ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Neurobiology

# Axiomatic methods, dopamine and reward prediction error

Andrew Caplin and Mark Dean

The phasic firing rate of midbrain dopamine neurons has been shown to respond both to the receipt of rewarding stimuli, and the degree to which such stimuli are anticipated by the recipient. This has led to the hypothesis that these neurons encode reward prediction error (RPE)—the difference between how rewarding an event is, and how rewarding it was expected to be. However, the RPE model is one of a number of competing explanations for dopamine activity that have proved hard to disentangle, mainly because they are couched in terms of latent, or unobservable, variables. This article describes techniques for dealing with latent variables common in economics and decision theory, and reviews work that uses these techniques to provide simple, non-parametric tests of the RPE hypothesis, allowing clear differentiation between competing explanations.

**Addresses**
Center for Experimental Social Science, Department of Economics, New York University, 19 West 4th Street, NY 10012, United States

Corresponding authors: Caplin, Andrew (andrew.caplin@nyu.edu) and Dean, Mark (mark.dean@nyu.edu)

**Current Opinion in Neurobiology** 2008, **18**:197–202

This review comes from a themed issue on
Cognitive neuroscience
Edited by Read Montague and John Assad

Available online 12th August 2008

0959-4388/$ – see front matter

**DOI** 10.1016/j.conb.2008.07.007

## Introduction

The reward prediction error (RPE) model has become the dominant paradigm for explaining the behavior of the neurotransmitter dopamine. It asserts that the phasic firing rate of midbrain dopamine neurons encodes the difference between the predicted and experienced 'reward' of an event [1,2•,3]. This signal is then used as a component of a reinforcement learning system that attaches values to different actions and thus guides choice behavior [4••–6]. Yet the RPE hypothesis remains controversial: Other explanations of dopamine activity include the 'salience' hypothesis (dopamine responds to how salient an event is) [7,8], the 'incentive salience' hypothesis (which differentiates between how much something is 'wanted' and how much something is 'liked') [9], and the 'agency' hypothesis (sensory prediction errors (with crude valence information) are used to reinforce the discovery of agency and novel actions) [10].

One reason that these hypotheses have proved difficult to disentangle is the current treatment of their 'latent' model elements, or variables that are not directly observable. Concepts such as 'rewards', 'predictions', 'incentive salience', 'salience', and 'valence' cannot be measured directly; they can only be identified through a relationship to something we *can* observe. This makes it difficult to test models that make use of such variables. Most current experimental analyses of the RPE hypothesis [11–18] get round this problem by adding to the RPE hypothesis specific assumptions about the nature of rewards and prediction that combine to give the RPE model testable implications. Typically, realized reward is assumed to be linearly related to some stimulus (such as money or fruit juice), while predictions are assumed to be driven by the temporal difference reinforcement learning model from computer science [19]. The time path of the RPE for a particular experiment can then be specified, and one can identify the extent to which this time series is correlated with activity in relevant regions of the brain. Strong correlation is taken as evidence in support of the RPE model.

There are five interrelated problems with this type of test [20]. First, they test both the broad RPE hypothesis and the auxiliary assumptions about the nature of rewards and predictions. Second, because of the flexibility of these auxiliary assumptions, it is difficult to provide a categorical rejection of any particular model of dopamine activity, or even to know whether different theories do make different predictions. Third, in practice, the various alternative models tend to produce predictions that are highly correlated, making it difficult to differentiate between them using regression techniques. Fourth, even if one does use statistical techniques to pick a 'winner' from the above models, this only tells us that this model is the best of the ones considered, not that it is a 'good' description of the data in a global sense. Fifth, the approach does little to guide model development in the face of a rejection by the data.

These problems do not derive from the use of latent variables *per se*, which are a valuable part of the modelers toolkit. Rather they stem from use of highly parameterized auxiliary assumptions, coupled with regressions, to test the resulting models. In order to address this, a new methodology has recently been proposed for testing the RPE hypothesis [21••,22•] derived from techniques commonly used in economics and decision theory for the modelling of latent variables. This approach largely overcomes the problems discussed above by discarding the need for auxiliary assumptions that link latent, or unobservable, concepts (such as reward) to observable variables (such as amount of fruit juice). Instead, it asks the question "if 'experienced' and 'predicted' rewards are completely unobservable, how can we test whether dopamine is encoding an RPE signal?". The latent variables in the RPE model are defined only in relation to our variable of interest —dopamine activity. In this manner, the entire class of RPE models can be characterized by a small number of empirical rules, or 'axioms'. These axioms are easily testable and provide stark and simple qualitative predictions that *must* hold for the RPE theory to be true for *any* definition of rewards and predictions. In other words, they provide a guide as to whether the latent concepts inherent in the RPE model provide a useful way of thinking about dopamine activity. Moreover, they provide a clear guide to how the RPE model differs from other explanations for dopamine activity.

This paper reviews the way in which these axiomatic techniques can be used to test the RPE model. Throughout, we refer to the variable of interest as dopamine. However, this technique can be used to test whether *any* candidate data series encodes an RPE signal, be it direct observations of dopamine spike rates, fMRI measurement of activity in the ventral striatum, or something else entirely. Thus, the thrust of this review is not to provide evidence for or against the RPE model of dopamine but to provide techniques for determining whether or not a particular signal can be thought of as an RPE encoder.

This 'axiomatic' approach to modelling latent variables has proved valuable within economics. We argue that it may be of general value to neurobiologists, not just with respect to dopamine activity. In particular, it has the potential to enhance the interaction between experiment and theory, thereby speeding up the process of scientific discovery.

## The axiomatic approach to reward prediction error

In its most basic form, the RPE hypothesis states that dopamine activity encodes the difference between the experienced and predicted reward of an event. Unfortunately, 'reward' itself is inherently unobservable; the amount of fruit juice you give a monkey is observable, the amount of money you give someone is observable. But these are not 'reward'; they are things that one might assume would lead to the feeling of 'reward'. Therefore, without a working definition of how 'experienced' or 'predicted' reward relate to things we can observe, this theory is incomplete. One way round this problem is to add to the theory a definition of rewards and predictions, which links them to something that we can directly measure. However, any test of the hypothesis is then a joint test of both the original RPE hypothesis and the (often strict, parametric) assumptions that operationalize the relevant latent, or unobservable, variables.

A typical methodology for this class of test [14,18] is to observe subjects (either monkey or human) making repeated choices between various options that lead to prizes (fruit juice, or money) according to some stochastic schedule. Reward is assumed to be linearly related to prize magnitude (i.e. how much fruit juice or money the subject receives) while predictions are assumed to follow some sort of reinforcement learning rule, such as the temporal difference (TD) algorithm, possibly calibrated using behavior. Under these assumptions, RPE is 'observable' (the difference between the magnitude of received prize, and expected prize as determined by the TD model) and can be correlated with brain activity in order to test the RPE model.

An alternative approach is to treat reward and predictions of subjects as completely unobservable and ask whether the theory still has any testable predictions. The only thing that the RPE theory claims about reward is that dopamine is positively related to experienced and negatively related to predicted reward. It makes no claim about how reward is related to fruit juice, or how predictions are formed. The question is therefore whether the theory's central claims are enough to put testable restrictions on dopaminergic activity. If the answer is 'yes', then we can construct tests of the RPE theory that are completely non-parametric, and do not rely on any auxiliary assumptions.

Figure 1 describes a set of three such rules for the RPE model in the simplest possible environment [21••]. The idealized data set that we consider comprises observations of dopamine activity when a subject receives various prizes drawn from well defined probability distributions, or 'lotteries', over such prizes. In a typical example, the subject will either win or lose $5 depending on the flip of a fair coin. This environment has the advantage of abstracting from the need to model learning, making the resulting model very simple. In such an environment, the RPE hypothesis can be characterized by three broad statements, or axioms. First, the ranking of different *prizes* in terms of dopamine activity must be independent of the lottery those prizes are received from—for a fixed lottery, better prizes should always lead to higher dopamine release (Axiom A1). Second, the ranking of different *lotteries* in terms of dopamine activity must be independent of the prizes received from those lotteries - for a fixed prize, better lotteries should always lead to lower dopamine release (Axiom A2). Third, if a prize is fully anticipated, then dopamine activity has to be independent of what the prize actually is (Axiom A3). These three conditions are *necessary* and *sufficient* for the RPE model; if they do not hold, then there is no possible definition of experienced and predicted reward that can make the RPE model fit the data. If they do hold, then we can find *some* way of assigning experienced and predicted reward such that dopamine encodes RPE with respect to these definitions.

This does not imply that the RPE model is the only one that satisfies these three axioms; it could be that some other potential explanation for dopamine activity also implies these properties (though, for example, the salience hypothesis would not, as we discuss below). In fact, one of the advantages of the axiomatic approach is that it allows one to determine the extent to which different models have different implications for a particular data set. Were one to find another model that implied A1–A3, the above result tells us that it would be impossible to falsify the RPE model at the expense of this new model, as any falsification of RPE would imply a violation of one of A1–A3, and so would also falsify the new model.

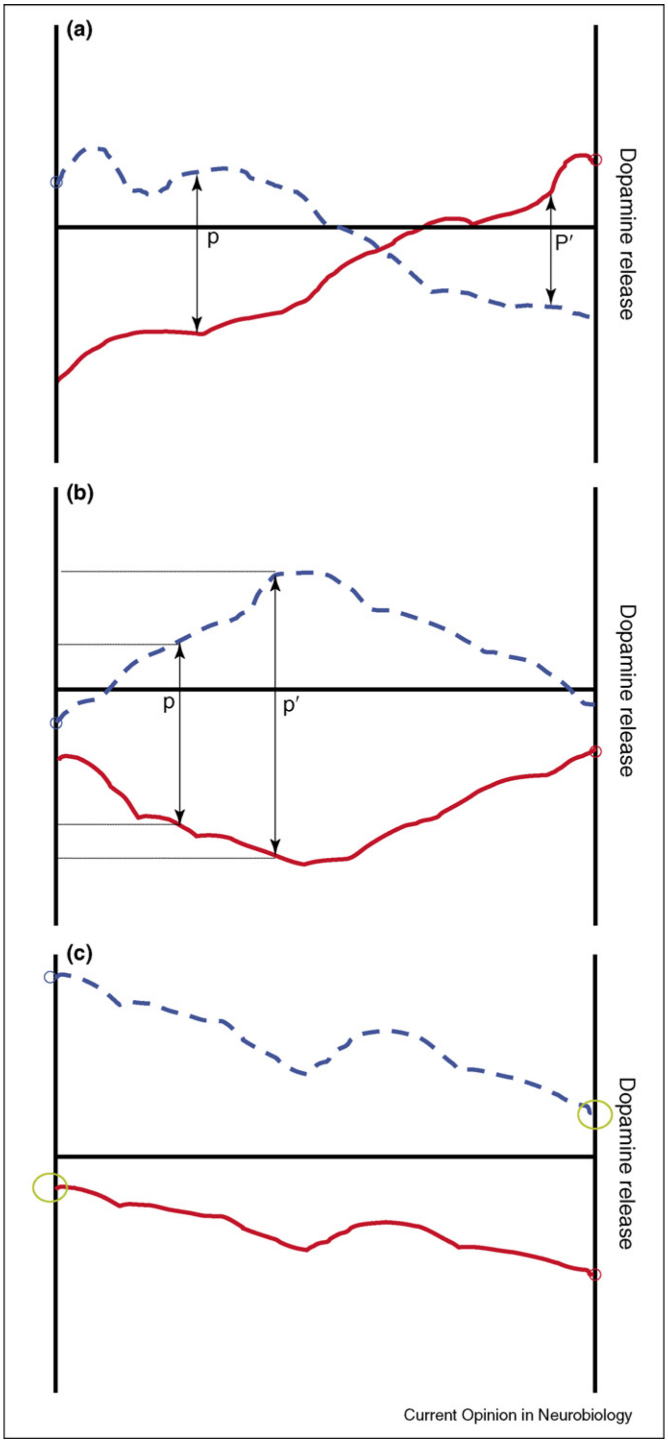
Again, it should be noted that while we describe these tests as being carried out on 'dopamine activity', they can be applied to any data series that is purported to encode an RPE signal, be it fMRI data of activity in the ventral striatum [15] or single unit recording from midbrain dopamine neurons in primates [12].

## Advantages of the axiomatic approach

The axiomatic approach to testing the RPE model has a number of advantages over the more traditional regression-based tests. First and foremost, because it defines 'experienced' and 'predicted' reward only non-parametrically, and only by their relation to the variable of interest, we provide a test of the *entire class* of RPE models —if these axioms are violated, then it is not because of some incorrect parametric assumption, or an incorrect model of reward, or how predictions are made. It means that there is something fundamentally wrong with the entire basis of the RPE model. In this sense, these tests are *weaker* than existing tests of the RPE hypothesis that impose a specific functional form for reward, and an explicit model for learning.

Second, this approach provides an easily testable set of conditions that divide the universe of possible observations into those that are in line with the RPE model

**Figure 1**



Current Opinion in Neurobiology

The axioms that characterize the RPE model can be illustrated graphically in the case in which the experiment has only two possible prizes. In this case, the set of all possible lotteries can be represented by a single number: The probability of winning prize 1 (the probability of winning prize 2 must be 1 minus the probability of winning prize 1). This forms the *x*-axis of these figures. We represent dopamine activity using two lines—the dashed line indicates the amount of dopamine released when prize 1 is obtained from each of these lotteries, while the solid line represents the amount of dopamine released when prize 2 is obtained from each lottery. (**Panel a**) A violation of A1: When received from lottery p, prize 1 leads to higher dopamine release than does prize 2, indicating that prize 1 has higher experienced reward. This order is reversed when the prizes are realized from lottery p′, suggesting prize 2 has higher experienced reward. Thus a DRPE representation is impossible. (**Panel b**) A violation of A2: Looking at prize 1, more dopamine is released when this prize is obtained from p′ than when obtained from p, suggesting that p has a higher predicted reward than p′. The reverse is true for prize 2, making a DRPE representation impossible. (**Panel c**) A violation of A3: The dopamine released when prize 1 is obtained from its sure thing lottery is higher than that when prize 2 is obtained from its sure thing lottery.

and those that are not. This allows the model to be tested in a *global* sense—either the model is an accurate description of the data, or it is not, rather than a *relative* sense—the model is a better description of the data than others we have considered. Furthermore, it forces the modeler to be explicit about exactly what it is their model says for a particular data set.

Third, by characterizing theories in this way we can draw clear demarcations between different models, allowing us to understand how one might test between them. For example, the RPE hypothesis states that the ranking of prizes in terms of dopamine release must be independent of the lottery that they are received from. Thus, if we consider two lotteries—one that has a 1% chance of winning $5 (and 99% chance of losing $5) and another that gives a 99% chance of winning $5 (and a 1% chance of losing $5), then the observation that more dopamine is released when $5 is won than when $5 is lost from the former implies that the same must be true in the latter. This would clearly not be true for any model of 'salience', since the salience of an event is related to its rarity [7].

This advantage is notable in comparison to the 'regression-based' approach described above, in which auxiliary assumptions are used to operationalize a theory. In such an approach, because these auxiliary assumptions are not central to the theory, researchers can try ever more elaborate relations between observable and latent variables in order to best fit the data (for example, could reward be a quadratic function of prize magnitude? Or a power function?). This flexibility can make differentiating between different models by statistical means very difficult.

A fourth advantage is that the axiomatic approach allows easily for a hierarchical testing structure for a particular model. For example, the above model can be refined to include the hypothesis that dopamine responds to the *difference* between experienced and predicted reward in the strict sense (i.e. experienced minus predicted reward), or that predicted reward is the mathematical expectation of the experienced reward of a lottery [21••]. Axiomatic representations exist for these nested models, and one can therefore test how far the RPE hypothesis can be extended.

Finally, the outlined approach offers guidance on what to do in the face of a rejection by the data. As one knows which particular axiom has been violated, one can adjust the model in precisely the right way to accommodate the data. Examples of this type of interaction between data and theory abound in economics and are discussed more below.

## An axiomatic approach to behavioral neurobiology

The axiomatic approach provides an alternative way of characterizing models couched in terms of latent variables, and one that has proved popular within economics and decision theory [23,24]. As neurobiology begins to model the processes that underlie choice and decision-making, we believe the same techniques may prove to be equally useful. To understand why, it is instructive to consider examples from economics in which the precision that axioms offer has spurred the joint development of theory and experimentation.

The classic example is the theory of 'utility maximization', which has been benchmark model of economic behavior almost since the inception of the field. Yet it was left to Paul Samuelson in 1938 to ask the question: ''Given that we do not observe 'utility', how can we test whether people are utility maximizers?'' [25]. The answer to this question is that their choices must obey the so-called 'Weak Axiom of Revealed Preference' (WARP), which basically states that if one chooses some option $x$ over $y$, then one cannot at some other point choose $y$ over $x$. Simple as this is, it turns out that WARP is the only testable prediction of utility maximization. In the wake of this pivotal insight, the axiomatic approach has been successfully used within economics to characterize and test other theories that share with utility maximization that they involve 'latent' variables.

Von Neumann and Morgensten [26] and Savage [27], extended the utility maximization hypothesis to the realm of risk and uncertainty, axiomatically modelling the behavior of agents who maximize *expected* utility. The failure of these axioms, demonstrated in famous experiments by Allais [28] and Ellsberg [29] led to the development of more sophisticated models of behavior, such as rank-dependent expected utility [30] and ambiguity aversion. [31], that are now used in analyzing economic policy [32,33].

Behavioral neurobiology shares with economics the assumption that variables of interest are influenced by latent variables that are not subject to direct empirical identification, such as rewards, beliefs, emotions, and motivations. While preliminary data analysis and research may be well guided by such intuitive constructs, when time for formalizing models comes, axiomatic methods have much to add. In particular, they discipline the introduction of new constructs into the theoretical canon. The axiomatic method calls for consideration of precisely how inclusion of these new concepts impacts observations of some idealized data set. If their inclusion does not expand the range of predicted behaviors, they are not seen as 'earning their keep'. If they do increase the range of predictions, then questions can be posed concerning whether such observations are commonly observed. Thus, the axiomatic method can be employed to ensure that any new latent variable adds new empirical predictions that had proven hard to rationalize in its absence.

Another reason that the axiomatic approach has proven so fruitful in economics is that our theories are very far from complete in their predictive power. There is little or no hope of constructing a simple theory that will adequately summarize all relevant phenomena: systematic errors are all but inevitable. The axiomatic method adds particular discipline to the process of sorting between such poorly fitting theories. In essence, the key to a successful axiomatic agenda involves maintaining a close connection between theoretical constructs and empirically observable phenomena. We believe that much of behavioral neurobiology may be similar in this respect.

Unfortunately, axiomatic methods have in the past earned something of a bad name in psychological theory, in which their use has not been associated with a progressive interaction between theory and data. We see no need for this pattern to continue. As we illustrate in the case of dopamine, we see use of axiomatic methods not as an end in and of itself, but rather as a guide to drive experimentation in the most progressive possible directions. Used in this manner, axiomatic modelling techniques strike us as an intensely practical weapon in the neuroscientific arsenal.

## Acknowledgements

The authors would like to thank Paul Glimcher, P. Read Montague, Paul Dean and Peter Redgrave for their helpful comments, and Ernst Fehr and Peter Bossaerts for their encouragement and suggestions.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of special interest

1. Wolfram S, Apicella P, Lungberg T: **Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task**. *J Neurosci* 1993, **13**:900-913.

2•. Mirenowicz J, Schultz W: **Importance of unpredictability for reward responses in primate dopamine neurons**. *J Neurophysiol* 1994, **72(2)**:1024-1027.
Establishes the role of beliefs in determining dopamine activity, by showing that phasic dopamine responses to a liquid reward occur only if that reward is unexpected.

3. Wolfram S, Dayan P, Montague PR: **A neural substrate of prediction and reward**. *Science* 1997, **275**:1593-1599.

4••. Montague PR, Dayan P, Sejnowski TJ: **A framework for mesencephalic dopamine systems based on predictive Hebbian learning**. *J Neurosci* 1996, **16**:1936-1947.
Introduces the idea that dopmine might encode a reward prediction error signal of the type used in reinforcement-style learning models, and develops a formal model of such a learning system which is consistent with the known physiological facts.

5. Waeltl P, Anthony D, Wolfram S: **Dopamine responses comply with basic assumptions of formal learning theory**. *Nature* 5 July 2001, **412**:43-48.

6. Montague PR, Hyman SE, Cohen JD: **Computational roles for dopamine in behavioural control**. *Nature* 2004, **431**:760-767.

7. Zink CF, Pagnoni G, Martin ME, Dhamala M, Berns G: **Human striatal response to salient nonrewarding stimuli**. *J Neurosci* 2003, **23**:8092-8097.

8. McClure SM, Daw N, Montague PR: **A computational substrate for incentive salience**. *Trends Neurosci* 2003, **26(8)**:423-428.

9. Berridge KC, Robinson TE: **What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience?** *Brain Res Rev* 1998, **28**:309-369.

10. Redgrave P, Gurney KN: **The short-latency dopamine signal: a role in discovering novel actions?** *Nature Reviews Neuroscience* 2006, **7**:967–975.

11. Montague PR, Berns GS: **Neural economics and the biological substrates of valuation**. *Neuron* 2002, **36**:265-284.

12. Bayer H, Glimcher P: **Midbrain dopamine neurons encode a quantitative reward prediction error signal**. *Neuron* 2005, **47**:129-141.

13. Bayer H, Lau B, Glimcher P: **Statistics of midbrain dopamine neuron spike trains in the awake primate**, *J Neurophysiol* 2007, **98**:1428–1439.

14. O'Doherty J, Dayan P, Friston KJ, Critchley HD, Dolan RJ: **Temporal difference models account and reward-related learning in the human brain**. *Neuron* 2003, **38**:329-337.

15. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ: **Dissociable roles of ventral and dorsal striatum in instrumental conditioning**. *Science* 2004, **304**:452-454.

16. O'Doherty J, Buchanan TW, Seymour B, Dolan R: **Predictive neural coding of reward preferences involves dissociable responses in human ventral midbrain and ventral striatum**. *Neuron* 2006, **49**:157-166.

17. Daw N, O'Doherty JP, Dayan P, Seymour B, Dolan RJ: **Polar exploration: cortical substrates for exploratory decisions in humans**. *Nature* 2006, **441**:876-879.

18. Li J, McClure SM, King-Casas B, Montague PR: **Policy adjustment in a dynamic economic game**, *PlosONE* **1(1)**: e103. doi:10.1371/journal.pone.0000103.

19. Sutton RS, Barto AG: Reinforcement Learning: An Introduction Cambridge, USA: MIT Press; 1998.

20]. Caplin A, Mark D: **Axiomatic neuroeconomics**, *Neuroeconomics, Decision Making and the Brain*. Edited by Glimcher P., Camerer C., Fehr E., Poldrack R., New York: Elsevier 2008, in press.

21••.Caplin A, Dean M: **Dopamine, reward prediction error, and economics**. *Quart J Econ* May 2008.
The authors derive an axiomatic foundation for the reward prediction error theory of dopaminergic activity, the first paper to use axiomatic methods to approach neuroeconomic data. The paper shows that the basic RPE model is equivalent to three easily testable axiomatic conditions.

22•. Caplin A, Dean M, Glimcher P, Rutledge R: **Measuring beliefs and experienced reward: a neuroeconomic approach**. NYU CESS/ CNE working Paper 2008.
This paper uses fMRI data on brain activity in the nucleus accumbens from human subjects to test the axiomatic conditions of [21••] above. The authors find evidence to support all three axioms, suggesting that such a signal can be thought of as encoding an RPE signal.

23. Kreps D: Notes on the Theory of Choice. Colorado: Westview Press; 1988.

24. Fishburn P: Utility Theory for Decision Making. New York: John Wiley and Sons; 1970.

25. Samuelson P: **A note on the pure theory of consumer's behaviour**. *Economica, New Series* 1938, **5(17)**:61-71.

26]. von Neumann J, Morgenstern O: *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press 1944 (sec. ed. 1947).

27]. Savage: *The Foundations of Statistics*. New York: Wiley; 1954.

28. Allais M: **Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine**. *Econometrica* 1953, **21**:503-546.

29. Ellsberg D: **Risk, ambiguity, and the savage axioms**. *Quart J Econ* 1961, **75**:643-669.

30. Quiggin J: **A theory of anticipated utility**. *J Econ Behav Org* 1982, **3(4)**:323-343.

31. Gilboa I, Schmeidler D: **Maxmin expected utility with a non-uniqueprior**. *J Math Econ* 1989, **18**:141-153.

32. Mukerji S, Tallon J-M: **Ambiguity aversion and incompleteness of financial markets**. *Rev Econ Stud* Oct, 2001, **68(4)**: 883-904.

33. Epstein Larry G, Martin S: **Ambiguity, information quality, and asset pricing**. *J Fin Am Fin Assoc* 2008, **63(1)**:197-228 02.