

**How to Be Rational:
How to Think and Act Rationally**

David Robert

Copyright © 2021 David Robert
All rights reserved

To my parents

Table of Contents

Introduction	1
Bibliography	2
Section 1: How to think rationally	3
Chapter 1: How to Acquire Rational Belief Attitudes	4
Bibliography	5
Chapter 2: Should You Be Skeptical of Climate Change: Letter to a Climate Skeptic.....	7
Bibliography	9
Section 2: How to act rationally.....	10
Chapter 3: How to Make Rational Choices	11
Bibliography	12
Chapter 4: Expected Comparative Utility Theory: A New Theory of Rational Choice Under Risk.....	13
Introduction	13
The Argument for ECU theory.....	14
Conclusion.....	36
Acknowledgements.....	37
Bibliography	37
Chapter 5: Should You Invest in Life-Extension Medical Research?.....	39
Bibliography	40
Chapter 6: The Existential Passage Hypothesis	41
1. Summary of the Conclusions	41
2. Assessment of the Arguments	43
3. Further Comments and Suggestions.....	48
4. Concluding Remarks.....	48
Acknowledgements.....	49
Bibliography	49

Introduction

It has always appalled me that really bright scientists almost all work in the most competitive fields, the ones in which they are making the least difference. In other words, if they were hit by a truck, the same discovery would be made by somebody else about 10 minutes later.

—*Aubrey de Grey*

To make *rational choices* involves choosing courses of action that best serve one's ends. As a good example, one's chances of having a positive impact on the world are significantly greater if one focuses one's efforts on solving the world's biggest, most urgent problems than if one focuses one's efforts on solving comparatively small, non-urgent problems that will not ultimately matter in the long run if the biggest, most urgent problems are left unsolved. Therefore, if one's goal is to make a difference in the world, then it is simply not rational to focus one's efforts on solving comparatively small, non-urgent problems.

It is true that functioning societies do need people working on comparatively small day-to-day problems. But a shortage of people working on comparatively small day-to-day problems is itself a big urgent problem and so, something that one would need to factor into any rational decision about where to focus one's efforts.

Moreover, it does not require an exceptionally intelligent person to survey the literature on the various problems that we may be facing and to arrive at a correct assessment of what problems are the biggest and most urgent and what problems are comparatively small and non-urgent. To adopt *rational belief attitudes* is to adopt belief attitudes (i.e., to believe, to disbelieve or to withhold belief) that best serve one's epistemic ends, namely those of acquiring true beliefs and avoiding false beliefs. So, provided that one's methodology for acquiring rational belief attitudes is robust, then there is surely a reasonably approachable threshold of analysis and synthesis beyond which more sophisticated analysis and synthesis will not likely yield a more correct assessment of what problems are the biggest and most urgent and what problems are comparatively small and non-urgent.

We face today an unprecedented number of big urgent problems. With those problems in mind, in this book, I will supply readers with a robust methodology for acquiring rational belief attitudes and making rational decisions. Readers will then be better equipped to identify and solve the world's biggest, most pressing problems.

This book is divided into 2 sections. In Section 1 (How to think rationally), I address how to acquire rational belief attitudes and, on that basis, I consider the question whether one ought to be skeptical of climate change. In Section 2 (How to act rationally), I address how to make rational choices and, on that basis, I consider the questions whether one is rationally required to do what one can to support life-extension medical research and, more broadly, whether one is rationally required to do what one can for the welfare of posterity.

Bibliography

- Gates, B., & Gates, M. (n.d.). *Annual Letters From Bill & Melinda Gates*. Bill & Melinda Gates Foundation. Retrieved April 8, 2021, from <https://www.gatesfoundation.org/Ideas/Annual%20Letters>
- MacAskill, W. (2015). *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. New York: Avery. <http://www.effectivealtruism.org/doing-good-better>
- Singer, P. (2010). *The Life You Can Save: How to Do Your Part to End World Poverty*. New York: Random House Trade Paperbacks,
- Singer, P. (2015). *The Most Good You Can Do: How Effective Altruism is Changing Ideas about Living Ethically*. Yale University Press.
- Todd, B. J. (2016). *80,000 Hours: Find a fulfilling career that does good*. CreateSpace Independent Publishing Platform. <https://80000hours.org/book/>

Section 1: How to think rationally

Chapter 1: How to Acquire Rational Belief Attitudes

How should a layperson think (or form belief attitudes) about any given statement? Is it enough to research reliable sources and use logical, unbiased thinking? In Chapter 1, I will argue that it is not and that a layperson should instead defer to the majority testimony of experts on that statement, if there are any experts on that statement.

For any statement (or proposition), p , there are three possible belief attitudes (or doxastic attitudes, in technical jargon) toward p : to believe p (which is consistent with having a range of degrees of confidence in p), to disbelieve p (to believe that p is false), and to withhold belief about p (to withhold belief about whether p is true or false).

Doxastic attitudes may not be directly voluntary, but they can be informed by guiding principles that are sufficiently *prima facie* plausible or evidenced. Consider the following *prima facie* plausible principle: Unless there are overriding moral objections, one should form *epistemically rational doxastic attitudes*—that is, doxastic attitudes that best serve one's *epistemic ends*, namely those of acquiring true beliefs and avoiding false beliefs to some respective extent.

Generally, when choosing between different courses of action, with a view to achieving one's ends, false beliefs are undesirable more so than true beliefs are desirable. One error or miscalculation is often enough to undermine such a decision. Nevertheless, true beliefs are still obviously essential to choosing the best courses of action in pursuit of one's goals. Therefore, one should aim to maximize one's subjective chances of holding a favorable ratio of true beliefs to false beliefs (Alston, 1985), rather than the highest absolute number of true beliefs or the lowest absolute number of false beliefs.

It follows that unless there are overriding moral objections, one should acquire *epistemically rational doxastic attitudes*—that is, doxastic attitudes that maximize one's subjective chances of holding a favorable ratio of true beliefs to false beliefs.

Now, avoiding fallacies, mitigating one's cognitive biases, and relying on trustworthy sources of information are the most commonly talked about ways of promoting truth-conducive thinking (or, in other words, acquiring epistemically rational doxastic attitudes). But for any given proposition, unless one is an expert on that proposition, there are ways of thinking about that proposition that are considerably more truth-conducive (or epistemically rational) but rarely if ever discussed outside of specialist academic circles—namely, following the rules of epistemic rationality that govern how laypeople should learn from the testimony of experts.

There are cases of credentialed experts defending fringe claims in reliable sources, such as academic journals, without apparently committing any logical fallacies and without showing any overt signs of bias (Curry, 2011; Lemonick, 2010). This is to be expected in the sciences and in philosophy, where progress often requires that individual scientists and philosophers challenge mainstream or even consensus views.

Laypeople who have learned how to discern logical fallacies, cognitive biases, as well as reputable and disreputable sources of information will be poorly equipped to rationally assess the statements made by those credentialed experts *unless and until* these laypeople are made acutely aware of the rules of epistemic rationality that govern how they should learn from the testimony of experts—that is, how they should adjust their own doxastic attitudes in light of what doxastic attitudes experts hold.

These rules of epistemic rationality can be summarised as follows: For any proposition, *p*, unless and until you become an expert on *p*, you are rationally required to hold whatever doxastic attitude is the *majority* doxastic attitude toward *p* among experts on *p*, unless there are not any experts on *p* or there is no *majority* doxastic attitude toward *p* among experts on *p*, in which cases you are rationally required to *withhold belief* about *p* as long as you are not yourself an expert on *p*. There are exceptions to the rule. See Lahno (2014). See also Coady (2012), Goldman (2001), and Huemer (2005).

This rule can be distilled down to the following rule: For any proposition, *p*, a layperson with respect to *p* is rationally required to believe *p* if and only if *p* is believed by the majority of experts on *p*. Of course, there is, in principle, nothing stopping a layperson from *becoming* an expert on *p*. In that case, there are several alternative belief-forming methodologies available to him or her in the published literature (Matheson, 2015).

Anyone can be an expert on any given proposition, regardless of their academic or professional credentials, provided that they have studied or researched sufficiently thoroughly and impartially the evidence that bears on that proposition (Croce, 2019). The key word here is “sufficiently”. Given the Dunning–Kruger effect (i.e., “difficulties in recognizing one’s own incompetence lead to inflated self-assessments”¹), those who lack the academic or professional credentials of credentialed experts should exercise caution in self-identifying as experts, and others should also exercise caution in identifying them as experts.

In the case of philosophical propositions, and especially ethical and religious propositions, it is arguable whether there are any genuine experts on those propositions (De Cruz, 2018; Frances, 2018; Matheson, McElreath, & Nobis, 2018).

Bibliography

- Alston, W. (1985). Concepts of epistemic justification. *The Monist*, 68(1), 57–89.
<https://doi.org/10.5840/monist198568116>
- Coady, D. (2012). *What to Believe Now: Applying Epistemology to Contemporary Issues*. Chichester: Wiley-Blackwell.
- Croce, M. (2019). On What it Takes to be an Expert. *The Philosophical Quarterly*, 69(274), 1–21.
<https://doi.org/10.1093/pq/pqy044>
- Curry, J. (2011). Reasoning about climate uncertainty. *Climatic Change*, 108, 723–732.
<https://doi.org/10.1007/s10584-011-0180-z>
- Curtis, G. N. (n.d.). The Fallacy Files. Retrieved April 8, 2021, from <http://www.fallacyfiles.org/>

¹ Kruger & Dunning, 1999.

- De Cruz, H. (2018). *Religious Disagreement*. Cambridge: Cambridge University Press.
- Dowden, B. (n.d.). Fallacies. *The Internet Encyclopedia of Philosophy*. Retrieved April 8, 2021, from <https://www.iep.utm.edu/fallacy/>
- Dunning, D. (2014). We are all confident idiots. *Pacific Standard*, 7(6), 46–54. <https://psmag.com/social-justice/confident-idiots-92793>
- Frances, B. (2018). Philosophical Expertise. In D. Coady & J. Chase (Eds.), *The Routledge Handbook of Applied Epistemology (Routledge Handbooks in Philosophy)* (pp. 297–306). Routledge. <https://doi.org/10.4324/9781315679099>
- Goldman, A. (2001). Experts: Which Ones Should You Trust? *Philosophy and Phenomenological Research*, 63(1), 85–110. <https://doi.org/10.1111/j.1933-1592.2001.tb00093.x>
- Huemer, M. (2005). Is Critical Thinking Epistemically Responsible? *Metaphilosophy*, 36(4), 522–531. <https://doi.org/10.1111/j.1467-9973.2005.00388.x>
- Is My Source Credible?* (n.d.). University of Maryland University College. Retrieved April 8, 2021, from <https://sites.umuc.edu/library/libhow/credibility.cfm>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Lahno, B. (2014). Challenging the majority rule in matters of truth. *Erasmus Journal for Philosophy and Economics*, 7(2), 54–72. <https://doi.org/10.23941/ejpe.v7i2.167>
- Lemonick, M. D. (November 1, 2010). Climate heretic: Judith Curry turns on her colleagues. *Nature*. <https://doi.org/10.1038/news.2010.577>
- Matheson, J. (2015). Disagreement and Epistemic Peers. *Oxford Handbooks Online*. <http://dx.doi.org/10.1093/oxfordhb/9780199935314.013.13>
- Matheson, J., McElreath, S., & Nobis, N. (2018). Moral Experts, Deference & Disagreement. In J. C. Watson & L. K. Guidry-Grimes (Eds.), *Moral Expertise: New Essays from Theoretical and Clinical Bioethics* (pp. 87–106). Springer. https://doi.org/10.1007/978-3-319-92759-6_5

Chapter 2: Should You Be Skeptical of Climate Change: Letter to a Climate Skeptic

This is a short letter in response to a Reddit post on climate change skepticism (see below):

Until recently, I never read much about climate change other than standard newspaper stuff. I think I have heard most of the arguments for it. I've also heard people say stuff like "everyone who actually studies/is an expert on climate change agrees". Well, Judith Curry studies it and doesn't. As I read through this presentation, it seems balanced and well thought out. In fact, searching through "climate sceptic" sites on the web, I find surprisingly sensible-seeming people with logical arguments and data references. This [Curry, 2018] walkthrough is a good example. So can anyone tell me; am I missing something obvious? Why are the arguments for climate scepticism wrong?

Any feedback is appreciated (u/HCAndersAnd, 2018).

Dear Skeptic,

I think that you are honest and fair-minded. Your skepticism of consensus climate change science is understandable, but it is misplaced. Since you are *not* an expert on the science of climate change, you are *rationaly required* to believe *consensus* climate change science.² But you might ask, "Why should I believe that the climate science claims made by proponents of climate change mitigation are consensus climate science?"

Signatory governments to the Paris climate change agreement base their support for climate change mitigation on the *Fifth Assessment Report (AR5)* of the Intergovernmental Panel on Climate Change (IPCC) (*The Paris Agreement*, n.d.).³ A key finding of the AR5 is quoted below:

Projections of changes in the climate system are made using a hierarchy of climate models ranging from simple climate models, to models of intermediate complexity, to comprehensive climate models, and Earth System Models. These models simulate changes based on a set of scenarios of anthropogenic forcings. [...]

² For any proposition, p , a layperson with respect to p is rationally required to believe p if and only if p is believed by the majority of experts on p . See Chapter 1 ("How to Acquire Rational Belief Attitudes"). See also Coady and Corry (2013).

³ Gao, Gao, and Zhang (2017) note that,

In 2008–2014, the IPCC's Fifth Assessment Report (AR5) made a comprehensive assessment of the climate system change, risks, emission budget, and mitigation pathway choice of 2°C global warming on the basis of the research results available. After scientific assessment and a series of political pushes, one of the three goals reached at the 2015 Paris Climate Change Conference was stated as "Holding the increase in the global average temperature to well below 2°C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5°C above pre-industrial levels" [3].

Relative to the average from year 1850 to 1900, global surface temperature change by the end of the 21st century is projected to *likely* exceed 1.5°C for RCP4.5, RCP6.0 and RCP8.5 (*high confidence*). Warming is *likely* to exceed 2°C for RCP6.0 and RCP8.5 (*high confidence*), *more likely than not* to exceed 2°C for RCP4.5 (*high confidence*), but *unlikely* to exceed 2°C for RCP2.6 (*medium confidence*). Warming is *unlikely* to exceed 4°C for RCP2.6, RCP4.5 and RCP6.0 (*high confidence*) and is *about as likely as not* to exceed 4°C for RCP8.5 (*medium confidence*). {12.4} (IPCC, 2013, p. 18)

You should believe that the above statement (call it *X*) is consensus climate change science because (i) *X* is clear and unambiguous and (ii) *X* is quoted from the work of the IPCC, and there is a consensus among recognized organizations of experts with expertise in climate science that “the work of the IPCC represents the consensus of the international scientific community on climate change science.” (The Science of Climate Change, 2001, p. 1261)⁴

Suppose however that *X* was consensus climate science at the moment of publication of the AR5, but *X* is *no longer* consensus climate science, and you have polling or other unequivocal evidence to support this. It then follows that (i) you should *disbelieve X* if consensus temperature projections have shifted since publication of the AR5 or (ii) you should *withhold belief about X* if there are no longer any consensus temperature projections and no such projections are believed by even the majority of climate scientists.

Suppose instead that *X* was *not* consensus climate science at the moment of publication of the AR5 (and that *X* is not currently believed by even the majority of climate scientists). This would be the most significant and egregious failure of science communication in history. All the climate scientists who were involved in signing off on *X*'s wording would have experienced a lapse or would have engaged in deception. This would have happened in a report, the AR5's *Summary for Policymakers*, that is designed to guide laypersons in making informed policy decisions on climate change. This is very improbable.

Therefore, unless you have polling or other unequivocal evidence that *X* is no longer consensus climate science, the likelihood that *X* is currently consensus climate science is significantly greater than the likelihood that *X* is *not* currently consensus climate science. And so, it is rational to believe that *X* is consensus climate science.

In conclusion, if you were a dissenting *expert* on the science of climate change—that is, if you regarded yourself as such or if you were a practicing climate scientist who disagreed with the conclusions of your peers on climate change science—then you would be entitled to *your own view* on those conclusions. But, judging from your question, you are clearly *not* a dissenting *expert*. As such, I hope to have convinced you that, *unless and until* you become an expert, you are *rationally required* to believe the statement quoted above from the IPCC's AR5.⁵

⁴ See also *Scientific consensus: Earth's climate is warming* (n.d.).

⁵ I noted three earlier uses of this “Letter to a climate skeptic” format: Kaulbars, 2010; Morrison, 2011; and Newman, 2016.

Bibliography

- Coady, D., & Corry, R. (2013). *The Climate Change Debate: An Epistemic and Ethical Enquiry*. London: Palgrave Macmillan. <https://doi.org/10.1057/9781137326287>
- Curry, J. (2018, June 12). *The Debate: My Presentation*. Climate Etc. <https://web.archive.org/web/20180630062836/https://judithcurry.com/2018/06/12/the-debate-mann-tittley-moore-curry/>
- Gao, Y., Gao, X., & Zhang, X. (2017). The 2 °C Global Temperature Target and the Evolution of the Long-Term Goal of Addressing Climate Change—From the United Nations Framework Convention on Climate Change to the Paris Agreement. *Engineering*, 3(2), 272–278. <https://doi.org/10.1016/J.ENG.2017.01.022>
- IPCC (2013). Summary for Policymakers. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P. M. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press. http://www.climatechange2013.org/images/report/WG1AR5_SPM_FINAL.pdf#page=18
- Kaulbars, M. (2010, June 14). *Letter To A Climate 'Skeptic'*. News Junkie Post. <https://web.archive.org/web/20180630073421/http://newsjunkiepost.com/2010/06/14/letter-to-a-climate-%E2%80%9Cskeptic%E2%80%9D/>
- Morrison, D. (2011). Letter to Climate Skeptics. *Skeptic*, 16(2), 10. <http://go.galegroup.com/ps/anonymous?id=GALE%7CA252289589>
- Newman, S. (2016, July 1). *An open letter to climate change sceptics*. 1 Million Women. <https://web.archive.org/web/20180423160439/https://www.1millionwomen.com.au/blog/open-letter-climate-change-sceptics/>
- Scientific consensus: Earth's climate is warming*. (n.d.) Global Climate Change: Vital Signs of the Planet. Earth Science Communications Team at NASA's Jet Propulsion Laboratory. Retrieved April 8, 2021, from <https://climate.nasa.gov/scientific-consensus/>
- The Paris Agreement* (n.d.). United Nations Climate Change. Retrieved April 8, 2021, from <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement>
- The Science of Climate Change (2001). *Science*, 292(5520), 1261. <https://doi.org/10.1126/science.292.5520.1261>
- u/HCAndersAnd (2018, June 12). *What is wrong with these arguments?* Reddit (r/climatechange). Retrieved April 8, 2021, from https://www.reddit.com/r/climatechange/comments/8qnq31/what_is_wrong_with_these_arguments/

Section 2: How to act rationally

Chapter 3: How to Make Rational Choices

Unless there are overriding moral objections, one should make *rational choices*—choices that best serve one’s ends full stop. *Expected utility (EU) theory* is generally accepted as a normative theory of rational choice under risk. EU theory advises agents to rank their choice options (from least to most choiceworthy) according to their EU, where the EU of an option is a probability-weighted sum of each of its possible utilities.⁶ According to Martin Peterson, today, nearly all decision theorists accept EU theory (Peterson, 2017, p. 66). As such, as long as that is the case, laypersons with respect to decision theory will be rationally required to accept EU theory.

Philosopher Johan Gustafsson has argued that choice options in a decision problem should be construed as sets of acts such that one could jointly intentionally perform, at any time t , all the acts in the set, but no additional acts (Gustafsson, 2014). One of the reasons given by Gustafsson is that if one construes choice options as individual acts, then one runs into the *Problem of act versions* (Bergström, 1966; Castaneda, 1968). Consider the following example:

It is raining outside, but Ann will feel invigorated if she takes a brisk walk around the block (10 utiles), more so than if she stays inside (2 utiles). However, Ann has an injured toenail which causes her a great deal of pain when she tries to walk with her rain boots on. She will therefore experience a great deal of pain if she goes out for a walk wearing her rain boots (−30 utiles), more so than if she stays inside wearing her rain boots (−2 utiles). Luckily, Ann has a very comfortable pair of shoes which do not cause her any pain. However, there is a problem: it is raining very hard and her feet will get soaked. Ann will experience considerable discomfort if she goes out for a walk not wearing her rain boots (−15 utiles), more so than if she stays inside not wearing her rain boots (0 utiles).

Let us suppose that Ann assigns probability 1 to the state of the world as described above. Although the utility of the act ‘Ann stays inside’ is lower than that of the act ‘Ann goes out for a walk’, the utility of at least one version of the act ‘Ann stays inside’—that is, ‘Ann stays inside and does not wear her rain boots’ ($2 + 0 = 2$ utiles)—is greater than the utility of all versions of the act ‘Ann goes out for a walk’—that is, ‘Ann goes out for a walk and wears her rain boots’ ($10 + -30 = -20$ utiles) and ‘Ann goes out for a walk and does not wear her rain boots’ ($10 + -15 = -5$ utiles). Thus, intuitively, Ann should stay inside. However, if choice options are construed as individual acts, then EU theory counsels Ann *not* to stay inside, but instead to go out for a walk.

Therefore, to be intuitively plausible, EU theory should be minimally cashed out as follows:⁷

⁶ For a defense of EU theory, see Gintis (2018).

⁷ Inspired by Gustafsson (pp. 593–594).

For any agent, *S*, faced with any decision under certainty or under risk and for any number of mutually exclusive and jointly exhaustive options, or sets of acts, *a*, *b*, *c*, *d* and *e*, such that, for each set, *S* could jointly intentionally perform, at any time, *t*, all the acts in the set, but no additional acts,

- *a* is more choiceworthy than *b*, for *S*, at *t*, if and only if the EU of *S* jointly intentionally performing *a* at *t* is greater than the EU of *S* jointly intentionally performing *b* at *t*, and
- *a* is just as choiceworthy as *b*, for *S*, at *t*, if and only if the EU of *S* jointly intentionally performing *a* at *t* is equal to the EU of *S* jointly intentionally performing *b* at *t*.

This implies the following derivative decision rule for individual acts:⁸

For any agent, *S*, faced with any decision under certainty or uncertainty and for any two mutually exclusive acts, *a* and *b*,

- *a* is more choiceworthy than *b*, for *S*, at any time, *t*, if and only if *a* is logically entailed by every set of acts such that, for each set, *S* could jointly intentionally perform, at *t*, all the acts in the set, but no additional acts and such that, in accordance with EU theory, the set of acts would be more choiceworthy for *S*, at *t* than each set of acts such that *S* could jointly intentionally perform, at *t*, all the acts in the set, but no additional acts and such that the set of acts logically entails *b*, and
- *a* is just as choiceworthy as *b*, for *S*, at any time, *t*, if and only if *a* is not more choiceworthy than *b*, and *a* is logically entailed by every set of acts such that, for each set, *S* could jointly intentionally perform, at *t*, all the acts in the set, but no additional acts and such that, in accordance with EU theory, the set of acts would not be less choiceworthy for *S*, at *t* than each set of acts such that *S* could jointly intentionally perform, at *t*, all the acts in the set, but no additional acts and such that the set of acts logically entails *b*.

Bibliography

- Bergström, L. (1966). *The Alternatives and Consequences of Actions*. Stockholm: Almqvist & Wiksell.
- Castaneda, H.-N. (1968). A Problem for Utilitarianism. *Analysis*, 28(4), 141–142.
<https://doi.org/10.1093/analys/28.4.141>
- Gintis, H. (2018). Rational Choice Explained and Defended. In G. Bronner & F. Di Iorio (Eds.), *The Mystery of Rationality: Mind, Beliefs and the Social Sciences* (pp. 95–114). Springer. https://doi.org/10.1007/978-3-319-94028-1_8
- Gustafsson, J. E. (2014). Combinative Consequentialism and the Problem of Act Versions. *Philosophical Studies*, 167(3), 585–596. <https://doi.org/10.1007/s11098-013-0114-x>
- Peterson, M. (2017). *An Introduction to Decision Theory*. Cambridge University Press.
<https://doi.org/10.1017/9781316585061>

⁸ Inspired by Gustafsson (p. 595).

Chapter 4: Expected Comparative Utility Theory: A New Theory of Rational Choice Under Risk

Abstract: In this paper, I argue for a new normative theory of rational choice under risk, namely expected comparative utility (ECU) theory. I first show that for any choice option, a , and for any state of the world, G , the measure of the choiceworthiness of a in G is the comparative utility (CU) of a in G —that is, the difference in utility, in G , between a and whichever alternative to a carries the greatest utility in G . On the basis of this principle, I then argue that for any agent, S , faced with any decision under risk, S should rank his or her decision options (in terms of how choiceworthy they are) according to their comparative expected comparative utility (CECU) and should choose whichever option carries the greatest CECU. For any option, a , a 's CECU is the difference between its ECU and that of whichever alternative to a carries the greatest ECU, where a 's ECU is a probability-weighted sum of a 's CUs across the various possible states of the world. I lastly demonstrate that in some ordinary decisions under risk, ECU theory delivers different verdicts from those of standard decision theory.

Introduction

Standard decision theory, otherwise known as *expected utility (EU) theory*, counsels agents to rank their choice options (in order of preference) according to their EU and to choose whichever option carries the greatest EU. The EU of an option is a probability-weighted sum of each of its possible utilities. EU theory has been the dominant normative theory of rational choice under risk since the 18th century (Bernoulli, 1738), and in more recent times (1920s–), has received foundational support from both economists and philosophers (Bolker, 1966; Jeffrey, 1983; Ramsey, 1931; Savage, 1954; Von Neumann & Morgenstern, 1947).⁹

In this paper, I will argue for a new normative alternative to EU theory. I will argue that from the fact that we need a graded, quantitative measure of *choiceworthiness* for decisions under certainty and decisions under risk,¹⁰ it follows that we need a new normative theory of rational choice under risk, namely *expected comparative utility (ECU) theory*. I will show that

This paper is an updated version of my published paper “Robert, D. (2018). Expected Comparative Utility Theory: A New Theory of Rational Choice, *The Philosophical Forum*, 49, 19–37. <https://doi.org/10.1111/phil.12178>”

⁹ According to Martin Peterson, today, nearly all decision theorists agree that the “principle of maximizing expected value is the appropriate decision rule to apply to decisions under risk [...] There are no serious contenders” (Peterson, 2017, p. 66).

¹⁰ A *decision under certainty* is a choice situation where an agent is subjectively certain about which state the world is in and where he or she assigns probability 1 to that state being actual, whereas a *decision under risk* is a choice situation where an agent is *not* subjectively certain about which state the world is in, but where he or she can nevertheless assign probabilities to the different possible states.

for any choice option, a , and for any state of the world, G , the measure of the choiceworthiness of a , in G , is the difference in utility, in G , between a and whichever alternative to a carries the greatest utility in G . This difference in utility is what I will call the *comparative utility* (CU) of a . For any agent, S , faced with any decision under risk, ECU theory counsels S to measure and rank his or her options (in terms of how choiceworthy they are) according to their *comparative expected comparative utility* ($CECU$) and to choose whichever option carries the greatest CECU. For any option, a , a 's CECU is the difference between its ECU and that of whichever alternative to a carries the greatest ECU, where a 's ECU is a probability-weighted sum of a 's CUs across the various possible states of the world. In this paper, I will show that in some ordinary decisions under risk, ECU theory gives different verdicts from those of EU theory and that EU theory therefore fails as a normative theory of rational choice under risk.

The idea of calculating differences between the utility of an option under consideration and the utilities of its alternatives in the choice situation—idea essentially similar to ECU theory—has been explored in the philosophical literature (Colyvan, 2008; Colyvan & Hájek, 2016)¹¹ and economic modeling literature (Zhang, Timmermans, Borgers & Wang, 2004; Zhang, 2015).

The Argument for ECU theory

This paper will argue for a new normative theory of rational choice under risk, namely ECU theory. The argument can be broken down into 15 steps, which are numbered below.

Let us begin with a preliminary argument (i.e., the *instrumental rationality argument*). First, note that, in what follows, I will use the words *choiceworthy* and *choiceworthiness* in a non-moral sense. Therefore, what is *choiceworthy* for S should be distinguished from what it is *morally good* or *morally right* for S to do, and *choiceworthiness* should be distinguished from *moral goodness* and *moral rightness*.¹²

1. For any agent, S , faced with any decision under certainty or any decision under risk and for any option, a , for S , a is *choiceworthy* for S if and only if a is worthy of being chosen by S in light of S 's rational preferences over each of the various possible states of the world, where S 's *rational preferences* are preferences that obey the series of rationality

¹¹ Colyvan (2008) has argued for a new decision theory that gives the right verdicts in decision problems where there are an infinite number of states with only finite utilities attached, such as the St-Petersburg game, and where EU theory gives no verdicts whatsoever. Colyvan's new theory, i.e., *relative expectation theory*, states that rational agents rank their choice options on the basis of their *relative expected utility*: for any agent, S , and for any two options, a and b , S prefers a to b if and only if the probability-weighted sum of the differences in utility between a and b for each possible state is positive, and S is indifferent between a and b if and only if the probability-weighted sum of the differences in utility between a and b for each possible state is zero. Relative expectation theory gives the same decision advice as EU theory in all decision cases where there are only a finite number of possible states and where the states are probabilistically independent of all choice options. See also Colyvan and Hájek (2016).

¹² *Moral rightness* arguably cannot be measured on a graded scale (Hurka, n.d.; Hurka, 2019; Olsen, 2018; Sinhababu, 2018).

conditions or axioms of standard decision theory. Moreover, the *degree* to which *a* is *choiceworthy* for *S*, or (i.e.) the *choiceworthiness* of *a* for *S*, is the degree to which *a* is worthy of being chosen by *S* in light of *S*'s rational preferences over each of the various possible states of the world.

2. *S*'s choosing *a* is *instrumentally rational* if and only if *S*'s choosing *a* is a suitable means to *S*'s ends (i.e., *S*'s rational preferences),¹³ and the *degree* to which *S*'s choosing *a* is *instrumentally rational* is the degree to which *S*'s choosing *a* is a suitable means to *S*'s ends (i.e., *S*'s rational preferences).
3. *a* is worthy of being chosen by *S* in light of *S*'s rational preferences over each of the various possible states of the world if and only if *S*'s choosing *a* is a suitable means to *S*'s ends (i.e., *S*'s rational preferences), and the degree to which *a* is worthy of being chosen by *S* in light of *S*'s rational preferences over each of the various possible states of the world is the degree to which *S*'s choosing *a* is a suitable means to *S*'s ends (i.e., *S*'s rational preferences).
4. Therefore, *a* is *choiceworthy* for *S* if and only if *S*'s choosing *a* is instrumentally rational, and the *choiceworthiness* of *a* for *S* is the *degree* to which *S*'s choosing *a* is instrumentally rational. (4 follows from 1 to 3.)

In what follows, I will assume that for any agent, *S*, and for any choice option, *a*, for *S*, *a*'s utility is a *cardinal* indicator of preference and is derived from *S*'s preferences as in standard decision theory, that is, via a *representation theorem*. This requires that *S*'s preferences obey a series of conditions or axioms of *rational preference*, one of which is the *Independence of irrelevant alternatives (IIA)* (for preferences): if an option, *a*, is preferred over some alternative option, *b*, then introducing a third option, *c*, in the choice situation will not change the preference ordering between *a* and *b*. For the present purposes, *rational preference* is analysed as satisfying the IIA. Note however that the IIA has been challenged (Sen, 1993; Wedgwood, 2013, pp. 2668–2670).

The argument for ECU theory proceeds as follows:

1. For any agent, *S*, and for any option, *a*, for *S*, *a* is choiceworthy for *S* if and only if *a* is *maximally choiceworthy* for *S* over the space of all alternatives in the choice set.
2. For any agent, *S*, and for any option, *a*, for *S*, *a* is *maximally choiceworthy* for *S* over the space of all alternatives in the choice set if and only if *a* *maximizes choiceworthiness* for *S* over the space of all alternatives in the choice set.
3. For any agent, *S*, and for any option, *a*, for *S*, *a* is choiceworthy for *S* if and only if *a* *maximizes choiceworthiness* for *S* over the space of all alternatives in the choice set (i.e., the *choiceworthiness maximization (CM) principle*). (3 follows from 1 and 2.)
4. For any agent, *S*, faced with any decision under *risk* and for any option, *a*, for *S*, the measure of the choiceworthiness of *a* for *S* is its CECU, that is, the difference between its

¹³ According to Kolodny & Brunero (2020), “Someone displays instrumental rationality insofar as she adopts suitable means to her ends.”

ECU and that of whichever alternative to a carries the greatest ECU (or one of them in the event that several alternatives are tied) (i.e., the *CECU principle*). a 's ECU is a probability-weighted sum of a 's CUs across the various states of the world, where, for any state of the world, G , a 's CU in G is the difference in utility, in G , between a and whichever alternative to a carries the greatest utility in G (or one of them in the event that several alternatives are tied).

5. For any agent, S , faced with any decision under *risk* and for any option, a , for S , a is choiceworthy for S if and only if a maximizes CECU. (5 follows from 3 and 4.)
6. For any agent, S , faced with any decision under *risk* and for any number of alternative options, a , b , c , d , and e , for S , it is rational for S to prefer a to b if and only if a 's EU is greater than b 's, it is rational for S to be indifferent between a and b if and only if a 's EU is equal to b 's, and the extent to which it is rational for S to prefer a to b is the difference in EU between a and b .
7. For any agent, S , faced with any decision under *risk* and for any option, a , for S , it is rational for S to (weakly¹⁴) prefer a over the alternative options in the choice set if and only if a maximizes EU. (7 follows from 6.)
8. In decisions under *risk*, what option(s) maximize CECU sometimes differ from what option(s) maximize EU.
9. In decisions under *risk*, what option(s) are choiceworthy sometimes differ from what option(s) it is rational to (weakly) prefer over the alternative options in the choice set. (9 follows from 5, 7 and 8.)
10. For any agent, S , faced with any decision under *risk*, it is a requirement of instrumental rationality that S should measure and rank his or her options in terms of how choiceworthy they are for S (i.e., how worthy of being chosen by S they are in light of S 's rational preferences over each of the various possible states of the world).
11. It is *not* the case that for any agent, S , faced with any decision under *risk*, it is a requirement of instrumental rationality that S should measure and rank his or her options in order of rational preference. (11 follows from 3, 9 and 10.)
12. For any agent, S , faced with any decision under *risk*, it is a requirement of instrumental rationality that S should measure and rank his or her options in terms of how choiceworthy they are for S , that is, according to their CECU, rather than in order of rational preference, that is, according to their EU. (12 follows from 4, 6, 8, 10 and 11.)
13. For any agent, S , faced with any decision under *risk*, it is a requirement of instrumental rationality that S should choose out of what option(s) are *choiceworthy* for S (i.e., what option(s) are worthy of being chosen by S in light of S 's rational preferences over each of the various possible states of the world).
14. It is *not* the case that for any agent, S , faced with any decision under *risk*, it is a requirement of instrumental rationality that S should choose out of what option(s) it is

¹⁴ For any agent, S , and for any two choice options, a and b , for S , if S weakly prefers a to b , then S either prefers a to b or is indifferent between a and b .

rational for S to (weakly) prefer over the alternative options in the choice set. (14 follows from 9 and 13.)

15. For any agent, S , faced with any decision under *risk* where what option(s) are choiceworthy for S differ from what option(s) it is rational for S to (weakly) prefer over the alternative options in the choice set, it is a requirement of instrumental rationality that S should choose out of what option(s) are choiceworthy for S (i.e., what option(s) maximize CECU), rather than out of what option(s) it is rational for S to (weakly) prefer over the alternative options in the choice set (i.e., what option(s) maximize EU). (15 follows from 5, 7, 9, 13 and 14, as well as from 3, 5, 7, 9 and 10.)

I will now discuss the different steps in the argument:

1. For any agent, S , and for any option, a , for S , a is choiceworthy for S if and only if a is *maximally choiceworthy* for S over the space of all alternatives in the choice set.

The question whether a given option is more (or less) choiceworthy than (or just as choiceworthy as) another option within a set of alternatives is well-formed and meaningful. Therefore, the question whether a given option is *maximally choiceworthy* within a set of alternatives is also well-formed and meaningful. I will assume that Step 1 is true without further argument.

2. For any agent, S , and for any option, a , for S , a is *maximally choiceworthy* for S over the space of all alternatives in the choice set if and only if a *maximizes choiceworthiness* for S over the space of all alternatives in the choice set.

For any number of alternative choice options, a , b , c , d , and e , we want to say that a (utility: 100) is more choiceworthy than b (utility: 5) even if a is not choiceworthy tout court (i.e., a does not maximize utility). We also want to say that the *extent* to which a is more choiceworthy than b is greater than the extent to which c (utility: 10) is more choiceworthy than b . In order to say that a is more choiceworthy than b (and to what extent), we cannot rely on a binary measure of choiceworthiness. Whether (and to what extent) a is more choiceworthy than b , and by implication, whether (and to what extent) any option is more choiceworthy than any other within a set of alternatives is necessarily a function of *how* choiceworthy each of the two options is within the set of alternatives (and not necessarily a function of one being choiceworthy tout court and the other unchoiceworthy tout court). To ask how choiceworthy an option is is to ask how desirable or worthy of being chosen that option is, how imperative it is to choose that option.¹⁵ Such a question is well-formed and meaningful. In order to answer the question, we

¹⁵ A qualification is in order: Let a and b denote two mutually exclusive and jointly exhaustive choice options. a is more choiceworthy than b if and only if choosing a is more imperative than choosing b , and a is just as choiceworthy as b if and only if choosing a is just as imperative as choosing b . However, if a is just as choiceworthy

require a graded, quantitative measure of how choiceworthy options are—i.e., we require a graded, quantitative measure of *choiceworthiness*.

3. For any agent, *S*, and for any option, *a*, for *S*, *a* is choiceworthy for *S* if and only if *a* maximizes choiceworthiness for *S* over the space of all alternatives in the choice set (i.e., the CM principle). (3 follows from 1 and 2.)
4. For any agent, *S*, faced with any decision under *risk* and for any option, *a*, for *S*, the measure of the choiceworthiness of *a* for *S* is its CECU, that is, the difference between its ECU and that of whichever alternative to *a* carries the greatest ECU (or one of them in the event that several alternatives are tied) (i.e., the *CECU principle*). *a*'s ECU is a probability-weighted sum of *a*'s CUs across the various states of the world, where, for any state of the world, *G*, *a*'s CU in *G* is the difference in utility, in *G*, between *a* and whichever alternative to *a* carries the greatest utility in *G* (or one of them in the event that several alternatives are tied).

4.1 The CU Principle

In order to establish the CECU principle, I first need to argue for a graded, quantitative measure of choiceworthiness for decisions under *certainty* (i.e. the *CU principle*). According to the CU principle, for any agent, *S*, faced with any decision under *certainty* and for any option, *a*, for *S*, the measure of the choiceworthiness of *a* for *S* is its comparative utility (CU). For any choice option, *a*, and for any state of the world, *G*, *a*'s CU in *G* is the difference in utility, in *G*, between *a* and whichever alternative to *a* carries the greatest utility in *G* (or one of them in the event that several alternatives are tied).¹⁶ In what follows, I will provide three arguments for the CU Principle:

To that end, I will assume that for any agent, *S*, faced with any decision under certainty and for any option, *a*, for *S*, *a* is *choiceworthy* for *S* if and only if *a* maximizes utility over the space of all alternatives in the state of the world to which *S* assigns probability 1. I will refer to this principle as the *utility maximization (UM) principle*. The UM principle defines a binary measure of choiceworthiness for decisions under certainty (i.e., whether an option is choiceworthy tout court or unchoiceworthy tout court).

as *b*, then both *a* and *b* are choiceworthy, whereas if choosing *a* is just as imperative as choosing *b*, then neither choosing *a* nor choosing *b* is imperative.

¹⁶ CU should be distinguished from the purely descriptive economic concept of *opportunity cost*. For any agent, *S*, let *a* be the highest-valued choice option available to *S*. The CU of *a*, for *S*, is the value of whatever *additional benefit* *S* would enjoy by choosing *a* over the highest-valued alternative to *a*. By contrast, the opportunity cost of *a*, for *S*, is the value of whatever *cost* *S* would incur by choosing *a* over the highest-valued alternative to *a*, where this includes the *total value* of the highest-valued alternative to *a* (Henderson, 2008).

4.1.1 Argument 1

It is generally accepted that for any agent, S , faced with any decision under certainty and for any number of alternative options, a , b , c , d , and e , available to S , the extent to which a is more choiceworthy than b , for S , is the extent to which S (rationally) prefers a to b , or equivalently the extent to which S (rationally) prefers a to b more than S (rationally) prefers b to a . However, intuitively, that is a mistake. Even though we are comparing a to b , we want to see how a and b measure up to the *very best alternative options on offer*, in the following way: the extent to which a is more choiceworthy than b , for S , is the extent to which S (rationally) prefers a to the most (rationally) preferred alternative to a (either b , c , d , or e) more than S (rationally) prefers b to the most (rationally) preferred alternative to b (either a , c , d , or e). After all, if S must choose an alternative to a , then S should choose the most (rationally) preferred alternative to a (either b , c , d , or e), and not necessarily the option to which S is comparing a (i.e., option b). The same goes for option b .

Therefore, the extent to which a is more choiceworthy than b , for S , is the extent to which {the difference in utility between a and whichever alternative to a carries the greatest utility} is greater than {the difference in utility between b and whichever alternative to b carries the greatest utility}. It follows that the extent to which a is choiceworthy for S (or [i.e.] the measure of *how* choiceworthy a is for S) is a 's CU , i.e., the difference in utility between a and whichever alternative to a carries the greatest utility (or one of them in the event that several alternatives are tied) (i.e., the CU principle). The same goes for option b . (Henceforth, c -utils are defined as units of CU .)

An alternative approach is to say that the extent to which a is more choiceworthy than b , for S , is the extent to which S (rationally) prefers a to the most (rationally) preferred option (either a , b , c , d , or e) more than S (rationally) prefers b to the most (rationally) preferred option (either a , b , c , d , or e). In other words, the extent to which a is more choiceworthy than b , for S , is the extent to which {the difference in utility between a and whichever option carries the greatest utility (either a , b , c , d , or e)} is greater than {the difference in utility between b and whichever option carries the greatest utility (either a , b , c , d , or e)}. If that is the case, then the extent to which a is choiceworthy for S (or [i.e.] the measure of *how* choiceworthy a is for S) is a 's CU^* , i.e., the difference in utility between a and whichever option carries the greatest utility (either a , b , c , d , or e) (or one of them in the event that several alternatives are tied) (henceforth, the CU^* principle).¹⁷ The same goes for option b .

The CU^* principle is however untenable since it results in a double standard. It entails that the degrees of choiceworthiness of all the option(s) that do *not* carry the greatest utility

¹⁷ For any choice option, a , and for any state of the world, G , a 's CU^* in G is the difference in utility, in G , between a and whichever option carries the greatest utility in G (either a , b , c , d , or e) (or one of them in the event that several alternatives are tied). The rule of maximizing *expected* CU^* (or ECU^*) counsels agents to choose whichever option in the choice set has the greatest ECU^* —i.e., a probability-weighted sum of an option's CU^* s across the

depend on what other options are available in the choice set—those degrees of choiceworthiness may be different negative numbers, but never 0—whereas the degrees of choiceworthiness of all the option(s) that *do* carry the greatest utility do *not* depend on what other options are available in the choice set—those degrees of choiceworthiness are 0 no matter what the utilities of the other options are. Moreover, the latter standard is implausible. It's as if the degrees of choiceworthiness of all the option(s) that do *not* carry the greatest utility did *not* depend on what other options are available in the choice set—it's as if those degrees of choiceworthiness were the same negative number, e.g., -1 , no matter what the utilities of the other options are. Contrary to the CU^* principle, the original CU principle does not suffer from these problems.

Let us now consider four choice situations involving decisions under certainty:

Table 1

	1	2	3	4
<i>a</i>	5	5	5	5
<i>b</i>	1	1	1	1
<i>c</i>	1	2	2	2
<i>d</i>	1	3	3	3
<i>e</i>	1	3	5	8

Compared to the *difference in utility* and *difference in CU^** , the *difference in CU* is a more plausible measure of the extent to which *a* is more choiceworthy than *b* in situations 1–4, as explained above. The differences in utility and CU^* between *a* and *b* are the same in all four situations (4 units), whereas the differences in CU between *a* and *b* are as follows (in situations 1–4):

1. $(5 - 1) - (1 - 5) = 8$ c-utiles
2. $(5 - 3) - (1 - 5) = 6$ c-utiles
3. $(5 - 5) - (1 - 5) = 4$ c-utiles
4. $(5 - 8) - (1 - 8) = 4$ c-utiles

The CU principle is therefore well-supported.

various states of the world. The rule of maximizing ECU^* is equivalent to the rule of maximizing EU (i.e., EU theory), which means that both rules deliver the same verdicts in all decision cases.

4.1.2 Argument 2

1. For any agent, S , faced with any decision under certainty and for any option, a , for S , a is choiceworthy for S if and only if a is worthy of being chosen by S over whichever alternative to a is the most choiceworthy for S . (True by definition)
2. a is choiceworthy for S if and only if a maximizes choiceworthiness for S over the space of all alternatives in the choice set (i.e., the CM principle). (Assumption)
3. The extent to which a is choiceworthy for S (or [i.e.] the measure of *how* choiceworthy a is for S) is the extent to which a is worthy of being chosen by S over whichever alternative to a is the most choiceworthy for S . (3 follows from 1 and 2.)
4. a is choiceworthy for S if and only if a maximizes utility over the space of all alternatives in the choice set (i.e., the UM principle). (Assumption)
5. a is the most choiceworthy option available to S in the choice set if and only if a carries the greatest utility in the choice set. (5 follows from 2 and 4.)
6. a is the most choiceworthy option available to S in a restricted choice set if and only if a carries the greatest utility in that restricted choice set. (6 follows from 5.)
7. Whichever alternative to a is the most choiceworthy for S is whichever alternative to a carries the greatest utility. (7 follows from 6.)
8. The extent to which a is worthy of being chosen by S over some alternative to a is the difference in utility between a and that alternative to a . (True by conceptual analysis)
9. Therefore, the extent to which a is choiceworthy for S (or [i.e.] the measure of *how* choiceworthy a is for S) is the difference in utility between a and whichever alternative to a carries the greatest utility (or one of them in the event that several alternatives are tied) (i.e., the *CU principle*). (9 follows from 3, 7 and 8.)

4.1.3 Argument 3

Let us now consider a much longer argument. The simplest attempt at defining a graded, quantitative choiceworthiness measure for decisions under certainty is as follows: for any agent, S , faced with any decision under certainty and for any option, a , for S , the measure of the choiceworthiness of a for S is the utility of a in the state of the world to which S assigns probability 1. I will refer to this as the *utility principle*. The UM principle is true if (*but not* only if) the utility principle is true. The utility principle is, however, untenable.

First, *measures* of quantities, for example 20°C for temperature, are meaningful (and *only* meaningful) relative to a given zero point and unit of measurement. (Let us call this the *measurement principle*.) In the case of temperature, the measure (e.g., 20°C) is defined in relation to the zero point and unit of measurement (i.e., the measure itself presupposes a given temperature unit and zero point of temperature). That is not the case for utility. In accordance with the measurement principle, the measure of a 's utility (e.g., 20 units of utility [or *utiles*]) is meaningful (and *only* meaningful) relative to a given utility unit and zero point of utility.

However, the measure (e.g., 20 units of utility) is *not* defined in relation to the unit and zero point (i.e., the utility measure itself does *not* presuppose a given utility unit and zero point of utility).¹⁸ These values must be explicitly specified. Hence, the utility principle is at best underspecified.

Second, even *relative to an explicitly given utility unit and zero point of utility*, the measure of the choiceworthiness of a for S is not necessarily its utility. In accordance with the measurement principle, for any given decision situation (under certainty) and for any specified utility unit and zero point of utility (for that situation), the measure of the choiceworthiness of any available option is its utility value if and only if it is possible to ascertain how choiceworthy any available option is (in that situation) by solely considering its utility value in relation to that specified utility unit and zero point of utility. In practical terms, what this means is that, for any given *decision setup* (i.e., any decision situation combined with any explicit specification of a utility unit and zero point of utility), the measure of the choiceworthiness of any available option is its utility value if and only if (i) any available option is choiceworthy just in case its utility value is equal to or greater than zero (and not choiceworthy otherwise) and (ii) the degree of choiceworthiness of any available option is its utility value. Now, it is straightforward to come up with decision situations where it is possible to select a specific zero point of utility (and a specific utility unit) such that it is *not* the case that any available option is choiceworthy (in that situation) if and only if its utility value is equal to or greater than zero. Per the UM principle, there are possible decision setups where an option has a positive utility value and is nevertheless unchoiceworthy, namely setups where that option does *not* maximize utility over the space of all available alternatives, and there are possible decision setups where an option has a negative utility value and is nevertheless choiceworthy, namely setups where that option *does* maximize utility over the space of all available alternatives. Therefore, per the measurement principle, there are possible decision setups such that it is *not* the case that the measure of the choiceworthiness of any available option (in that setup) is its utility value.

In light of the preceding considerations and in accordance with the measurement principle, it is necessarily the case that for any agent, S , faced with *any* decision situation under certainty and for any option, a , for S , the measure of the choiceworthiness of a for S depends on a unit of measurement of choiceworthiness as well as a zero point of choiceworthiness (or *benchmark*) in the following way: the measure of the choiceworthiness of a for S (relative to *any* explicitly given utility unit and zero point of utility) is the *difference in utility* between a and some *benchmark* for a , such that (i) a is choiceworthy for S if and only if the difference in utility between a and the benchmark for a is equal to or greater than zero (and not choiceworthy

¹⁸ “[S]ince the utilities of options, whether ordinal or interval-valued, can only be determined relative to the utilities of other options, there is no such thing as the absolute utility of an option, at least not without further assumptions. The further assumptions would need to relate particular options to particular privileged levels of utility; for instance, one would need to argue that a rational agent’s preference ordering should incorporate, say, a privileged zero-utility option, in which case ratios of utility distances from this option would be meaningful.” (Steele & Stefánsson, 2020) “The zero point and the unit in an expected utility representation are arbitrary; utility values become meaningful only once they have been fixed.” (Colyvan & Hájek, 2016, pp. 838–39)

otherwise), and (ii) the degree of choiceworthiness of a for S is the difference in utility between a and the benchmark for a . In other words, the measure of the choiceworthiness of a for S is the degree to which a is worthy of being chosen over the benchmark for a . The benchmark for a can be, for example, some option in the set of available options, such as whichever option has the highest utility, whichever option has the lowest utility, or the status quo, or some average of the utilities of the available options. Choiceworthiness is thus a relative concept.¹⁹ As will become clear in what follows, the concept of choiceworthiness itself presupposes a given benchmark (or zero point of choiceworthiness).

If there are any alternatives to a which carry a greater utility than does a , then the benchmark for a is whichever alternative to a carries the greatest utility. Indeed, if there are any alternatives to a with a greater utility than a , then, in accordance with the UM principle, a is not choiceworthy for S . But if a is not choiceworthy for S , then how choiceworthy a is for S is simply how a compares to whichever alternative is choiceworthy for S (or, per the UM principle, whichever alternative to a carries the greatest utility). I will now argue that if there are *not* any alternatives to a which carry a greater utility than does a , then the benchmark for a still has to be whichever alternative to a carries the greatest utility. As far as I know, this idea has not been explored by others in the published literature.

Let us consider two decision situations (or setups): 1 and 2. In each situation, S is faced with the same three options: a , b , and c . What's more, in each situation, S assigns probability 1 to a given state of the world (but not the same state for both situations). If that state of the world is realized, then S assigns the following utilities to the set of options:

Table 2

	1	2
a	100	100
b	-10,000	99
c	-10,000	99

Per the UM principle, a is choiceworthy for S in both situations 1 and 2. a is also more choiceworthy for S in 1 than in 2—that is to say, it is more imperative for S to choose a if S is in situation 1 than if S is in situation 2. In 2, S misses out on only 1 utile by not choosing a , but instead choosing the best alternative to a (i.e., b or c), whereas in 1, S misses out on 10,100 utiles by not choosing a , but instead choosing the best alternative to a (i.e., b or c). Another way of putting it is that a is more choiceworthy in 1 than in 2 because a is more worthy of being chosen over the best alternative to a in 1 than in 2.

¹⁹ Ralph Wedgwood (2017) relies on considerations of incommensurability to argue for the same idea: “the choiceworthiness of options is relative to choice situations”. Larry S. Temkin (2012) also addresses this idea: what he calls the “Essentially Comparative View.”

Let us now briefly introduce Ralph Wedgwood’s *benchmark theory (BT)* (Wedgwood, 2013). The basic idea of BT is to rank choice options (in terms of how choiceworthy they are) according to their *expected comparative value*, where the comparative value of an option is its *value* (broadly construed) in some state of the world compared to a benchmark for that state of the world. Wedgwood identifies the benchmark as an average of the options’ values within a given state of the world. He emphasizes that all statewise dominated options and more generally, “all the options that *do not deserve to be taken seriously*” (2664) should be excluded from consideration at the outset.²⁰ Wedgwood explicitly rejects the idea that the value of an option is its utility. Nevertheless, it is interesting to see how BT (henceforth, *BT**) fairs when the value of an option is understood to be its utility.

Coming back to our example, we can see that *BT** agrees with the verdict that *a* is choiceworthy for *S* in situations 1 and 2, but *not* with the verdict that *a* is more choiceworthy for *S* in 1 than in 2. According to *BT**, *a* is equally choiceworthy for *S* in situations 1 and 2 since *b* and *c* are strictly dominated by *a* in both 1 and 2 and are therefore excluded from consideration at the outset. If *b* and *c* are *not* excluded from consideration and the benchmark is identified as an average of the values (or utilities) of all the available options, then this alternative approach agrees with our verdict: *a* is more choiceworthy for *S* in 1 than in 2.

Here is a different example:

Table 3

	1	2
<i>a</i>	100	100
<i>b</i>	−100	−500
<i>c</i>	−100	100

Per the UM principle, *a* is choiceworthy for *S* in both situations 1 and 2. *a* is also more choiceworthy for *S* in 1 than in 2—that is to say, it is more imperative for *S* to choose *a* if *S* is in situation 1 than if *S* is in situation 2. In 2, *a* is merely optional—*S* misses out on *zero* utiles by not choosing *a*, but instead choosing the best alternative to *a* (i.e., *c*)—whereas in 1, *a* is *not* optional—*S* misses out on 200 utiles by not choosing *a*, but instead choosing the best alternative to *a* (i.e., *b* or *c*). Again, *a* is more choiceworthy in 1 than in 2 because *a* is more worthy of being chosen over the best alternative to *a* in 1 than in 2.

*BT** agrees with both verdicts: *a* is choiceworthy for *S* in 1 and 2, and *a* is more choiceworthy for *S* in 1 than in 2. However, if the benchmark is defined as an average of the values (or utilities) of all the available options (whether strictly dominated or not), then this alternative approach does *not* agree with our verdict: *a* is more choiceworthy for *S* in 1 than in 2. The two examples just laid out, when taken together, make for an effective counterexample to *BT**.

²⁰ For critiques of BT, see Bassett (2015) and Briggs (2010).

Another very similar example:

Table 4

	1	2
<i>a</i>	100	100
<i>b</i>	-100	-500
<i>c</i>	-100	99

Per the UM principle, *a* is choiceworthy for *S* in both situations 1 and 2. *a* is also more choiceworthy for *S* in 1 than in 2—that is to say, it is more imperative for *S* to choose *a* if *S* is in situation 1 than if *S* is in situation 2. In 2, *S* misses out on only 1 utile by not choosing *a*, but instead choosing the best alternative to *a* (i.e., *c*), whereas in 1, *S* misses out on 200 utiles by not choosing *a*, but instead choosing the best alternative to *a* (i.e., *b* or *c*). Once again, *a* is more choiceworthy in 1 than in 2 because *a* is more worthy of being chosen over the best alternative to *a* in 1 than in 2.

BT* agrees with the verdict that *a* is choiceworthy for *S* in 1 and 2, but *not* with the verdict that *a* is more choiceworthy for *S* in 1 than in 2. According to BT*, *a* is equally choiceworthy for *S* in situations 1 and 2 since *b* and *c* are strictly dominated by *a* in both 1 and 2 and are therefore excluded from consideration at the outset. If *b* and *c* are *not* excluded from consideration and the benchmark is identified as an average of the values (or utilities) of all the options, then *a* is more choiceworthy for *S* in 2 than in 1. I take this to be a further counterexample to BT*.

One final example:

Table 5

	1	2
<i>a</i>	100	100
<i>b</i>	-100	99
<i>c</i>	100	100

Per the UM principle, *a* is choiceworthy for *S* in both situations 1 and 2. *a* is also equally choiceworthy for *S* in both situations—that is to say, it is just as imperative for *S* to choose *a* if *S* is in situation 1 as it is if *S* is in situation 2. In both situations, *a* is merely optional—*S* misses out on *zero* utiles by not choosing *a*, but instead choosing the best alternative to *a* (i.e., *c*). To put it another way, *a* is just as choiceworthy in 1 as it is in 2 because *a* is just as worthy of being chosen over the best alternative to *a* in 1 as it is in 2. (BT* agrees with both verdicts.)

These four examples serve to illustrate that if there are *not* any alternatives to *a* with a greater utility than *a*, then how choiceworthy *a* is depends on how much utility *S* would miss out

on by not choosing a , but instead choosing the best alternative to a . The greater the amount of utility S would miss out on by not choosing a , but instead choosing the best alternative to a , the more choiceworthy a becomes. Thus, the benchmark for a must be whichever alternative to a carries the highest utility.

What follows is that whether or not there are any alternatives to a which carry a greater utility than does a , the benchmark for a has to be whichever alternative to a carries the greatest utility. This means that there is no unique benchmark for a given choice situation. Instead, the benchmark is relative to a specific choice option. The benchmark for a may be some alternative, b , and the benchmark for b may be a .

Therefore, for any agent, S , faced with any decision under certainty and for any option, a , for S , the measure of the choiceworthiness of a for S (relative to any explicitly given utility unit and zero point of utility) is the *CU* of a (in the state of the world to which S assigns probability 1). The *CU* of a is the difference in utility between a and whichever alternative to a carries the greatest utility (or one of them in the event that several alternatives are tied). As previously indicated, I will refer to this principle as the *CU principle*. Like the utility principle, the *CU* principle entails the *UM* principle.

In light of the *CU* principle, the utility principle can be falsified. If the utility principle were true, then in accordance with the measurement principle, it would be the case that for *any* given decision situation, there is at least one specification of a utility unit and zero point of utility such that it is possible to ascertain how choiceworthy any available option is (for S) by solely considering its utility value in relation to that specification of a utility unit and zero point of utility. In other words, it would be the case that for *any* given decision situation, there is at least one specification of a utility unit and zero point of utility such that (i) any available option is choiceworthy (for S) if and only if its utility value is equal to or greater than zero (and not choiceworthy otherwise) and (ii) *the degree of choiceworthiness of any available option (for S) is its utility value*. As we will now see, that is not the case. Let us consider the following decision setup: S is faced with three options: a , b , and c . What's more, S assigns probability 1 to a given state of the world. If that state of the world is realized, then S assigns the following utilities to the available options: a (0), b (-100), c (-1,000). Therefore, no matter what zero point of utility is selected, S assigns the following utility intervals between the available options: between a and b , S assigns a positive interval of 100 utiles, between b and c , S assigns a positive interval of 900 utiles and between a and c , S assigns a positive interval of 1,000 utiles. Per the *CU* principle, the degrees of choiceworthiness of the available options are as follows: a (100), b (-100), c (-1,000). Therefore, the differences between the degrees of choiceworthiness of the available options are as follows: between a and b , the difference is 200 c-utiles, between b and c , the difference is 900 c-utiles and between a and c , the difference is 1,100 c-utiles. Since the utility intervals and the differences in degrees of choiceworthiness are at variance, we have a decision situation where no matter what zero point of utility (and what utility unit) is selected, it is *not* the case that the degree of choiceworthiness of any available option is its utility value.

The utility principle, let us recall, states that for any agent, S , faced with any decision under certainty and for any option, a , for S , the measure of the choiceworthiness of a for S is the utility of a in the state of the world to which S assigns probability 1. Since the utility principle is false and since the expected utility of a equals the utility of a in the state of the world to which S assigns probability 1, it follows that for any agent, S , faced with any decision under certainty or any decision under risk and for any option, a , for S , the measure of the choiceworthiness of a for S *cannot* be the expected utility of a .

4.2 The CECU Principle

As I argued in discussing Step 2, we require a graded, quantitative measure of how choiceworthy options are. When we move from decision-making under certainty to decision-making under risk, we can, in light of the CU principle, identify the measure of an option's choiceworthiness as expressing that option's *expected choiceworthiness*, or *ECU*, that is to say, the *expected value*, or the probability-weighted sum of all possible values, of that option's choiceworthiness, or CU, *in the actual state of the world*. That roughly encapsulates ECU theory.

As a first approximation, then, ECU theory says that for any agent, S , and for any choice option, a , for S , the measure of the choiceworthiness of a for S (relative to any explicitly given utility unit and zero point of utility) is the ECU of a . The *ECU* of an option, a , in a decision problem with n states is formally defined as:

$$\text{ECU}(a) = \sum_{i=1}^n (U(a, s_i) - U(\text{bm}(a), s_i))P(s_i)$$

where $U(a, s_i)$ denotes the utility of option a when state s_i is actual, $U(\text{bm}(a), s_i)$ denotes the utility of the benchmark for a when state s_i is actual (i.e., the utility in state s_i of whichever alternative(s) to a have the highest utility in state s_i), and $P(s_i)$ denotes the probability assigned to state s_i . In other words, for any number of alternative options, a, b, c, d , and e , one calculates the ECU of a as follows: for each state of the world, one subtracts a 's utility from the utility of b, c, d , or e , whichever of b, c, d , and e maximizes utility in that state, and one multiplies the result by the probability that one assigns to that state; finally, one sums the totals for every state.

The CU principle is straightforwardly entailed by ECU theory. Furthermore, ECU theory presupposes that the states of the world in any decision problem are probabilistically independent of the agent's choices.

ECU theory, as formulated above, is not quite right though. In accordance with the measurement principle, if the measure of the choiceworthiness of options is their ECU, then only options with ECU equal to or greater than zero can be choiceworthy. However, as I will illustrate in my discussion of Step 8, there will always be cases (regardless of what utility unit and zero point of utility are specified) where every option in a decision situation *under risk* has negative ECU. Since at least one option in a decision situation must be choiceworthy—the one with the

highest degree of choiceworthiness (or one of them in the event that several alternatives are tied) (i.e., the *CM principle*)—ECU theory, as defined above, is false in decision cases *under risk*.

By the same lines of reasoning as employed in my discussion of Step 4 (Section 4.1.3), we reach the following conclusion: for any agent, *S*, faced with any decision *under risk* and for any choice option, *a*, for *S*, the measure of the choiceworthiness of *a* for *S* (relative to any explicitly given utility unit and zero point of utility) is the *comparative expected choiceworthiness*, or *comparative expected comparative utility (CECU)*, of *a*, that is to say, the difference in ECU between *a* and whichever alternative to *a* carries the greatest ECU (or one of them in the event that several alternatives are tied). Let us call this principle the *CECU principle*. For any two alternative options, *a* and *b*, *a*'s CECU is greater than *b*'s if and only if *a*'s ECU is greater than *b*'s, and *a*'s CECU is equal to *b*'s if and only if *a*'s ECU is equal to *b*'s. We are now in a position to precisely define ECU theory: ECU theory is the conjunction of the CU principle (for decisions under certainty) and the CECU principle (for decisions under risk).

To demonstrate how to apply this new decision rule (i.e., ECU theory) to a concrete decision problem, let us consider the following case: An agent, *S*, is faced with a choice between two independent options or gambles: one option, *a*, offering a 0.01 probability of winning a prize worth 1,500 utiles (and nothing otherwise), and one option, *b*, offering a 0.02 probability of winning a prize worth 700 utiles (and nothing otherwise). According to ECU theory, *S* should choose option *a*, since its CECU is equal to or greater than zero ($1 - [-1] = 2$).

The ECUs of options *a* and *b* are given by the two equations below. The following notation is used: *A* denotes the state “If *S* chooses *a*, then *S* will not win the prize (0 utiles) and if *S* chooses *b*, then *S* will not win the prize (0 utiles)” (probability: $0.99 \times 0.98 = 0.9702$), *B* denotes the state “If *S* chooses *a*, then *S* will not win the prize (0 utiles) and if *S* chooses *b*, then *S* will win the prize (700 utiles)” (probability: $0.99 \times 0.02 = 0.0198$), *C* denotes the state “If *S* chooses *a*, then *S* will win the prize (1,500 utiles) and if *S* chooses *b*, then *S* will not win the prize (0 utiles)” (probability: $0.01 \times 0.98 = 0.0098$), *D* denotes the state “If *S* chooses *a*, then *S* will win the prize (1,500 utiles) and if *S* chooses *b*, then *S* will win the prize (700 utiles)” (probability: $0.01 \times 0.02 = 0.0002$), $P(A)$ denotes the probability of state *A*, and $U(a, A)$ denotes the utility of option *a* when state *A* is actual.

Table 6

	A (0.9702)	B (0.0198)	C (0.0098)	D (0.0002)
<i>a</i>	0	0	1,500	1,500
<i>b</i>	0	700	0	700

$$\text{ECU}(a) = (U(a, A) - U(b, A)) \times P(A) + (U(a, B) - U(b, B)) \times P(B) + (U(a, C) - U(b, C)) \times P(C) + (U(a, D) - U(b, D)) \times P(D) = 1 \text{ c-utile}$$

$$\text{ECU}(b) = (U(b, A) - U(a, A)) \times P(A) + (U(b, B) - U(a, B)) \times P(B) + (U(b, C) - U(a, C)) \times P(C) + (U(b, D) - U(a, D)) \times P(D) = -1 \text{ c-utiles}$$

One line of argument in support of ECU theory is that, contrary to EU theory, ECU theory agrees with and entails Wedgwood's *Gandalf's principle*: the choiceworthiness of an option *in a given state of the world* should be measured only relative to the values of the other options *in that state*, and not to the values of the options in other states. According to Wedgwood (2013, p. 2654),

to make a rational choice in [cases involving risk], one *does not need to consider* whether one is in a nice state of nature or a nasty one. All that one needs to consider are the *degrees* to which each of the available options is better (or worse) than the available alternatives *within* each of the relevant states of nature. Admittedly, when one is uncertain which state of nature one is in, one must make *some* comparisons across the states of nature. But since one does not even need to know whether one is in a nice state of nature or a nasty one, it seems that the only relevant comparisons are comparisons of the *differences* in levels of goodness between the various options *within* each state of nature with the *differences* between those options within each of the other states of nature – not any comparisons of *absolute* levels of goodness across different states of nature.

Although Wedgwood uses terms such as “better,” “worse,” and “levels of goodness” in his explication of Gandalf's principle, the principle can be expressed equally well using replacement terms such as “preferred,” “dispreferred,” and “levels of utility.”

Gandalf's principle is an eminently reasonable principle.²¹ In a paper critiquing Wedgwood's BT, Robert Bassett (2015) concurs: “Gandalf's principle strikes me as an eminently sensible principle to incorporate into rational decision-making.” There is, however, one alternative decision theory which agrees with and entails both the CU principle and Gandalf's principle and which has some *prima facie* plausibility—*maximum likelihood comparative utility (MLCU) theory*: for any agent, *S*, and for any option, *a*, for *S*, the measure of the choiceworthiness of *a* for *S* (relative to any explicitly given utility unit and zero point of utility) is the *most likely value* of *a*'s choiceworthiness (or CU) in the actual state of the world, and in cases where there is more than one maximally likely value of *a*'s choiceworthiness (or CU) in the actual state of the world, the measure of the choiceworthiness of *a* for *S* (relative to any explicitly given utility unit and zero point of utility) is *a*'s CECU across the maximally likely states of the world. We require a further argument to rule out MLCU theory.

This brings me to the following decision case: Let us suppose that an agent, *S*, is faced with three choice options: *a*, *b*, and *c*. *S* assigns probability 0.51 to a state of the world, *A*, and 0.49 to a state of the world, *B*. If state *A* or state *B* is realized, then *S* assigns the following utilities to the set of options:

²¹ For an illuminating discussion of Gandalf's principle, see Wedgwood (2013), pp. 2652–55.

Table 7

	A (0.51)	B (0.49)
<i>a</i>	110	-1,000
<i>b</i>	80	110
<i>c</i>	100	100

According to MLCU theory, *a* is uniquely choiceworthy for *S*, since state *A* is more likely to obtain than state *B* and the CU of option *a* in state *A* is greater than that of any other available option. Yet, it is clear that choosing option *a* is a mistake, since state *B* is almost as likely to obtain as state *A* and the comparative *disutility* of option *a* in state *B* is very high (-1,110 c-utiles). I take this to be an effective counterexample to MLCU theory.

5. For any agent, *S*, faced with any decision under *risk* and for any option, *a*, for *S*, *a* is choiceworthy for *S* if and only if *a* maximizes CECU. (5 follows from 3 and 4.)
6. For any agent, *S*, faced with any decision under *risk* and for any number of alternative options, *a*, *b*, *c*, *d*, and *e*, for *S*, it is rational for *S* to prefer *a* to *b* if and only if *a*'s EU is greater than *b*'s, it is rational for *S* to be indifferent between *a* and *b* if and only if *a*'s EU is equal to *b*'s, and the extent to which it is rational for *S* to prefer *a* to *b* is the difference in EU between *a* and *b*.

Decision-theoretic representation theorems—such as those of Von Neumann and Morgenstern (1947), Savage (1954), and Bolker (1966) and Jeffrey (1983)—show that if an agent fails to prefer choice options with higher EU, then that agent violates at least one of a series of axioms of *rational preference*,²² one of which is the IIA.

7. For any agent, *S*, faced with any decision under *risk* and for any option, *a*, for *S*, it is rational for *S* to (weakly) prefer *a* over the alternative options in the choice set if and only if *a* maximizes EU. (7 follows from 6.)
8. In decisions under *risk*, what option(s) maximize CECU sometimes differ from what option(s) maximize EU.

Let a *finite decision* be a decision problem where there are only finitely many states and no infinite utilities. In all finite decisions under risk requiring a choice between *only two* alternative options, ECU theory delivers the same verdicts as EU theory. However, in some finite decisions under risk requiring a choice between *more than two* alternative options, ECU theory

²² Following the formulation of Briggs (2019).

gives different verdicts from those of EU theory.²³ Let us consider the following example: an agent, S , is faced with five choice options: a , b , c , d , and e . S assigns probability 0.5 to a state of the world, A , and 0.5 to a state of the world, B . If state A or state B is realized, then S assigns the following utilities to the set of options:

Table 8

	A (0.5)	B (0.5)
a	2	10
b	5	8
c	6	6
d	8	4
e	10	2

According to EU theory, b is more choiceworthy than a , for S , since the EU of b (6.5 utiles) is greater than that of a (6 utiles). In fact, according to EU theory, b is choiceworthy tout court since its EU is greater than that of every other option.

$$EU(a) = (U(a, A) \times P(A)) + (U(a, B) \times P(B)) = 6 \text{ utiles}$$

$$EU(b) = (U(b, A) \times P(A)) + (U(b, B) \times P(B)) = 6.5 \text{ utiles}$$

$$EU(c) = (U(c, A) \times P(A)) + (U(c, B) \times P(B)) = 6 \text{ utiles}$$

$$EU(d) = (U(d, A) \times P(A)) + (U(d, B) \times P(B)) = 6 \text{ utiles}$$

$$EU(e) = (U(e, A) \times P(A)) + (U(e, B) \times P(B)) = 6 \text{ utiles}$$

By contrast, according to ECU theory, a is more choiceworthy than b , for S , since the ECU of a (-3 c-utiles) is greater than that of b (-3.5 c-utiles). In fact, according to ECU theory, a is choiceworthy tout court, since its CECU is equal to or greater than zero ($[-3] - [-3] = 0$).

$$ECU(a) = ((U(a, A) - U(e, A)) \times P(A)) + ((U(a, B) - U(b, B)) \times P(B)) = -3 \text{ c-utiles}$$

$$ECU(b) = ((U(b, A) - U(e, A)) \times P(A)) + ((U(b, B) - U(a, B)) \times P(B)) = -3.5 \text{ c-utiles}$$

$$ECU(c) = ((U(c, A) - U(e, A)) \times P(A)) + ((U(c, B) - U(a, B)) \times P(B)) = -4 \text{ c-utiles}$$

$$ECU(d) = ((U(d, A) - U(e, A)) \times P(A)) + ((U(d, B) - U(a, B)) \times P(B)) = -4 \text{ c-utiles}$$

$$ECU(e) = ((U(e, A) - U(d, A)) \times P(A)) + ((U(e, B) - U(a, B)) \times P(B)) = -3 \text{ c-utiles}$$

ECU theory gives different verdicts from those of EU theory because ECU theory, contrary to EU theory, violates the IIA (for choiceworthiness evaluations). According to this principle, for any decision situation, T , and for any choice option, a , in T , if a is choiceworthy in

²³ In a number of decision cases where there are infinitely many states with only finite utilities attached (e.g., the St. Petersburg game), ECU theory inherits the advantages of Mark Colyvan's relative expectation theory over EU theory. More specifically, in such (infinite) decision cases, ECU theory delivers the intuitively correct verdicts, whereas EU theory delivers none (Colyvan, 2008; Colyvan and Hájek, 2016, pp. 838–39).

T , then a is also choiceworthy in T if some other option(s) are eliminated from the pool of options in T . Likewise, if a is not choiceworthy in T , then a is also not choiceworthy in T if some other option(s) are added to the pool of options in T . Let us consider again the previous decision situation. In that situation, ECU theory dictates that a is choiceworthy. However, if options c , d , and e are eliminated from the pool of options, then b is choiceworthy according to ECU theory, as shown below:

Table 9

	A (0.5)	B (0.5)
a	2	10
b	5	8

b is choiceworthy tout court since its CECU is equal to or greater than zero ($[0.5] - [-0.5] = 1$).

$$\text{ECU}(a) = ((U(a, A) - U(b, A)) \times P(A)) + ((U(a, B) - U(b, B)) \times P(B)) = -0.5 \text{ c-utiles}$$

$$\text{ECU}(b) = ((U(b, A) - U(a, A)) \times P(A)) + ((U(b, B) - U(a, B)) \times P(B)) = 0.5 \text{ c-utiles}$$

Here is another example where ECU theory violates the IIA:²⁴ An agent, S , is faced with two choice options: a and b . S assigns probability 0.001 to a state of the world, A , and 0.999 to a state of the world, B . If state A or state B is realized, then S assigns the following utilities to the set of options:

Table 10

	A (0.001)	B (0.999)
a	1000	1
b	0	2

According to ECU theory, a is choiceworthy tout court, since its CECU is equal to or greater than zero ($[0.001] - [-0.001] = 0.002$).

$$\text{ECU}(a) = ((U(a, A) - U(b, A)) \times P(A)) + ((U(a, B) - U(b, B)) \times P(B)) = 0.001 \text{ c-utiles}$$

$$\text{ECU}(b) = ((U(b, A) - U(a, A)) \times P(A)) + ((U(b, B) - U(a, B)) \times P(B)) = -0.001 \text{ c-utiles}$$

Let us now introduce a third choice option (c) in the decision situation, all else being the same:

²⁴ Thanks to an anonymous reviewer for giving this example.

Table 11

	A (0.001)	B (0.999)
<i>a</i>	1000	1
<i>b</i>	0	2
<i>c</i>	900	0

In this new decision situation, *b* is choiceworthy tout court according to ECU theory, since *b*'s CECU is equal to or greater than zero ($[-0.001] - [-0.899] = 0.898$).

$$\text{ECU}(a) = ((U(a, A) - U(c, A)) \times P(A)) + ((U(a, B) - U(b, B)) \times P(B)) = -0.899 \text{ c-utiles}$$

$$\text{ECU}(b) = ((U(b, A) - U(a, A)) \times P(A)) + ((U(b, B) - U(a, B)) \times P(B)) = -0.001 \text{ c-utiles}$$

$$\text{ECU}(c) = ((U(c, A) - U(a, A)) \times P(A)) + ((U(c, B) - U(b, B)) \times P(B)) = -2.098 \text{ c-utiles}$$

This example is particularly telling because option *c* is statewise dominated by *a*. Whether state *A* or state *B* is actual, option *a* is strictly preferred to option *c*. Yet, introducing option *c* in the decision situation changes ECU theory's verdict: *b*, instead of *a*, is uniquely choiceworthy. ECU theory thus violates the *Irrelevance of statewise dominated alternatives (ISDA)* (Quiggin, 1994).

This gives rise to a worry. Without the IIA (and ISDA), it is possible to make up alternatives in any choice set and these manufactured alternatives would be altering the degrees of choiceworthiness of reasonable options.²⁵ This opens the door to strategic manipulation in the decision process. The worry can be overcome, however, if we accept Nicholas Smith's *theory of rationally negligible probabilities*: for any given decision, any outcome with probability $\leq p$, where *p* is very close to 0, can be rationally excluded from consideration in the decision process (Chalmers, 2017; Monton, 2019; Smith, 2014; Smith, 2016). As such, the *very improbable* outcomes of manufactured alternatives cannot alter the degrees of choiceworthiness of the other available options in the choice set.

Just as ECU theory delivers verdicts which are at odds with EU theory, ECU theory also supplies a more discriminating measure of the intervals in rankings of *more than two* choice options. Let us consider four choice situations involving decisions under certainty:

²⁵ Thanks to Douglas Lackey for raising this point and for wording suggestions.

Table 12²⁶

	1	2	3	4
<i>a</i>	5	5	5	5
<i>b</i>	1	1	1	1
<i>c</i>	1	2	2	2
<i>d</i>	1	3	3	3
<i>e</i>	1	3	5	8

The difference in CU between *a* and *b* is greater in situation 1 ($((5 - 1) - (1 - 5)) = 8$ c-utiles) than in situation 2 ($((5 - 3) - (1 - 5)) = 6$ c-utiles), and is greater in situation 2 than in situation 3 ($((5 - 5) - (1 - 5)) = 4$ c-utiles) and situation 4 ($((5 - 8) - (1 - 8)) = 4$ c-utiles), whereas the difference in utility between *a* and *b* is the same in all four situations (4 utiles). Therefore, compared to utility, CU is a more discriminating measure of the intervals between *a* and *b* in situations 1 to 4. What's more, there are not any contrary cases where CU (or CECU) gives a *less* differentiated picture than does utility (or EU).

9. In decisions under *risk*, what option(s) are choiceworthy sometimes differ from what option(s) it is rational to (weakly) prefer over the alternative options in the choice set. (9 follows from 5, 7 and 8.)

What the foregoing comparisons between EU theory and ECU theory show is that rational preference is not a reliable indicator of choiceworthiness. That is because whereas the criterion of rational preference (i.e., EU) satisfies the IIA (as assumed in this paper), the criterion of choiceworthiness (i.e., CECU) violates that principle (as demonstrated above). It is important to emphasize that the proposed criterion of choice (i.e., choiceworthiness) is independent from the standard choice criterion (i.e., rational preference). The latter is not shown here to violate the assumptions, for example, the IIA, which are needed to derive utilities from preferences via a representation theorem.

10. For any agent, *S*, faced with any decision under *risk*, it is a requirement of instrumental rationality that *S* should measure and rank his or her options in terms of *how choiceworthy* they are for *S* (i.e., how worthy of being chosen by *S* they are in light of *S*'s rational preferences over each of the various possible states of the world).

²⁶ Table 12 is identical to Table 1.

Whether (and to what extent) any option is more choiceworthy than any other within a set of alternatives is necessarily a function of *how* choiceworthy each of the two options is within the set of alternatives. (See the discussion of Step 2.)

Let us now consider again the *instrumental rationality argument*:

1. For any agent, *S*, faced with any decision under certainty or any decision under risk and for any option, *a*, for *S*, the *degree* to which *a* is *choiceworthy* for *S*, or (i.e.) the *choiceworthiness* of *a* for *S*, is the degree to which *a* is worthy of being chosen by *S* in light of *S*'s rational preferences over each of the various possible states of the world, where *S*'s *rational preferences* are preferences that obey the series of rationality conditions or axioms of standard decision theory.
 2. The *degree* to which *S*'s choosing *a* is *instrumentally rational* is the degree to which *S*'s choosing *a* is a suitable means to *S*'s ends (i.e., *S*'s rational preferences).
 3. The degree to which *a* is worthy of being chosen by *S* in light of *S*'s rational preferences over each of the various possible states of the world is the degree to which *S*'s choosing *a* is a suitable means to *S*'s ends (i.e., *S*'s rational preferences).
 4. Therefore, the *choiceworthiness* of *a* for *S* is the *degree* to which *S*'s choosing *a* is instrumentally rational. (4 follows from 1 to 3.)
11. It is *not* the case that for any agent, *S*, faced with any decision under *risk*, it is a requirement of instrumental rationality that *S* should measure and rank his or her options in order of rational preference. (11 follows from 3, 9 and 10.)
12. For any agent, *S*, faced with any decision under *risk*, it is a requirement of instrumental rationality that *S* should measure and rank his or her options in terms of how choiceworthy they are for *S*, that is, according to their CECU, rather than in order of rational preference, that is, according to their EU. (12 follows from 4, 6, 8, 10 and 11.)

I can offer two further arguments for Step 12. First, compared to the criterion of rational preference (i.e., EU), the criterion of choiceworthiness (i.e., CU/CECU) supplies a more plausible measure of the intervals in rankings of more than two choice options (see Section 4.1.1 and the discussion of Table 12). Second, contrary to the criterion of rational preference (i.e., EU), the criterion of choiceworthiness (i.e., CU/CECU) agrees with and entails Gandalf's principle, "an eminently sensible principle to incorporate into rational decision-making" (Bassett, 2015).

13. For any agent, *S*, faced with any decision under *risk*, it is a requirement of instrumental rationality that *S* should choose out of what option(s) are *choiceworthy* for *S* (i.e., what option(s) are worthy of being chosen by *S* in light of *S*'s rational preferences over each of the various possible states of the world).

Let us consider again the *instrumental rationality argument*:

1. For any agent, *S*, faced with any decision under certainty or any decision under risk and for any option, *a*, for *S*, *a* is *choiceworthy* for *S* if and only if *a* is worthy of being chosen by *S* in light of *S*'s rational preferences over each of the various possible states of the world, where *S*'s *rational preferences* are preferences that obey the series of rationality conditions or axioms of standard decision theory.
 2. *S*'s choosing *a* is *instrumentally rational* if and only if *S*'s choosing *a* is a suitable means to *S*'s ends (i.e., *S*'s rational preferences).
 3. *a* is worthy of being chosen by *S* in light of *S*'s rational preferences over each of the various possible states of the world if and only if *S*'s choosing *a* is a suitable means to *S*'s ends (i.e., *S*'s rational preferences).
 4. Therefore, *a* is *choiceworthy* for *S* if and only if *S*'s choosing *a* is instrumentally rational. (4 follows from 1 to 3.)
14. It is *not* the case that for any agent, *S*, faced with any decision under *risk*, it is a requirement of instrumental rationality that *S* should choose out of what option(s) it is rational for *S* to (weakly) prefer over the alternative options in the choice set. (14 follows from 9 and 13.)
15. For any agent, *S*, faced with any decision under *risk* where what option(s) are choiceworthy for *S* differ from what option(s) it is rational for *S* to (weakly) prefer over the alternative options in the choice set, it is a requirement of instrumental rationality that *S* should choose out of what option(s) are choiceworthy for *S* (i.e., what option(s) maximize CECU), rather than out of what option(s) it is rational for *S* to (weakly) prefer over the alternative options in the choice set (i.e., what option(s) maximize EU). (15 follows from 5, 7, 9, 13 and 14, as well as from 3, 5, 7, 9 and 10.)

Conclusion

In recent years, several alternatives to EU theory have been proposed, for example Mark Colyvan's (2008) *relative expectation theory (RET)*, Paul Bartha's (2007; 2016) *relative utility theory (RUT)*, and Lara Buchak's (2013) *risk-weighted expected utility (REU) theory*. In all finite decision cases, RET and RUT deliver the same rankings and recommendations as EU theory. As for REU theory, it *can* deliver the same rankings and recommendations as EU theory, depending on the risk attitude of the agent equipped with the REU decision rule. These alternative "*rational preference*" tracking decision theories are therefore subject to the same objection as that leveled here against EU theory: they fall short as theories of *instrumental rationality*.

Acknowledgements

Thanks to Douglas Lackey, Kenny Easwaran and several anonymous reviewers for very helpful comments on earlier drafts of this paper.

Bibliography

- Bartha, P. (2007). Taking Stock of Infinite Value: Pascal's Wager and Relative Utilities. *Synthese*, 154, 5–52. <https://doi.org/10.1007/s11229-005-8006-z>
- Bartha, P. (2016). Making do Without Expectations. *Mind*, 125, 799–827. <https://doi.org/10.1093/mind/fzv152>
- Bassett, R. (2015). A Critique of Benchmark Theory. *Synthese*, 192, 241–267. <https://doi.org/10.1007/s11229-014-0566-3>
- Bernoulli, D. (1738/1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22, 23–36.
- Bolker, E. (1966). Functions Resembling Quotients of Measures. *Transactions of the American Mathematical Society*, 2, 292–312. <https://doi.org/10.2307/1994401>
- Briggs, R. (2010). Decision-Theoretic Paradoxes as Voting Paradoxes. *Philosophical Review*, 119, 1–30. <https://doi.org/10.1215/00318108-2009-024>
- Briggs, R. A. (2019). Normative Theories of Rational Choice: Expected Utility. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2019/entries/rationality-normative-utility/>
- Buchak, L. (2013). *Risk and Rationality*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199672165.001.0001>
- Chalmers, A. (2017). *An Offer You Can't (Rationally) Refuse: Systematically Exploiting Utility-Maximisers with Malicious Gambles* (Unpublished honours thesis). The University of Sydney, Australia. <https://hdl.handle.net/2123/16131>
- Colyvan, M. (2008). Relative Expectation Theory. *Journal of Philosophy*, 105, 37–44. <https://doi.org/10.5840/jphil200810519>
- Colyvan, M. & Hájek, A. (2016). Making Ado Without Expectations. *Mind*, 125, 829–857. <https://doi.org/10.1093/mind/fzv160>
- Henderson, D. R. (2008). Opportunity Cost. *The Concise Encyclopedia of Economics*. Library of Economics and Liberty. <https://www.econlib.org/library/Enc/OpportunityCost.html>
- Hurka, T. (n.d.) *More seriously wrong* [Unpublished manuscript]. <https://www.brown.edu/academics/philosophy/ethical-inquiry/sites/brown.edu.academics.philosophy.ethical-inquiry/files/uploads/seriously%20wrong%202016.pdf>
- Hurka, T. (2019). More Seriously Wrong, More Importantly Right. *Journal of the American Philosophical Association*, 5, 41–58. <https://doi.org/10.1017/apa.2018.41>
- Jeffrey, R. (1983). *The Logic of Decision*, 2nd ed. Chicago: University of Chicago Press.
- Kolodny, N. & Brunero, J. (2020). Instrumental Rationality. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/spr2020/entries/rationality-instrumental/>
- Monton, B. (2019). How to Avoid Maximizing Expected Utility. *Philosophers' Imprint*, 19, 1–25. <https://hdl.handle.net/2027/spo.3521354.0019.018>
- Olsen, K. (2018). Subjective Rightness and Minimizing Expected Objective Wrongness. *Pacific Philosophical Quarterly*, 99, 417–441. <https://doi.org/10.1111/papq.12188>
- Peterson, M. (2017). *An Introduction to Decision Theory*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316585061>

- Quiggin, J. (1994). Regret Theory with General Choice Sets. *Journal of Risk and Uncertainty*, 8, 153–165.
<https://doi.org/10.1007/BF01065370>
- Ramsey, F. P. (1931). Truth and Probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and other Logical Essays* (pp. 156–184). London: Routledge and Kegan Paul.
- Savage, L. (1954). *The Foundations of Statistics*. New York: Wiley.
- Sen, A. (1993). Internal Consistency of Choice. *Econometrica*, 61, 495–521. <https://doi.org/10.2307/2951715>
- Sinhababu, N. (2018). Scalar Consequentialism the Right Way. *Philosophical Studies*, 175, 3131–3144.
<https://doi.org/10.1007/s11098-017-0998-y>
- Smith, N. J. J. (2014). Is Evaluative Compositionality a Requirement of Rationality? *Mind*, 123, 457–502.
<https://doi.org/10.1093/mind/fzu072>
- Smith, N. J. J. (2016). Infinite Decisions and Rationally Negligible Probabilities. *Mind*, 125, 1199–1212.
<https://doi.org/10.1093/mind/fzv209>
- Steele, K. & Stefánsson, H. O. (2020). Decision Theory. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/win2020/entries/decision-theory/>
- Temkin, L. S. (2012). *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. New York: Oxford University Press.
- Von Neumann, J. & Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- Wedgwood, R. (2013). Gandalf’s Solution to the Newcomb Problem. *Synthese*, 190, 2643–2675.
<https://doi.org/10.1007/s11229-011-9900-1>
- Wedgwood, R. (2017). Must Rational Intentions Maximize Utility? *Philosophical Explorations*, 20, 1–20.
<https://doi.org/10.1080/13869795.2017.1356352>
- Zhang, J., Timmermans, H., Borgers, A., & Wang, D. (2004). Modeling Traveler Choice Behavior Using the Concepts of Relative Utility and Relative Interest. *Transportation Research Part B*, 38, 215–234.
[https://doi.org/10.1016/S0191-2615\(03\)00009-2](https://doi.org/10.1016/S0191-2615(03)00009-2)
- Zhang, J. (2015). Relative Utility Modelling. In S. Rasouli & H. Timmermans (Eds.), *Bounded Rational Choice Behaviour: Applications in Transport*. Emerald Group Publishing Limited. <https://doi.org/10.1108/978-1-78441-072-820151014>

Chapter 5: Should You Invest in Life-Extension Medical Research?

The right to choose to live or to die is the most fundamental right there is; conversely, the duty to give others that opportunity to the best of our ability is the most fundamental duty there is.

—Aubrey de Grey

Should you invest in healthy life-extension medical research? Aubrey de Grey thinks that you should. De Grey is a biomedical gerontologist and the Chief Science Officer and Co-founder of SENS Research Foundation, a 501(c)(3) public charity that funds biomedical research into “regenerative medicine to repair the damage underlying the diseases of aging” (*About SENS Research Foundation*, n.d.). For the past 20 years or so, de Grey has been researching and promoting regenerative medicine solutions to the problem of aging.

In 2007, de Grey published the book *Ending Aging: The Rejuvenation Breakthroughs that Could Reverse Human Aging in Our Lifetime*. *Ending Aging* details how age-related decline and illness could be reversed in the near future if sufficient investments were made in what de Grey calls “SENS” research. SENS (or “Strategies for Engineered Negligible Senescence”) refers to a set of regenerative medicine approaches developed by de Grey which directly target age-related tissue damage at the microscopic level. While some gerontologists believe that the maximum lifespan can be extended through metabolic interventions (Olshansky, 2017; Zorn, 2018), de Grey thinks that metabolism is too complex to be effectively controlled for the foreseeable future. Just as we can keep vintage cars looking and operating like new by repairing broken parts or replacing them with new ones, de Grey points out that we can keep humans in a youthful state indefinitely by periodically removing or repairing damaged tissues at the cellular and molecular levels.

While de Grey’s “SENS” proposal to extend healthy human life remains controversial among gerontologists, SENS Research Foundation’s 30-member expert Advisory Board has endorsed de Grey’s conclusions that,

Recent biotechnological progress indicates that many aspects of aging may indeed be effectively treatable by regenerative medicine in the foreseeable future. We cannot yet know whether all aspects will be, but extensive scrutiny has failed to identify any definite exceptions. Therefore, at this point there is a significant chance that such therapies would postpone age-related decline by several years, if not more, which constitutes a clear case for allocating significant resources to the attempt to develop those therapies. (*SENS Research Foundation Research Advisory Board*, n.d.)

On this basis, de Grey has argued that Investments in regenerative medicine to extend healthy human life are morally imperative. One of his arguments can be put as follows:

What gives us the right to decide for future generations whether they should or should not live healthily past the age of 120 years? If we refrain from *investing* in healthy life-extension

technologies today because of moral concerns, then we are in effect denying future generations (our children and grandchildren) the right to live healthy lives past our current maximum lifespan. What is considered moral today might be considered loathsome tomorrow, either because circumstances have changed or because values have shifted (Schwitzgebel, 2018). So if we decide today not to *implement* those technologies because of overriding moral concerns, then we are rationally entitled not to do so, but what we are *not* rationally entitled to do is to prevent future generations from benefiting from those technologies should they deem them morally obligatory. Therefore, we can have a reasoned debate about whether it would be moral to *implement* healthy life-extension technologies today (and de Grey has forcefully argued that it is morally obligatory to do so), but what we cannot rationally refrain from doing is *investing* in healthy life-extension technologies today for the benefit of future generations (our children and grandchildren) should they deem those technologies morally necessary.²⁷

This suggests a more general principle of ethics and rational choice: Values and preferences may change over time. Therefore, to the extent that no current values or preferences are violated, a robust decision rule should leave open as many choice opportunities as possible to accommodate future changes in values or preferences.

Bibliography

- About SENS Research Foundation* (n.d.). SENS Research Foundation. Retrieved April 9, 2021, from <http://www.sens.org/about>
- De Grey, A., & Rae, M. (2007). *Ending Aging: The Rejuvenation Breakthroughs That Could Reverse Human Aging in Our Lifetime*. New York: St. Martin's Press.
- Olshansky, S. J. (2017). Is Life Extension Today a Faustian Bargain? *Frontiers in Medicine*, 4(215), 1–3. <https://dx.doi.org/10.3389%2Ffmed.2017.00215>
- SENS Research Foundation Research Advisory Board* (n.d.). SENS Research Foundation. Retrieved April 9, 2021, from <http://www.sens.org/about/leadership/research-advisory-board>
- Schwitzgebel, E. (2018, June 30). *Will Future Generations Find Us Especially Morally Loathsome?* The Splintered Mind. <https://schwitzsplinters.blogspot.com/2018/06/will-future-generations-find-us.html>
- Wareham, C. S. (2018). What is the ethics of ageing? *Journal of Medical Ethics*, 44, 128–132. <http://dx.doi.org/10.1136/medethics-2017-104374>
- Zorn, E. (2018, July 13). A 150-year-old human? Neither side is folding in The Great Longevity Wager. *Chicago Tribune*. <http://www.chicagotribune.com/news/opinion/zorn/ct-perspec-zorn-longevity-aging-olshansky-austad-20180715-story.html>

²⁷ For a helpful introduction to the ethics of ageing, see Wareham (2018).

Chapter 6: The Existential Passage Hypothesis

BOOK REVIEW: Metaphysics by Default, by Wayne Stewart, 1999, <http://mbdefault.org/>.

This online philosophical monograph is divided into 20 chapters, 8 appendices and a list of Works Cited. For an abstract of the entire work, see Wayne Stewart, “Metaphysics by Default: Naturalism and Metaphysics Reconciled,” *Metaphysics by Default*, <http://mbdefault.org/lectures/abstract2.asp>.

This review will focus on **Chapters 9 and 11** (accessed July 6, 2018). The review is divided into four sections: (1) summary of the Chapters’ conclusions, (2) assessment of the Chapters’ arguments, (3) further comments on the Chapters, including suggestions for future research, and (4) concluding remarks.

1. Summary of the Conclusions

In Chapter 9, Stewart defends the thesis that if non-reductive physicalism is true, then, contrary to a widespread belief, death does *not* bring about eternal oblivion, a permanent cessation of the stream of consciousness at the moment of death. Stewart argues that the stream of consciousness continues after death—devoid of the body’s former memories and personality traits—and it does so as the stream of consciousness of new, freshly conscious bodies (other humans, animals, etc., that are conceived and develop consciousness). And so, any permanent cessation of the stream of consciousness at the moment of death is impossible as long as new, freshly conscious bodies come to exist. *Consciousness* is defined here as awareness, and is not limited to self-awareness (i.e., the recognition of one’s awareness). This general thesis does not specify *when* in the future those *new, freshly conscious bodies* must have come into being. This thesis has been independently defended by several authors (Clark, 1994; Clark, n.d.; Darling, 1996; Sharlow, 2009; Spaulding, 1982).

In Chapter 9, Stewart argues *more specifically* for what he calls the *existential passage hypothesis*, expressing the concept in several ways, e.g.:

Where nature assembles necessary and sufficient conditions for a phenomenon, we trust nature to deliver the phenomenon. That trust applies to essay conditions, as everywhere.²⁸ [pers. comm.] It applies for example to William James’ unfelt time-gap; delivering the unfelt time-gap wherever nature assembles conditions for it, even if conditions are assembled across separate persons.²⁹

The following is my own restatement of that hypothesis:

For any conscious body, *x*, *x*’s stream of consciousness *continues*, following *x*’s permanent cessation of consciousness (or *death*), as the stream of consciousness of some

²⁸ http://mbdefault.org/9_passage/3.asp (Accessed July 6, 2018).

²⁹ Personal communication, July 15, 2018.

other body (or *passage recipient*), y , namely the *first body* to have *gained* (or first experienced) consciousness following x 's death.³⁰

Importantly, by “ x 's stream of consciousness *continues*...” what I mean (from this point onward) is that x 's stream of consciousness continues *unimbed* with x 's former memories and personality traits.

Given my restatement of the existential passage hypothesis, the notion of *existential passage* (as used by Stewart) can be stated as follows:

For any conscious body, x , the passage that occurs when x 's stream of consciousness *continues*, following x 's death, as the stream of consciousness of some other body (or *passage recipient*), y , namely the *first body* to have gained consciousness following x 's death.

In an appended chapter, Stewart argues that existential passage is unaffected by spatial distances and differences in central nervous systems (CNSs) and that this passage can thus theoretically occur between vastly distant worlds harboring living organisms with vastly different CNSs.³¹

In Chapter 11, Stewart explains and illustrates how four *types of existential passage* logically follow from the existential passage hypothesis. These four types are restated as follows.³²

- A *unitary passage*: For a given conscious body, x_1 , x_1 's stream of consciousness *continues*, following x_1 's death, as the stream of consciousness of some other body, y , namely the first body to have gained consciousness following x_1 's death; and neither a *merged* passage nor a *split* passage has occurred.
- A *merged passage*: For at least two conscious bodies, $\langle x_1 \dots x_n \rangle$, $\langle x_1 \dots x_n \rangle$'s streams of consciousness *continue*, following $\langle x_1 \dots x_n \rangle$'s deaths, as the stream of consciousness of some other body, y , namely the first body to have gained consciousness following $\langle x_1 \dots x_n \rangle$'s deaths.
- A *split passage*: For a given conscious body, x_1 , x_1 's stream of consciousness *continues*, following x_1 's death, as the streams of consciousness of at least two other bodies, $\langle y_1 \dots y_n \rangle$, namely the first bodies to have gained consciousness following x_1 's death, where those bodies have gained consciousness at the exact same moment in time. Stewart believes that split passages are probably unlikely since “developmental timings cannot approach the perfect synchronization posited in the split passage.”³³
- An *ex nihilo passage*: y 's stream of consciousness is not the continuation of any antecedent stream of consciousness. An *ex nihilo* passage occurs *if and only if* y achieves

³⁰ http://mbdefault.org/9_passage/2.asp (Accessed July 6, 2018).

³¹ http://mbdefault.org/20_proof/default.asp (Accessed July 6, 2018).

³² http://mbdefault.org/11_types/default.asp; http://mbdefault.org/12_grammar/default.asp (Accessed July 6, 2018).

³³ http://mbdefault.org/11_types/default.asp (Accessed July 6, 2018).

consciousness, but neither a unitary passage, nor a merged passage nor a split passage has occurred.³⁴

In Chapter 11, Stewart also discusses alternatives to his existential passage hypothesis. He calls these alternatives the *permeable identity hypotheses*.³⁵ He argues that they are conceivable, though unlikely to be true. Stewart identifies two: (I have restated these hypotheses and given them unique names.)

- The *strongly permeable identity hypothesis*: For any conscious body, x , x 's stream of consciousness *continues*, following x 's death, as the stream of consciousness of at least one other body, y , namely any body that is conscious following x 's death.
- The *weakly permeable identity hypothesis*: For any conscious body, x , x 's stream of consciousness *continues*, following x 's death, as the stream of consciousness of some other body, y , namely the first body to have gained or *regained* consciousness following x 's death.

Importantly, by *regained consciousness*, what I mean (from this point onward) is that y (the *body*) was previously conscious, then y lost consciousness and, after an unspecified amount of time, y returned to consciousness. According to *MedlinePlus*,

“Unconsciousness is when a person is unable to respond to people and activities. Doctors often call this a coma or being in a comatose state. [...] Being asleep is not the same as being unconscious. A sleeping person will respond to loud noises or gentle shaking. An unconscious person will not.” (*Unconsciousness—first aid*, n.d.)

2. Assessment of the Arguments

In Chapter 9, Stewart argues I think successfully for the core intuition of his existential passage hypothesis, i.e., the intuition that the passage between x 's death and y 's birth is “understood as unfelt time-gap, with nothing superadded—rather, and critically, with individuation *subtracted*. All that has ‘passed’ is a shift of perceived existential ‘moment’—a natural relocation of the awareness of existence.”³⁶ Philosopher Thomas W. Clark calls this *generic subjective continuity* (Clark, 1994; Clark, n.d.). In arguing for this core intuition, Stewart refers to two plausible, central concepts: *time-gaps* and the *stream of thought*. (I refer to the latter concept as the *stream of consciousness*.) Both concepts are credited to William James (James, 1890). Stewart's achievement is especially remarkable on account of how revolutionary the core intuition is.

In Chapter 11, Stewart does a really great job of explaining and illustrating the various types of existential passage and showing how they jointly exhaust the possibilities. This is crucial work for the development of a complete theory of generic subjective continuity.

³⁴ http://mbdefault.org/11_types/11_6.htm (Accessed July 6, 2018).

³⁵ http://mbdefault.org/11_types/default.asp#fn7 (Accessed July 6, 2018).

³⁶ http://mbdefault.org/9_passage/2.asp (Accessed July 6, 2018).

In a lengthy footnote (that would have merited discussion in a separate section), Stewart objects to both permeable identity hypotheses on the grounds that “the stream of thought persists unbroken throughout life” and that “we ourselves perceive subjective experience as a deeply unified whole”.³⁷ But these arguments are undercut by Stewart’s later acknowledgment that “the passage recipient would be ignorant of any such [passage] events, just as he or she would have been ignorant of the existential passage which transpired at conscious birth.”³⁸

In the same footnote, Stewart also objects to both permeable identity hypotheses on the grounds that they are not supported by any strong arguments or intuitions, and because “subjectivity is conserved in the thalamocortical system, even during sleep,” whereby that system sets a “baseline integrity of subjective experience.” The latter objection is addressed below in my counterargument to objection “(2) Potentiality”.

With regard to the former objection, let us assume for the sake of argument that Stewart is correct in claiming that no supporting arguments or intuitions are forthcoming. And let us define *permeable identities* as identities (or bodies) that are capable of *receiving* existential passages during the course of their lives, even after having initially achieved consciousness. Whether permeable identities *do* or *do not exist* we should *not* expect to have strong *supporting* arguments or intuitions about whether they exist, *either way*. Again, I need only point to Stewart’s own acknowledgement that if permeable identities *did* exist, then these identities would be completely oblivious to any passage events. Therefore, a lack of strong *supporting* arguments or intuitions about permeable identities cannot be counterevidence of permeable identities.

Actually, though, I believe that we *do* have strong supporting arguments and intuitions about permeable identities, more specifically, ***weakly permeable identities***—i.e., permeable identities that are capable of *receiving* existential passages only when they gain or *regain* consciousness. We can begin by noting that, all else being equal, hypotheses that posit weakly permeable identities are more parsimonious, and as such, more plausible than hypotheses that posit *strongly permeable identities*—i.e., permeable identities that are capable of *receiving* existential passages at *any and every* moment.

On this basis, strong arguments (or counterarguments) will be made (in what follows) in support of an *extended* hypothesis that posits *weakly permeable identities*. From now onward, I will refer to that hypothesis simply as the ***extended existential passage hypothesis***. I define it as follows:

For any conscious body, *x*, *x*’s stream of consciousness *continues, following any (permanent or temporary) loss of consciousness by x*, as the stream of consciousness of at least one body (or *passage recipient*), *y*, namely the first body to have gained or *regained* consciousness following *x*’s *loss of consciousness*, where instances of *x* can be instances of *y* (i.e., the passage recipient can be *x* itself).

³⁷ http://mbdefault.org/11_types/default.asp#fn7 (Accessed July 6, 2018).

³⁸ http://mbdefault.org/11_types/default.asp#fn7 (Accessed July 6, 2018).

The notion of *extended existential passage* is defined as follows:

For any conscious body, x , the passage that occurs when x 's stream of consciousness *continues, following any (permanent or temporary) loss of consciousness by x* , as the stream of consciousness of at least one body (or *passage recipient*), y , namely the first body to have gained or *regained* consciousness following x 's *loss of consciousness*, where instances of x can be instances of y (i.e., the passage recipient can be x itself).³⁹

Using Chapter 11's four passage types as a template, we can identify four *types of extended existential passage*:

- A *unitary passage*: For a given conscious body, x_1 , x_1 's stream of consciousness *continues*, following any loss of consciousness by x_1 , as the stream of consciousness of some body, y , namely the first body to have gained or *regained* consciousness following x_1 's loss of consciousness (where x_1 can be y); and neither a *merged* passage nor a *split* passage has occurred.
- A *merged passage*: For at least two conscious bodies, $\langle x_1 \dots x_n \rangle$, $\langle x_1 \dots x_n \rangle$'s streams of consciousness *continue*, following any losses of consciousness by $\langle x_1 \dots x_n \rangle$, as the stream of consciousness of some body, y , namely the first body to have gained or *regained* consciousness following $\langle x_1 \dots x_n \rangle$'s losses of consciousness (where $\langle x_1$ or $\dots x_n \rangle$ can be y).
- A *split passage*: For a given conscious body, x_1 , x_1 's stream of consciousness *continues*, following any loss of consciousness by x_1 , as the streams of consciousness of at least two bodies, $\langle y_1 \dots y_n \rangle$, namely the first bodies to have gained or *regained* consciousness following x_1 's loss of consciousness, where those bodies have (respectively) gained and *regained* consciousness at the exact same moment in time (where x_1 can be $\langle y_1$ or $\dots y_n \rangle$).
- An *ex nihilo passage*: y 's stream of consciousness is not the continuation of any antecedent stream of consciousness. An *ex nihilo* passage occurs *if and only if* y gains or *regains* consciousness, but neither a unitary passage, nor a merged passage nor a split passage has occurred.

Stewart does not *explicitly* acknowledge this *extended* existential passage hypothesis.

Other than his arguments against strongly permeable identities and weakly permeable identities (which I have addressed above), Stewart's *implicit* rejection of this *extended* existential passage hypothesis in favor of his more *restricted* existential passage hypothesis appears to be based on Arguments (1) and (2) as they are restated below. Here is a telling passage from Chapter 9:

Subjectively, Nicos' *unfelt* time-gap *continues*, indefinitely. [...]

This particular time-gap is unusual in that it is open-ended. Nicos' inanimate body cannot restore subjectivity to Nicos in future; as a result, it cannot end the time-gap which Nicos' death has initiated.

³⁹ Contrary to existential passage, *extended existential passage* can thus occur between living bodies.

Hereafter I will refer to this special type of unfelt time-gap as a ‘mortal amnesia’: it is the forgetfulness of existence we can associate with failure of the criteria of personal identity. By prior reasoning this amnesia is irreversible. Having encountered mortal amnesia, Nicos afterwards lacks the means of perceiving any aspect of his condition, or of recovering in future any of the memories which death has destroyed.⁴⁰

Arguments (1) and (2) and my counterarguments to each are as follows:

- (1) *Backward causation*: The *future* restoration of x ’s personal identity (or alternatively, the *future* restoration of key attributes guarantying the continuity of x ’s personal identity), upon or after x ’s return to consciousness, prevents x ’s extended existential passage (to another passage recipient) from occurring in the present.

Note: In Chapter 8, Stewart argues at length that x ’s personal identity is best understood as a combination of three key attributes: physical continuity, episodic memory and subjectivity.⁴¹

My reply to (1): Backward causation is only possible if we accept a *tenseless theory of time* (or *B-theory of time*)—where the past, present and future are equally real. But the notion of a *stream of consciousness* (as it is used in Chapters 9 and 11) seems to necessitate a *tensed theory of time* (or *A-theory of time*)—where the present is real, but *not* the future. This is because the tensed (or A-) theory of time is seemingly the only theory of time that allows for the *objective passage of time* (or objective becoming) that is needed to make sense of the notion of a *stream of consciousness*. Without objective temporal passage (or objective passage of time), conscious experience is nothing more than a set of counterfactually-related conscious experiences superimposed on a set of time coordinates. Consequently, without objective temporal passage, there can be no stream of consciousness and so, no generic subjective continuity—i.e., no existential passage and no *extended* existential passage. According to *The Stanford Encyclopedia of Philosophy*,

A proper notion of backward causation requires a static account of time in the sense that there is no objective becoming, no coming into being such that future events exist on the par with present and past events. It means that the future is real, the future does not merely consist of unrealised possibilities or even nothing at all. [...] If backward causation is to be conceptually possible it forces us to be realists with respect to the future. The future must contain facts, events with certain properties, and these facts can make sentences about the future true or false. Such a realist account is provided by static and tenseless theories of time. (Faye, 2018)

Some recent metaphysical work has however challenged the widely accepted view that, under a tenseless (or B-) theory of time, time does not objectively pass:

Most B-theorists defend the reality of both time and change. Overwhelmingly, however, they deny that time genuinely passes, insisting that the passage of time is some kind of cognitive illusion. In this chapter it is argued that, while A-theoretic accounts of the passage of time are indeed mistaken, there is no reason for the B-theorist to resist the idea of mind-independent temporal passage. This mistake stems from two sources: first, the implicit acceptance of the A-theory’s

⁴⁰ http://mbdefault.org/9_passage/default.asp (Accessed July 6, 2018).

⁴¹ http://mbdefault.org/8_identity/default.asp (Accessed July 6, 2018).

understanding of passage; secondly, from the unnecessary assumption that temporal passage is best understood as some kind of motion. A tenseless, relational account of passage that is based on tenseless, temporal relations is presented and defended. It is further argued that the B-theory is compatible with an objective direction of time. (Mozersky, 2015)

If this new perspective is correct, then the idea of a *stream of consciousness* appears to be fully consistent with a tenseless (or B-) theory of time. So, let us suppose for the sake of argument that the idea of a *stream of consciousness* is entirely consistent with a tenseless (or B-) theory of time. Is (1) then salvageable?

I do not believe so. The reason I say this is that Argument (1) presupposes Argument (2) (see below), and Argument (2) is unsustainable. Let me explain: (1) states that personal identity or key attributes thereof have the power to retro-cause x 's prior stream of consciousness to continue in x . But as I will explain in my reply to (2), we have no reason to think that personal identity or key attributes thereof have any such potentiality—i.e., whether we take the cause to precede its effect (forward causality) or the effect to precede its cause (backward causality).

(2) *Potentiality*: When x has all the markers of temporary unconsciousness, x has the potential to *receive* x 's continued stream of consciousness, and so, upon x 's return to consciousness, x 's prior stream of consciousness seamlessly continues as x 's renewed stream of consciousness.

My reply to (2): As long as x 's neural and cognitive machinery remains intact, x has the potential to experience a stream of consciousness imbued with x 's memories and personality traits. *That*, we can all agree on. However, we have no justifiable reason for claiming that upon x 's return to consciousness, x 's stream of consciousness must have *this or that* origination *on the basis of* x 's neural and cognitive machinery.

Since streams of consciousness *unimbued* with the bodies' memories and personality traits differ only in their originations and since x 's neural and cognitive machinery cannot discriminate among originations, unimbued streams of consciousness are indistinguishable to x 's neural and cognitive machinery. Therefore, as long as x 's neural and cognitive machinery remains intact, then upon x 's return to consciousness, x has the potential to *receive* any *unimbued* stream of consciousness—i.e., either *ex nihilo* or from any conscious body—*within* the passage rules entailed by generic subjective continuity.

It is also worth noting that personal identity abstracted from x 's neural and cognitive machinery has no potentiality as it is *abstracta*—abstract objects (e.g. numbers, sets, propositions, etc.) are considered causally inert (Rosen, 2017).

For all these reasons, I believe that (2) is unsustainable. And if we apply the same line of reasoning to y , then we also have strong reasons for positing a *weakly permeable identity*—i.e., where y is the first body to have gained *or regained* consciousness following x 's loss of consciousness.

To sum up, I have argued that as long as generic subjective continuity is *itself* plausible, then among the various *alternative hypotheses* that posit generic subjective continuity (e.g. the weakly permeable identity hypothesis, the strongly permeable identity hypothesis, the (restricted)

existential passage hypothesis, etc.), the *extended* existential passage hypothesis is the only one of those hypotheses that can be considered plausible. As such, since Stewart argues successfully for generic subjective continuity, we have every reason to give credence *not* to the *restricted* existential passage hypothesis but rather to the *extended* existential passage hypothesis (as defined and explicated above).

3. Further Comments and Suggestions

In some *multiverse theories* (i.e., physical theories positing multiple universes), such as the many-worlds interpretation of quantum mechanics, causal interactions between different universes are a theoretical possibility (Plaga, 1997). Such multiverse theories thus certainly do not rule out extended existential passage between different universes.

In Chapter 18, Stewart explains in what ways his passage hypothesis could profitably inform decision theory and ethics. Stewart points out that if we accept his passage hypothesis, then we are rationally required to do what we can for the welfare of *posterity*—i.e., future generations of conscious creatures—because we ourselves will join posterity via existential passage.⁴²

Building on Chapters 9 and 11, one worthwhile project would be to research how likely it is for someone’s unimbued stream of consciousness to pass—via extended existential passage—to a comparatively *worse* stream of consciousness. One would need to calculate the odds of various experiential outcomes of extended existential passage—e.g. the passage recipient lives a pleasant life, the passage recipient is plunged into misery, etc.—under various assumptions—i.e., one or multiple spatiotemporal universe(s) (i.e. multiverse theory). With this knowledge, every person would be able to determine, in light of their individual circumstances and on the basis of rational choice theory, whether it would be rational for them to endeavor to prolong their conscious life *indefinitely* in order to avoid risky extended existential passages. For example, this research could count as a pragmatic reason to make investments in life-extension medical research.

4. Concluding Remarks

On the basis of my review of Chapters 9 and 11, I can confidently say that these chapters are a must-read for any person interested in their own existential fate and in that of human-kind and

⁴² “For [Peter] Singer, the critical anatomic structure is just the central nervous system (CNS) itself. The CNS makes possible the sensation of pleasure and pain. Consequently the CNS makes a creature deserving of natural rights and ethical treatment. Singer’s ethical conclusion dovetails with the metaphysical conclusion of Chapter 17. In that chapter we found that Metaphysics by Default would seem to apply not to *Homo sapiens* alone, but to CNS species generally. CNS species have been shown to meet the criteria of personal identity: it follows that creatures of all CNS species may be thought to participate in the web of existential passages described by Metaphysics by Default.” (http://mbdefault.org/18_benefits/default.asp. Accessed July 6, 2018.)

conscious-kind more generally. These chapters deserve to be widely cited in the philosophical literature, especially on metaphysics, ethics and decision theory, where Stewart's existential passage hypothesis and extensions thereof have significant implications.⁴³

Acknowledgements

Many thanks to Wayne Stewart for helpful comments on an earlier draft of this chapter.

Bibliography

- Baker, A. (2016). Simplicity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/win2016/entries/simplicity/>
- Briggs, R. A. (2017). Normative Theories of Rational Choice: Expected Utility. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/spr2017/entries/rationality-normative-utility/>
- Clark, T. W. (1994). Death, Nothingness, and Subjectivity. *The Humanist*, 54(6), 15–21.
<http://www.naturalism.org/philosophy/death/death-nothingness-and-subjectivity>
- Clark, T. W. (n.d.). *A Notable Theoretical Convergence*. Naturalism.Org. Retrieved April 9, 2021, from
<http://www.naturalism.org/philosophy/death/a-notable-theoretical-convergence>
- Darling, D. J. (1996). *Zen Physics, the Science of Death, the Logic of Reincarnation*. Harpercollins.
- Faye, J.. (2018). Backward Causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/sum2018/entries/causation-backwards/>
- James, W. (1890). *The Principles of Psychology*. Volume 1. New York: Holt.
<https://archive.org/details/theprinciplesofp01jameuoft>.
- Markosian, N. (2016). Time. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/fall2016/entries/time/>
- Mozersky, J. (2015). The B-theory and the Passage of Time. *Time, Language, and Ontology: The World from the B-Theoretic Perspective*. Oxford University Press.
<http://dx.doi.org/10.1093/acprof:oso/9780198718161.003.0008>
- Plaga, R. (1997). On a possibility to find experimental evidence for the many-worlds interpretation of quantum mechanics. *Foundations of Physics*, 27(4), 559–577. <https://doi.org/10.1007/bf02550677>
<https://arxiv.org/pdf/quant-ph/9510007.pdf>
- Rosen, G. (2017). Abstract Objects. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/win2017/entries/abstract-objects/#CausInefCrit>
- Sharlow, M. F. (2009). Why Science Cannot Disprove the Afterlife. *Patchworks*.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.638.1519&rep=rep1&type=pdf>
- Spaulding, W. C. (1982). *The Creation of I's*. Williamspaulding.com. <https://williamspaulding.com/philosophy/i.htm>
- Stewart, W. (n.d.). *Lectures*. Metaphysics by Default. Retrieved April 9, 2021, from
<http://mbdefault.org/lectures/transcript.asp>
- Stewart, W. (n.d.). *Metaphysics by Default: Naturalism and Metaphysics Reconciled*. Metaphysics by Default. Retrieved April 9, 2021, from <http://mbdefault.org/lectures/abstract2.asp>
- Sturm, T. (2012). Consciousness regained? Philosophical arguments for and against reductive physicalism. *Dialogues in Clinical Neuroscience*, 14(1), 55–63. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3341650/>

⁴³ http://mbdefault.org/18_benefits/default.asp (Accessed July 6, 2018).

Unconsciousness—First Aid (n.d.). MedlinePlus. U.S. National Library of Medicine (NLM), National Institutes of Health (NIH), U.S. Department of Health & Human Services. Retrieved April 9, 2021, from <https://medlineplus.gov/ency/article/000022.htm>