EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 21(1)
March 2024: 69–78

Rejoinder on Ergodicity Economics

Matthew C. Ford[1] and John A. Kay[2]

LINK TO ABSTRACT

In our critique of ergodicity economics (Ford and Kay 2023), we had three major aims. Firstly, we wanted to set out the reasons ergodicity economics (EE) and expected utility theory (EUT) differ. Secondly, we wanted to set out difficulties with the justification and practical implementation of EE, difficulties that EUT does not face. Finally, much of the literature on EE has criticised EUT, either explicitly or implicitly. We wanted to respond to these criticisms where they were erroneous and in doing so show that the two theories may not be as far apart as they appear in many practical situations, which has implications for experimental work.

In a response (Hulme et al. 2023) to our critique, eleven authors (hereafter 'the authors') involved in EE research raise two issues with our paper: our treatment of the mapping between EE and EUT; and our treatment of the link between a converging growth rate and final wealth. Understanding these, they suggest, accounts for many of our conclusions. We are grateful for the generosity and fullness of their reply, but we believe that it is predicated on a misreading of our paper—for which we must take some of the blame, since we made a mistake in one aspect of our analysis which they helpfully correct. We address this below.

Having clarified these issues, we believe that the main points we made about the justification of EE and its practical applicability remain salient and unanswered. The conclusions of our original paper remain: it is still the case that, unlike EUT, there is no clear justification for using EE; there is no clear guidance on which situations it is applicable in; and there is good reason to believe that, compared to EUT, it is often a poor model of the world.

1. St. John's College, Oxford OX1 3JP, UK.
2. St. John's College, Oxford OX1 3JP, UK.

Finally, the authors note that they are in at least partial agreement with our concerns about experimental work done on EE and that some of them are conducting a new experiment which addresses these issues (Hulme et al. 2023, 345; Skjold et al. 2023). We discuss the new experiment with two aims: to clarify what we might learn from its results, and to show the practical importance of several of the theoretical issues we raised in our original paper.

Clarifying concerns

An important section of our original paper addressed an unintuitive result: there are gambles which, in the limit, converge almost surely on a multiplicative growth rate below 1. A multiplicative growth rate below 1 means that wealth will fall over time. Despite this, there exist utility functions which show that agents will take those gambles. But these agents prefer having more wealth to less wealth. Knowing this, it might seem that in the limit these utility functions will not show that agents will take these gambles—that is, perhaps they will take them when the gamble length is short but not when it is long—and therefore EUT and EE converge.

In fact, this intuition is incorrect. In our original paper, we write:

> "Samuelson (1971)'s contribution is to show that **this is not the case** for many utility functions… In the limit case those high-value outcomes are suppressed, explaining why EUT might **appear** to coincide with growth optimality… risk-neutral EUT is **not** growth-optimal… Samuelson (1971) can be extended to show that growth-optimal behaviour **necessarily violates** the axioms of EUT because it is **incompatible** with **any** utility function." (Ford and Kay 2023, 318–319; boldface added here)

The authors, however, appear to think that we argue the opposite. They write: "A mapping between the models exists, but the key condition which needs to be satisfied for this mapping to hold is that the utility function in EUT is chosen to be the ergodicity transformation of EE. It seems that the authors believe that the key condition is merely sufficiently long time scales" (Hulme et al. 2023, 336). We believe that both their points—on the EE-EUT mapping and on the distinction between convergence in growth and convergence in wealth—are based on the misunderstanding contained in that quotation.

Regarding the mapping: we do not dispute the fact that there is a unique mapping between EE and EUT, and the existence of this unique mapping is a result we drew on in our paper to show the difference between EE and EUT. We wrote:

> Choosing a gamble to maximise the growth rate of wealth looks equivalent to choosing a specific utility function: if the dynamics are multiplicative we maximise the expected value of $u(w) = \ln(w)$; if they are additive we maximise the expected value of $u(w) = w$. Peter Carr and Umberto Cherubini (2020) draw on this insight and vary a stochastic clock to show that a variety of utility functions can be justified in this way, and Peters and Adamou (2021) sets out a general way to find a correspondence between utility functions and dynamics (where this exists). (Ford and Kay 2023, 319)

We believe that the misunderstanding stems from a difference in emphasis. Towards the end of our paper, we pull back from the theoretical analysis and raise the question of what the models imply for "real decisions which real people face" (Ford and Kay 2023, 326). We do believe it is *likely* that, for many of the choices people face, applying an EUT analysis having estimated a utility function and applying an EE analysis will result in much the same ordering over the choices available: this is what we meant when we wrote that "utility approaches and growth-optimal approaches are likely to give the same answer in many cases" (ibid.). We were implicitly assuming these choices are relatively small in number: in everyday life we do not face a continuum of marginally different stochastic processes. Of course we agree that, faced with the full set of possible stochastic processes, EE and EUT in general do not produce the same ordering.

Furthermore, one of our reasons for stressing this point was that much of the academic literature on EE implicitly misunderstands EUT by treating it as equivalent to myopic EUT. As we will discuss below, this was the case in the experimental analysis run by David Meder et al. (2021) and proposed by Benjamin Skjold et al. (2023). Furthermore, Ole Peters wrote that "by wrongly assuming ergodicity, wealth is often replaced with its expectation value before growth is computed", and, referring to the experimental work subsequently published as Meder et al. (2021), "the experiment may be flawed in a way we don't yet understand" (Peters 2019, 1216, 1220). In fact, EUT does not replace wealth with its expectation value before calculating growth, and this was indeed a flaw in Meder et al. (2021). In their reply, the authors claim that "EUT…does not take dynamic information into account" (Hulme et al. 2023, 341). This is simply incorrect, for reasons we explain in our original paper. (We would be happy to expand on this point if the authors could explain why they believe that this is not the case, but no such argument is offered.)

In general, myopic EUT and EUT properly applied will not give the same answers; we discuss the special case where they do below. This is what we meant when we wrote that the fact that the expected value and time average may not equate is "mechanically incorporated" in an EUT analysis (Ford and Kay 2023, 318). We did not dispute that, even with this incorporation, EUT and EE will in general give different answers.

Regarding convergence: the authors identify a mistake we made when discussing the properties of terminal wealth, and we are grateful for the correction. As they note, we wrote that "The time averages used in [the EE model] correspond to a situation where there is no measurable uncertainty—final wealth will almost always be what the time average predicts" (Ford and Kay 2023, 317). They are correct to note that this is not true: as they write, even though the growth rate converges, terminal wealth diverges, and this uncertainty is measurable.

Without wishing to downplay the importance of this correction, we do not believe that it invalidates our argument. Our claim was that both growth and wealth *are* uncertain and that this explains why the intuition that EE and EUT should, in the limit, give the same result is incorrect. As we will discuss below, EE's justification relies on the growth rate's convergence; and of course this only matters because of its implications for terminal wealth. In the authors' words, "EE focuses on maximising the time-average (or expected) growth rate of wealth. In doing so, it guarantees that EE agents, unlike EUT agents, maximize not only wealth but also utility as time passes" (Hulme et al. 2023, 340). Whilst the specific claim we made about the convergence of terminal wealth was wrong, the broader point we were making was correct.

Unresolved issues

We believe two important questions raised in our original paper remain: what is the justification for using EE; and which gambles is it appropriate to apply it to? The authors write that EE's motivation is that agents who maximise the time average of the appropriately defined growth rate "maximize the long-term growth rate of their wealth. In the long run, agents who act in this way become wealthier than agents who act differently" (Hulme et al. 2023, 337–338). Elsewhere, they claim that EE "guarantees that EE agents, unlike EUT agents, maximize not only wealth but also utility as time passes. … As time passes, EE agents are guaranteed to do better than EUT agents, in terms of wealth and utility" and that "maximisation [of the time average] guarantees that we end up with greater wealth (and utility) in the long-time limit" (ibid., 340, 343). So the justification appears to be that using EE means your wealth grows faster and ends up being larger than using any other approach.

It is important to be clear on what the authors are saying here. Consider Peters' bet, played for $t < \infty$ rounds. An EE agent will decline to play this bet and keep their initial wealth w_0. In contrast, an EUT agent with linear utility will accept this bet. As long as $t > 1$, they will probably end up with terminal wealth $w_t < w_0$; nevertheless it is possible that they will end up with $w_t > w_0$, for example if the coin lands heads on every toss. "Guarantees" and "maximize," therefore, do not mean that there exists any physically possible situation, i.e., a finite gamble, in which an EE agent will certainly end up with more wealth than an agent who takes the bet. Formally, they are saying P($w_t < w_0$ for all sufficiently large t) = 1. But for any finite value of t, this probability will be less than 1.

In our original paper we quoted Henry Latané who explicitly discussed the "probability of adverse dominance" (Latané 1979, 310; quoted in Ford and Kay 2023, 322), and we believe that this is a much clearer perspective on the problem. His formulation also makes it clear that for an agent playing Peters' bet P($w_t > w_0$) > 0 for every $t > 0$. Given this, we do not think looking at the limit result provides much insight, and certainly does not justify claims that EE "guarantees" that EE agents do better than an agent who takes the bet.

In the light of this, we should update EE's justification: using EE means your wealth *probably* grows faster and ends up being larger than using any other approach. But it is unclear why it makes sense to rank choices by their most probable outcomes. This is not how we generally act in life: many people fasten seatbelts and insure their house against fire even though the most probable outcome is that they are not involved in an accident and their house does not burn down; some, including some of the same people, buy lottery tickets for the pleasure of imagining a prize they do not expect to win. We apply for a good job which we know we will probably not get because the upside and chance that we *will* get it are sufficiently large relative to the costs of applying; we invest in a startup which we know probably won't succeed because the potential payoff is enormous. There is no clear answer to the question 'how probable must an outcome be before we can disregard all other outcomes?' short of 'it must be certain.'

In fact, as we note in our original paper, it is unclear why we would ever want to base a decision rule on any kind of average value, rather than examining the possible outcomes and their associated probabilities. Past work on EE (Peters and Gell-Mann 2016; Peters 2019) implicitly justified EE by arguing that EUT based decisions on an inappropriate average, whereas EE based them on an appropriate average. But as we explained in our original paper, this is not the justification generally made for EUT, which merely represents preferences. Under special circumstances, therefore, EUT's use of an average is justified. But there is no equivalent justification for EE's use of an average.

Our second question is about which finite gambles it is appropriate to apply EE to. Let us assume, for the sake of argument, that there is a good answer to why we want to use EE in the first place. Is there a rule with a convincing justification for which gambles are admissible?

The authors appear to address this: the model of EE they set out has no restrictions on t, suggesting that it applies even to gambles of length 1. But as we note in our original paper, this leads to bizarre results. Firstly, it implies that a gambler could develop a gamble where $t = 1$ and toss a coin to decide whether he describes it additively or multiplicatively. This is just an issue of labelling, of course: the gamble is exactly the same either way. And yet the EE agent will accept or decline the bet based on that description! The authors argue that EE preferences are complete as the dynamic in such cases is "a hidden variable" (Hulme et al. 2023, 341), but since this hidden variable has no impact on what actually occurs we do not find this a very persuasive defence.

In their appendix, the authors take another approach. They note that we can find a value, christened T^*, by equating the expectation and standard deviation of the growth rate of a particular gamble. Using this, they claim they can "quantif[y] what we mean by 'large' time" (Hulme et al. 2023, Appendix p. 4). But it is not apparent why doing so tells us what we want to know. Indeed, they immediately undercut the idea that it does in three ways: they write that "'large time' corresponds to a number of iterations much larger than [T^*]" (without specifying what "much larger" might mean); that this number will also depend on "the required level of certainty of the growth rate"; and that in any case "standard deviation is not a robust measure of the dispersion" (Hulme et al. 2023, Appendix pp. 5, 7). The only thing this makes clear is that EE doesn't guarantee a higher growth rate in real situations.

Secondly, whatever a gamble might look like in the limit, it seems clearly unpersuasive to say that we should take a certain gamble because doing so will maximise the long-term growth rate of our wealth, when in the time that *actually* elapses there is, for example, a 0.5 or 0.25 probability of it *not* maximising our growth rate.

Thirdly, having no minimum value of t results in EE agents engaging in bizarre behaviour. It is easy to construct pairs of additive gambles where one has a marginally higher growth rate than the other and much higher variance. Thus we have a situation where gambles which are obviously very risky are taken in preference to gambles which are only marginally less generous and substantially safer. In our original paper we present an example and note that it implies EE may not apply even to gambles with very large values of t (Ford and Kay 2023, 324).

Finally, we note in our original paper that EE violates completeness, meaning that there are gambles where it behaves inconsistently or offers no ranking. As we discuss above, cases where $t = 1$ are an example of the former issue. But, as we discuss at greater length in our original paper, there are other cases where EE certainly violates completeness (in both an EUT and EE sense) because it cannot rank gambles. Such cases include gambles with different dynamics and time-varying gambles, even in cases where these gambles do converge to a particular growth rate in the limit.

Experimental issues

As a result of the issues we discuss above, as well as others, we noted that two experimental studies on EE (Meder et al. 2021; Vanhoyweghen et al. 2022) had serious problems which meant that we did not follow their conclusions. The authors, several of whom were involved in those studies, accept at least some of these criticisms (Hulme et al. 2023, 345), and five of them have designed and preregistered a new experiment (Skjold et al. 2023).

The Skjold et al. (2023) experiment involves subjects making a series of choices between lotteries in additive and multiplicative dynamics; these lotteries change the number of experimental points they have, and their payout in real money is determined by their points. The choices are then analysed to estimate each subject's coefficient of relative risk aversion, on the grounds that EE predicts these will differ between dynamics whereas EUT predicts that, for each individual subject, they will not. Unfortunately, this experimental design also appears to have problems, in large part because the issues we raised in our original paper—in general, and more specifically in our appendix—have not all been considered.

Two issues relate to our broader critique of EE. Firstly, as in Meder et al. (2021), the experiment uses myopic EUT to estimate the coefficient of relative risk aversion, resulting in the same issues we discussed in our original paper. In a different context, the authors discuss Peters' bet under EUT and write that "we let the agent evaluate utility after one round, although nothing changes if the agent were to evaluate utility after an arbitrary number of rounds." (Hulme et al. 2023, 342). It is true in the case they set up there that the two approaches are equivalent, but this is a special result which only occurs when the agent is staking their entire wealth on the gamble in each period and has a utility function from the constant relative risk aversion (CRRA) family. We mention this because, as we noted in our original paper, these conditions are not met in the experiments: firstly, agents may not have a CRRA utility function; secondly, they certainly have significant external wealth which is not subject to the experimental dynamics. Only considering CRRA utility functions and ignoring outside wealth also biases the experimental estimates for other reasons: readers interested in more detail about both of these issues should examine Appendix B in our original paper (Ford and Kay 2023, 328–332).

Secondly, the issues we noted above about when it is appropriate to use EE for finite gambles are relevant here. The experimental analysis appears to put equal weight on every data point gathered when estimating the coefficient of relative risk aversion. As we note above, gambles where $t = 1$, i.e., the last choice in each session, can be described as additive or multiplicative, so it is unclear what result

EE predicts for them.[3] For gambles where $t > 1$ but still small, it is unclear whether EE should apply to them; at the very least it seems that proportionally less weight should be put on those choices. Furthermore, we note that subjects' performance in the task is positively correlated with their payout in real money. In the additive session, we can think of each choice as being over an amount of money which will be taken out of the experiment into the real world, where the wealth dynamics the subjects will experience are presumably multiplicative. Applying the EE model, it seems that in this session they should treat every gamble as being part of a multiplicative dynamic and that their coefficient of relative risk aversion should equal 1, rather than the 0 they predict.

There is a further problem with the experimental design in Skjold et al. (2023). Similarly to Vanhoyweghen et al. (2022), gambles in the experiment are over experimental points rather than money. At the end of a session, subjects receive a share of a ~365USD prize pool based on their performance relative to 9 other subjects. (For example, if the subject had finished with 3,000 points and the other 9 subjects had collectively finished with 12,000 points, the subject would receive ~73USD.) The utility functions estimated are therefore over experimental points, not wealth. They would only be equivalent if there was a linear relationship between points and wealth (and if, as mentioned above, outside wealth was considered). This is not the case. If a subject has a very large point total, then (even holding the points earned by others in their group equal) they gain much less from getting additional points than they do when their point totals are lower. This asymmetry creates an incentive to be risk averse (when measured in terms of experimental points) even if a subject is really risk neutral (when measured in terms of wealth). The specific issue here is distinct from that in Vanhoyweghen et al. (2022), where the use of a payout based on experimental points and relative performance incentivised risk seeking, but the general issue is the same: the experimenters have not considered how, in EUT terms, the problem should be set up, and so make mistakes in their estimation.

Conclusion

We hope that this exchange has resulted in the debate making some progress. For our part, we are glad to have the opportunity to clarify what we meant in various passages about the practical differences between EUT and EE, and are happy to affirm that EE and EUT are only exactly equivalent when the conditions for the specific mapping between them is met—a result we relied on in our original

3. As we explained above, we do not find the claim that EE includes the dynamic as an additional variable persuasive. Regardless of one's view on this, the experiment does not specify to the subjects that they are in an additive or multiplicative dynamic, so appealing to this additional variable does not deal with this issue.

paper. Similarly, we are grateful for the correction regarding diverging terminal wealth. Fortunately, our mistake did not have any serious implications for our analysis. Whilst we cannot make assumptions about what the authors leave unstated, we hope that the lack of discussion of the metaphysical issues with EUT means that this particular concern has been alleviated, and the comparison of the two theories can take place on less exalted territory.

Nonetheless, our original argument—hopefully clarified by this exchange—remains unanswered. It has three main elements: we do not believe that there is a compelling justification for EE; we do not believe it is at all clear when it is appropriate to use EE as a model; and we believe that EUT has been seriously misrepresented by proponents of EE, which undermines one of the justifications offered for EE. We further believe that, as a direct consequence of these issues, experimental work planned on EE is seriously flawed.

No single model can explain human behaviour and economists should use a toolbox of different models. Very few decisions—from what detergent to buy to which career to follow—are purely financial; most of the risks individuals face—such as those of disappointment or injury—are non-financial. EUT, which derives from axioms of consistent but idiosyncratic preferences, accommodates this; EE, by contrast, seems to have nothing to say here.

We remain confident in our original conclusion. Adapting George Box: all models are wrong, but some are useful; some are more wrong and less useful than others. The limited use case for EUT (and similar theories) is clear: it is a convenient mathematical representation of some sets of preferences which in some cases serves as a helpful way of representing the behaviour of modelled agents. Because it is axiomatic, it is clear when the situation we want to analyse does not match the model, and in those cases we must use our judgement to establish both how helpful and how limited its application will be. In contrast, the use case for EE is not obvious, and it remains unclear whether it is supposed to be a normative theory offering individuals advice, a descriptive theory, or some combination of the two.

References

Carr, Peter, and Umberto Cherubini. 2020. Generalized Compounding and Growth Optimal Portfolios: Reconciling Kelly and Samuelson. Working paper, April 6. **Link**

Ford, Matthew C., and John A. Kay. 2023. The Limitations of Growth-Optimal Approaches to Decision Making Under Uncertainty. *Econ Journal Watch* 20(2): 314–334. **Link**

Hulme, Oliver, Arne Vanhoyweghen, Colm Connaughton, Ole Peters, Simon Steinkamp, Alexander Adamou, Dominik Baumann, Vincent Ginis, Bert Verbruggen, James Price, and Benjamin Skjold. 2023. Reply to "The Limitations of Growth-Optimal Approaches to Decision Making Under Uncertainty." *Econ Journal Watch* 20(2): 335–348. **Link**

Latané, Henry A. 1979. The Geometric Mean Criterion Continued. *Journal of Banking and Finance* 3: 309–311. **Link**

Meder, David, Finn Rabe, Tobias Morville, Kristoffer H. Madsen, Magnus T. Koudahl, Ray J. Dolan, Hartwig R. Siebner, and Oliver J. Hulme. 2021. Ergodicity-Breaking Reveals Time Optimal Decision Making in Humans. *PLoS Computational Biology* 17(9): e1009217. **Link**

Peters, Ole. 2019. The Ergodicity Problem in Economics. *Nature Physics* 15: 1216–1221. **Link**

Peters, Ole, and Alexander Adamou. 2021. The Time Interpretation of Expected Utility Theory. Working paper, February 28. **Link**

Peters, Ole, and Murray Gell-Mann. 2016. Evaluating Gambles Using Dynamics. *Chaos* 26: 023103. **Link**

Samuelson, Paul A. 1971. The "Fallacy" of Maximizing the Geometric Mean in Long Sequences of Investing or Gambling. *PNAS* 68(1): 2493–2496. **Link**

Skjold, Benjamin, Simon Steinkamp, Colm Connaughton, Oliver Hulme, and Ole Peters. 2023. Are Risk Preferences Optimal? Working paper, September 7. **Link**

Vanhoyweghen, Arne, Brecht Verbeken, Cathy Macharis, and Vincent Ginis. 2022. The Influence of Ergodicity on Risk Affinity of Timed and Non-Timed Respondents. *Scientific Reports* 12(1): 3744. **Link**

About the Authors

Matthew Ford read History and Economics at St. John's College, Oxford, and then worked with John Kay and Mervyn King as a research assistant on their book *Radical Uncertainty*. Having got interested in decision theory, he returned to St. John's to write *Intellectual Capital*, an academic and financial history of the College, and to investigate how useful decision theories were in understanding the actual situation the College found itself in. His email is matthewcford@icloud.com.

John Kay has been a Fellow of St. John's College, Oxford, since 1970. In 1979 he became research director and then director of the Institute for Fiscal Studies. In 1986 he founded London Economics, an economic consultancy. He was the first dean of Oxford's Saïd Business School and has held chairs at London Business School, the University of Oxford, and the London School of Economics. He is a Fellow of the British Academy and the Royal Society of Edinburgh. His email address is johnkay@johnkay.com.

Go to archive of Comments section
Go to March 2024 issue