Instead, we standardly first order the outcomes into a sequence $\langle o_1, \dots \rangle$, then take the limit of the partial sum $\sum_{i=1}^m \Pr(o_i)U(o_i)$ as m goes to infinity. Whichever way we order the outcomes, we obtain ∞ . If choice goes by maximal expected monetary value, we should be prepared to pay any finite sum for an opportunity to play this game. But that seems like a gross overvaluation.

Colyvan, M. [2008]: ‘Relative Expectation Theory’, *Journal of Philosophy* 105(1), pp. 37-44.

- ★ Martin, R. [2008]: ‘The St. Petersburg Paradox’, In Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/fall2008/entries/paradox-stpetersburg/>.

Neugebauer, T. [ms]: ‘Moral Impossibility in the Petersburg Paradox: A Literature Survey and Experimental Evidence.’ Introduction & Sections 1-2.

9. *The Pasadena and Altadena Games* As in the St Petersburg game, we toss a fair coin until it lands heads. But this time, if the first head appears on toss n , the payoff is $(-1)^{n-1}2^n/n$ euros. Remarkably, for any real number r , we can order the outcomes so that the limit of the relevant partial sum is equal to r . So what value should we put on this gamble? One may of course say that the expectation is undefined here. But if so, then so too is that of the Altadena game, in which the payoffs for each outcome are augmented by one euro, and which is clearly a preferable gamble.

- ★ Nover, H. & Hajek, A. [2004]: ‘Vexing expectations’, *Mind* 113(450), pp. 237-249.

Colyvan, M. [2006]: ‘No expectations’, *Mind* 115(459), pp. 695-702.

Hájek, A. & H. Nover [2006]: ‘Perplexing Expectations’ *Mind* 115(459), pp. 703-720;

Fine, T.L. [2006]: ‘Evaluating the Pasadena, Altadena, and St Petersburg Gambles’, *Mind* 117, pp. 613-632.

10. *The Two Envelopes* ‘You are presented with two sealed envelopes, *A* and *B*, and you know that one of them contains a cheque for twice as much money as the other. You are allowed to select one of them at random. You are then offered the chance to swap and take the other instead. If your selected envelope contains x , and your swap is lucky, you get $2x$, but if you are unlucky you get $0.5x$. So your expected utility if you swap is $2x/2 + 0.5x/2$, which is $1.25x$. So it looks as if you should swap. However, exactly the same argument would have been available if you had picked the other envelope in the first place.’⁵

★ Broome, J. [1995]: ‘The two-envelope paradox’, *Analysis* 55(1), pp. 6-11.

Chalmers, D.J. [2002]: ‘The St. Petersburg Two-Envelope Paradox’, *Analysis* 62, pp. 155-157.

Clark, M. & N. Shackel [2000]: ‘The Two-Envelope Paradox’, *Mind* 109(435), pp. 415-442.

★ Jackson, F., P. Menzies & G. Oppy [1994]: ‘The two envelope paradox’, *Analysis* 54(1), pp. 43-45.

Priest, G. & G. Restall [2008]: ‘Envelopes and Indifference’, in C. Dégrement, L. Keiff & H. Rückert (eds.) *Dialogues, Logics and Other Strange Things, essays in honour of Shahid Rahman*, College Publications, pp. 283-290.

11. *Infinite Utility Streams* It seems plausible that value is temporally additive, i.e. that the value of a state of affairs is the sum of the values of its temporal parts. But now ‘suppose... that time extends infinitely into the future and that an agent has a choice between producing two units of value at each time or one unit of value at each time. Intuitively, it would seem that the former outcome is better than the latter outcome. [By temporal additivity, T]he total value produced, however, is the same: infinity in each case.’⁶

★ Vallentyne, P. & S. Kagan [1997]: ‘Infinite Value and Finitely Additive Value Theory’, *The Journal of Philosophy* 94(1), pp. 5-26

Hamkins, J. & B. Montero [2000] ‘With Infinite Utility, More Needn’t be Better’, *Australasian Journal of Philosophy* 78(2), pp. 231-240.

Lauwers, L. & P. Vallentyne [2004]: ‘Infinite Utilitarianism: More Is Always Better’, *Economics and Philosophy* 20, pp. 307-330.

Oppy, G. [2006]: *Philosophical Perspectives on Infinity*. Ch. 6 ‘Probability and Decision Theory’, section 6.5 ‘Infinite Utility Streams’.

⁵Clark, M. [2007]: *Paradoxes from A to Z, 2nd Edition*, London: Routledge.

⁶Vallentyne, P. [2001]: ‘Infinity in Ethics’, in E. Craig (ed.) *Routledge Encyclopedia of Philosophy, electronic expansion*, London: Routledge.

12. *Decision-Theoretic Supertasks* ‘Suppose that at time 0 there are infinitely many one dollar bills that have been numbered using the natural numbers (1, 2, 3, etc.). . . Suppose that at time 0 God has possession of all the dollar bills, but offers to transfer them to you by the following scheme: At each of the times, $1/2$, $3/4$, $7/8$, . . . , $(2^n-1)/2^n$, . . . , she will give you two arbitrarily chosen bills from those not yet given to you, and she will also then destroy the lowest numbered bill in your possession. . . You are not permitted to use these bills until time 1. . .’⁷ Is God’s offer worth paying for?

★ Barrett J. & F. Arntzenius [1999]: ‘An infinite decision puzzle’, *Theory and Decision* 46, pp. 101-103.

★ Machina, M.J. [2000]: ‘Barrett and Arntzenius’s, infinite decision puzzle’, *Theory and Decision* 49(3), pp. 293-297.

★ Pulier, M.L. [2000]: ‘A flawed infinite decision puzzle’, *Theory and Decision* 49(3), pp. 291-292.

Barrett J. & F. Arntzenius [2002]: ‘Why the Infinite Decision Puzzle is Puzzling’, *Theory and Decision* 52(2), pp. 139-147.

Allis, V. & T. Koetsier [1991]: ‘On Some Paradoxes of the Infinite’, *The British Journal for the Philosophy of Science* 42(2), pp. 187-194.

van Bendegem, J.P. [1995]: ‘Ross’ Paradox is an Impossible Supertask’, *The British Journal for the Philosophy of Science* 45(2), pp. 743-748.

13. *The Surprise Exam* ‘A teacher announces to her students that there will be a surprise examination one day next week—a surprise in the sense that on the morning of the exam, the students will not know that the exam will be on that day. The students reason. . . as follows. The first such exam cannot be on Friday. For if, on Thursday night, the exam has not yet been held, we will know that it must be on Friday; and so it will not be a surprise. But then, it cannot be on Thursday either; for if on Wednesday night the exam has not yet been held, we will know that it must be on Thursday, so it will not be a surprise. But then it cannot be on Wednesday either. . . So it must be on Monday; in which case it will not be a surprise either. So there can be no such exam.’⁸ But surely there can!

★ Priest, G. [2000]: ‘The logic of backwards inductions’, *Economics and Philosophy* 16: 267-285. Sections 1, 2, 3, 5 and 7.

Wright, C. & A. Sudbury [1977]: ‘The Paradox of the Unexpected Examination’ *Australasian Journal of Philosophy* 55, pp. 41-58.

Sorensen R.A. [1982]: ‘Recalcitrant versions of the prediction paradox’, *Australasian Journal of Philosophy* 69, pp. 355-362.

⁷Vallentyne [2001]

⁸Priest [2000]

Sorensen R.A. [1984]: 'Conditional blindspots and the knowledge squeeze: a solution to the prediction paradox', *Australasian Journal of Philosophy* 62, pp. 126-135.

14. *The Centipede* 'There are two players, you and I. There are 100 one dollar notes on the table (and this is known to us). We take it in turns until either the money runs out or someone finishes the game. On any turn, one can take either 1, in which case it is the other's turn; or else take 2, finishing the game. Each of us wants to come away with as much money as possible.' 'If the game ever gets to move 99, I will take 2, ending the game. That way, I am better off. But you know this, so if the game ever gets to move 98, you will end the game. That way, you are better off. But I know this, so if the game reaches move 97... Hence on move 1, I will take 2 and end the game.'⁹ This line of reasoning may seem impeccable but empirical research shows that, in practise, very few people follow the recommendation that ensues.

★ Priest, G. [2000]: 'The logic of backwards inductions', *Economics and Philosophy* 16, pp. 267-285. Sections 4 and 6.

Pettit, P. & R. Sugden [1989]: 'The backward induction paradox', *Journal of Philosophy* 86, pp. 169-82.

Bicchieri, C. [1989]: 'Self-refuting theories of strategic interaction: a paradox of common knowledge', *Erkenntnis* 30, pp. 69-85.

⁹Priest [2000]