

A Paradox of Risk Management

Alex. Arthur

The two primitive models that underlie rational risk management methods are not equivalent, but are complementary. As commonly formulated, they generate a pair of complementary paradoxes – a circumstance which formally renders the concept of 'rational risk management' incoherent.

This paper explicates the paradoxes, and considers their possible relationships with other self-referential paradoxes that appear in the philosophies of language, truth, knowledge and logic. These include the liar paradox, the paradoxes associated with formulating reliable theories of truth and knowledge (among these Goodman's paradox), and, more speculatively, Gödel's paradoxes.

Finally the paper will consider a possible resolution of the paradox that, in a qualified way, rescues the rationality of risk management and has a bearing on how other self-referential paradoxes should be approached. An aspect of this resolution, however, is to undermine some of our more comforting intuitions about the relationship between language and reality.

Introduction – Two models of Risk Management

To make rational choices in an uncertain world, we try to estimate the range and probability of the future outcomes of these choices. The formal methods for doing this depend upon two different approaches to estimating probabilities, which I will call (1) the engineering approach and (2) the statistical approach.

The engineering approach depends upon the construction of a formal model of the system whose behaviour we wish to predict, and deducing the determinate and indeterminate aspects of that behaviour from our description of the model. A simple example is the engineering model of a normal die which has, as one of its consequences, the $1/6^{\text{th}}$ probability that a tossed die will land on any given one of its sides.

The statistical approach depends on the measurement of regularities that exist in a known example of the type of system whose behaviour we wish to predict, and the determination of the range and probabilities of its future behaviours from these measurements. Studies of the behaviour of large human populations are typical of this approach.

With respect to risk management, either of these methods, so long as it is reliable, will provide the information about probabilistic outcomes required for the calculation of expected values, and for making economically rational choices. This paper will not consider the practical aspects of constructing the appropriate engineering models, or carrying out the appropriate statistical studies. Neither will it consider the contentions surrounding categories such as 'economic rationality' or 'rational choice' or 'economic value'.

Instead, it will focus on formal features of these two approaches as they are represented here, and consider their rationality in their own terms. It will then consider the consequences of the outcomes of this investigation for some other related problems.

The rationality of system modelling

The engineering approach to calculating probabilities is formally unexceptionable, so long as it remains abstract. The mathematical model of the die from which the probable outcomes of tossing it can be calculated embodies no mysteries. A paradox associated with using this approach for risk management in the 'real world' is easily constructed, however.

In order to use engineering models to calculate real probabilities, we must construct these models in a way that ensures that their behaviour closely matches the behaviour of the real systems that we wish to manage. This is a reasonably straightforward business for simple physical systems, such as dice and pennies, which is why these simple systems appear so often as examples in textbooks on probability. Consider, however, how we might use the engineering approach to model the risk that some specific engineering model does not, in some relevant but presently invisible way, correctly correspond to the real system whose behaviour it is being used to predict. In order to rationally manage this risk, we would need to construct a new model which formally represented both the original model and the system we are trying to predict, so that we could evaluate the possibility that they might

diverge. This new model would either (a) suffer from exactly the same defect as the original, since it would be trying to model the same underlying reality or (b) be a replacement for the original, since it would clearly incorporate a demonstrably better model of the risky reality than the original did. In case (b), however, we are simply left with the further issue of managing the risk that the *new* model was unreliable.

In other words, not only is there always some underlying risk that the engineering model in use is unreliable, but if engineering approaches to risk management are the only approaches available, then that residual risk cannot be rationally managed, since it is exactly the risk that the risk management model is inapplicable. The consequences of this are fairly serious, because any 'common sense' judgements that the risk of a good model being unreliable were small must be either (a) intuitive and unsupportable or (b) based on some rational modelling process, and therefore exposed to the same paradox. If the engineering approach to probability calculation were the only one available for real world risk modelling, then we would be exposed to an incalculable probability that our predictions were completely unreliable.

At this stage, a more uncommon sense might intervene and point out that; after all, we can test our engineering model by doing some scientific observations in the real world.

The rationality of statistical prediction

Clearly, we can and do make reliable predictions about the probable ranges of real world behaviour, and we are often able to do this when we have very poor formal models of the behaviour we are trying to predict. Instead of modelling the behaviour, we simply observe it over wide enough number of cases so that we can establish within certain ranges of certainty how the system generally behaves. Statistical studies of large complex populations often reveal very reliable regularities of this type.

In addition to this, we can use these statistical studies to gauge the reliability of our engineering models, where these *do* exist. Unfortunately, however, the statistical method for establishing real world probabilities also generates a paradox.

Clearly even demonstrably random systems will produce spurious apparent regularities (they would not be truly random if they did not). Tossing coins, rolling dice, and spinning roulette wheels may all produce runs of heads, sixes, or blacks which challenge our confidence that they are not loaded or crooked. Consider the question of how we would use a statistical approach to reassure ourselves that any regularities that we observed in a population of measurements was not spurious in this way. Continuing to make measurements could not comfort us, because we could never eliminate, *nor statistically evaluate*, the residual probability that we were just on a lucky (or unlucky) run.

In real cases, of course, we comfort ourselves by constructing underlying explanatory models to support these regularities. We address the residual formal incoherence of the statistical method by appealing to the engineering method, and vice versa.

In consequence, if the statistical and engineering methods exhaust our repertoire of rational risk management techniques, then we must accept that we are always facing completely unmanageable risks of utterly indeterminate size.

Assumptions and practical consequences

The underlying assumptions

Before drawing radical conclusions from the fact that this compound paradox can be constructed, it is worth considering the assumptions that underlie it. Two important disputable assumptions are the following:

- (1) The engineering method and the statistical method are all the methods that are available for establishing probabilities in the real world.
- (2) Establishing real world probabilities is an essential component in any rational risk management regime.

The first assumption is perhaps the most problematic. After all, we manage many day to day risks in completely ad hoc or intuitive ways without resorting to sophisticated modelling or making complex measurements. If we were to call all of these methods irrational, we would render irrational most of our unreflective behaviour.

This objection to assumption (1) may fail, however, because when we come to argue for the rationality of our unreflective actions we do this by showing that however little we might have consciously reflected on them at the

time, we can show that these actions *would have been rational even after reflection*. In other words, if I lie in bed for a few more minutes in the morning and risk missing a bus, I can show this to be rational by doing the relevant statistical study of bus arrival times and the time it takes me to get dressed, and by constructing models of my behaviour, and the behaviour of the buses, which support these studies. My claim that lying in bed is rational is a claim that further research of this kind would support my decision with statistical and/or engineering argument that I was unlikely to miss my bus.

If 'engineering methods' are taken to mean 'valid forecasting arguments', and 'statistical measurement' is taken to mean 'valid observation', it is hard to see how any rational method of calculating synthetic probabilities can avoid being some combination of these. Whether or not we actually do the complete calculation, or undertake the observations, our assessments of probabilities will always be subject to appraisal against these standards.

The issue of whether there are unreflective actions of any other kind that can be shown to be rational in some other way has a role to play in a later argument, and it will be deferred until that argument is developed.

The second assumption, that calculating real probabilities is an essential part of risk management is more or less equivalent to the question whether calculating expected values is a necessary element in making economic choices. Notoriously, the issue whether some particular piece of human behaviour is subject to economic evaluation is the entry point to a conceptual quagmire. What seems difficult to avoid, however, is the conclusion that *if* a particular piece of behaviour is to be analysed as a choice and *if* that analysis incorporates a correct evaluation of the subject's utility function, then we can only call the choice a rational choice if it conforms to the optimum outcome available from the conjunction of that utility function and the known real constraints – including probabilistic ones. These are both big 'ifs', but the relationship between rational choice and probability can be determined without determining whether any particular action counts as a choice, or whether any particular actor is well-informed. In so far as rational risk management is a matter of making rational choices, the need to be able to establish probabilities in the real world seems unavoidable.

Practical consequences

A common misunderstanding about this kind of fundamental enquiry – a misunderstanding fostered by the delusions of philosophers since pre-Socratic times – is that radical conceptual conclusions should immediately be incorporated into the methods of real world practitioners. It is not part of the purpose of this paper to render practical risk management absurd.

In many scientific disciplines, absurdities are managed in an ad-hoc way. The infinities in quantum mechanical computations are cancelled out, or their locations avoided, in practical applications; while at the same time theoreticians ponder their meanings in pursuit of greater conceptual coherence. No one thinks we are deceived about the reliability of these slightly cooked sums, nor, on the other hand, that the cooking is rendered mathematically valid by the empirical results. Instead, both types of researcher continue to work on their disparate problems, each recognising they must proceed tentatively until the theoretical incoherence and the practical success can be reconciled.

Theoretical consequences

Classical contexts

In order to find a theoretical productive approach to the paradox outlined above, it is worth putting it in the context of some other similar paradoxes.

Because the construction of the paradox depends upon the application of a method to itself, we can call it a 'self-referential' paradox. Other self-referential paradoxes include the paradox of the Liar; Russell's paradox (or the class paradox); a set of paradoxes associated with theories of truth, meaning and knowledge; and Gödel's paradoxes concerning the completeness and consistency of arithmetic.

The paradox of the liar (Gupta, 2001) is well known, and has classical origins (Eubulides of Miletus). Consider the statement 'What I am now saying is untrue'. If it is true, then it is false, and vice versa. Note that a liar paradox can be constructed in any language that can be used to consider questions of the truth of statements in general (i.e. without restriction). Such a language needs to be able to make statements about the truth and validity of its own statements, and so can be used to construct 'This statement is untrue'. Another way of saying this is to say that any language that can be used to address general questions of truth, and indeed validity, must be inconsistent – i.e. can be used to construct contradictions. The rules of the language itself cannot rule them out. Another interesting thing about such languages is that they can also be used to construct statements which

are true, but whose truth cannot be demonstrated within the language. Consider the statement 'This statement cannot be proved to be true'. If it is true, then it cannot be proved to be true (because that's what its truth would entail). If it is false, then it can be proved to be true, so its falsehood is self-contradictory. The self-contradictoriness of its falsehood guarantees its truth, and its truth guarantees that it can't be proved. For a logician, incidentally, this outcome simply demonstrates that the concept of proof in ordinary language is not well defined - we are not able to formally specify what we mean by proof. Unfortunately, a consequence of Gödel's theorem (see below) is that any language rich enough to use for constructing scientific theories will have exactly the defects of inconsistency and incompleteness outlined here.

Russell's paradox (Frege, 1964) is equally easy to construct, and may be closely related to the liar paradox. It is seriously damaging to Frege's proposal to found arithmetic on the logic of set theory, because it demonstrates that the formal axioms of set theory can be used to generate a contradiction, and are therefore inconsistent. The paradox, in brief, is that although the concept of a set which is a member of itself makes sense (the set of large sets is a member of itself), the concept of a set which is not a member of itself does not make sense since the membership of the set of all such sets cannot be established. The set of all sets which are not members of themselves belongs to itself, since it is such a set, but does not, because of its definition.

There is a general paradox associated with theories of truth, which can also be constructed in any language that can make unconstrained truth claims. We might imagine that one thing a theory of truth could do for us would be to set out criteria for distinguishing true statements from false ones. The paradox arises when we try to test the truth of this theory. Since we have no concept of truth independent of the theory, any assessment of the theory is immediately question-begging.

A similar paradox arises with respect to theories of knowledge that incorporate criteria for establishing reliable knowledge. Again, the paradox arises when we ask how we establish that this theory is one we can reliably know.

There is also a related paradox associated with theories of meaning, which is slightly more elliptical, but is worth outlining here. Again, it depends upon having a theory of meaning that explains how the meaning of any particular statement of expression in a language can be determined. The paradox is generated by asking what the theory of meaning means - and we cannot, of course, determine this without understanding its meaning. This is not such a directly contradictory outcome, but it leaves the concept of meaning systematically un-decidable rather than clarifying it (as a theory of meaning might be expected to do).

Gödel's paradox (Smullyan, 2001; Gödel 1992) is a good deal more complicated in its construction than these others, but it depends ultimately on showing that arithmetic can be used to consider general questions of validity (or questions strictly isomorphic to these), and so must contain both self-contradictory and un-provable statements (as discussed above). What Gödel does is show that a unique number can be constructed for every possible logical statement, and that an arithmetical function can be defined which relates these numbers, and is equivalent to logical demonstration. In other words, if you feed the numbers for two logical statements into the function, you can, in principle, calculate *arithmetically* whether or not one of the statements demonstrates the validity of the other one. The computations are, for most proofs, unmanageable complex. This does not matter, however, since all that Gödel needs to do is to demonstrate that they make arithmetical sense. The method he uses is to show how numbers can be constructed for expressions and proofs in such a way that if one set of expressions proves another expression, then the Gödel number for the statement of the premises will be an arithmetical factor of the Gödel number for the statement of the whole proof. Since this is all that is needed to demonstrate that arithmetic is a language that can be used to address general questions of truth and validity, then it follows that it must be either inconsistent or incomplete.

As well as being a self-referential paradox, the risk management paradox is also an inductive paradox, in the same sense as Goodman's paradox. Goodman's paradox (Goodman 1983), directed against naive inductionism goes like this: If we have inductive evidence for some object's having a quality A, then we have equally good evidence for it's having quality B, which has the character of being indistinguishable from A until *t* (some time in the future) and different thereafter. More colloquially, and adopting a neologism coined by Goodman, if we have evidence of parrots being green, then this is equally good evidence that parrots are *grue*, where *grue* is a colour indistinguishable from green *now*, but which becomes indistinguishable from blue after time *t*. Notice, again, that Goodman's criticism is of the logic of induction as it is classically presented, rather than against practical inductive reasoning.

Self-referential paradoxes are generated within certain validity testing systems when they are used to adjudicate on their *own* validity. If mathematics can be reduced to the logic of classes, and the logic of classes is a part of mathematics, then we should expect paradoxical results. For theories of truth, knowledge, and meaning, the results are generated directly. If we try reducing logic to arithmetic, rather than arithmetic to logic, we again have a self-referential validity testing system, and should expect some bizarre consequences. These little black

holes in the logical firmament are also black holes of meaning and referentiality. With the class paradox, the apparent meaning of 'the class of all classes which are not members of themselves' seems perfectly clear and well constructed, but the *reference* of this description is systematically elusive – the very act of trying to give it a reference ensures that that reference will be the wrong one. Setting out methods of determining meaning, when the methods themselves must have meaning in order to be methods, immediately makes meaning completely indeterminable. Instead of indeterminable references (as in the class paradox) we have indeterminable meanings. If we seek to demonstrate the consistency of a system some part of which can be shown to be logically isomorphic with any system capable of demonstrating consistency, we make the issue of the consistency of that system immediately indeterminable. When we try to figure out what we're talking about when we talk about what we're talking about, then we have no way of figuring out what we're talking about, except to say that it's what we're talking about (whatever that is).

Curiously, inductive paradoxes can also be shown to result from playing a game with the possibility of valid demonstration, (albeit in the physical rather than the logical sciences). Giving the 'grue' paradox a chronological inductive setting actually obscures its scope. A more general version of the paradox can be generated by *any* attempt to make completely general empirical statements on the basis of relatively local evidence, because any local evidence will count equally as evidence for qualities that are like this locally, but like something else elsewhere. Another way of putting this is to say that the hypothesis that local is typical is an untestable empirical hypothesis upon which the whole of general science depends. This is precisely the hypothesis underlying our dependence on statistical determinations of probability, when we assume that the patterns that we find are not just random local anomalies. Indeed, a possible response to Goodman – that we predict on the basis of hypothesised mechanisms, and not just on the basis of regularity (there is no mechanism that predicts grueness, while there is a mechanism predicting greenness) looks very like our switch to the engineering method of calculating probabilities.

A little reflection, however, reveals that even 'local is typical' is not sufficiently general to underpin empirical science. 'Local' and 'typical' are both rather vague, contingent terms to work at this level of abstraction. Local can't be spatially local, or temporally local (despite Goodman's exposition) because we can observe spatially and temporally distant things. We cannot 'see' into the future but our computations of what will happen there are not intrinsically less reliable than our arguments for past events, or events which occur beyond our immediate sensory bubble. A more general and more directly defensible underlying hypothesis, and one which goes some way towards dealing with Goodman, is the simple hypothesis that the world is, after all, the kind of place we can theorise about. If things could be grue as easily as green, this would not be the case. This hypothesis is more general because it does not refer to contingent categories such as space and time, but only to the possibility of theorising. In effect, it refers only to the possibility of talking about the world at all. Some statement like 'we can talk about the world' needs to be true in order for us to be able to construct scientific theories.

In addition to this, and to ultimately justify this particular line of thought, a statement such as 'we can talk about the world' does not need empirical justification. It's denial, being a statement about the world, is self-refuting *a priori*.

Reconstructing the validity of risk management

It is appropriate, now, to work back through the jungle of self-referential paradoxes to a point where we can eliminate the inconvenient black hole in the theory of risk management.

First of all, it is worth observing that the paradoxes that refer explicitly to the possibility of fundamental linguistic categories – truth and meaning – and fundamental epistemic ones - such as knowledge - cannot be used to argue for the vacuity of these categories. The incoherence of the possibility of a theory of truth cannot be used to demonstrate the incoherence of the concept of truth itself. After all, we would want to say that it is *true* that the possibility of a theory of truth is incoherent. In all these cases, in fact, statements that deny the coherence of the problematic categories themselves are just as paradoxical as statements that attempt to elucidate them. If we conclude from the impossibility of theories of truth, for instance, that truth itself is an incoherent concept, then we deny the truth of our conclusion. If we conclude from the impossibility of theories of knowledge that we can have no such thing as reliable knowledge, then we deny our capacity to know that this is the case. If we conclude from the impossibility of theories of meaning that meaning itself is impossible, we deny the possibility of meaning that meaning is impossible.

There is, of course, no general difficulty about being able to do things about which we cannot theorise. Not many people learn to ride a bicycle just by memorising a set of instructions, or developing an articulate physical theory of bicycle control. Most people can throw balls more competently than they can manage the differential calculus needed to design a ball-throwing machine. If we could not talk without first being able to articulate a theory of language, we'd still be tiger snacks on the savannah.

A reasonable starting point then, is the compound observation that we can talk about the world and that we don't need to be able to say how. There are a lot of things we can do without being able to say how, and we have no way of denying our ability to talk without, by demonstrating it, falsifying our denial.

With respect to the two models of rational risk management, the supposition that the world is radically unpredictable is equivalent to the supposition that the world is impossible to talk about – as, indeed, a radically unpredictable world would be. So while we can talk, we can rationally manage risk. If we find we can't, we won't be discussing it.

Bibliography:

Davis, M (ed.) *The Undecidable* (Raven Press, New York, 1964)

Frege, Gottlob 'The Russell Paradox', in Frege, Gottlob, *The Basic Laws of Arithmetic*, Berkeley: University of California Press, 1964, 127-143. Abridged and reprinted in Irvine, A.D., *Bertrand Russell: Critical Assessments*, vol. 2, London: Routledge, 1999, 1-3.

Goble, L (ed.) *The Blackwell Guide to Philosophical Logic* (Blackwell, Oxford, 2001)

Gödel, K *On Formally Undecidable Propositions of Principia Mathematica and Related Systems* (translated by E Mendelson from the original German Paper of 1931 in Davis (1964). Also available as its own book, translated by B Meltzer (Dover Publications, New York, 1992)

Goodman, N *Fact, Fiction and Forecast*, 4th edition, chapter 3 (Cambridge, Harvard University Press, 1983)

Gupta, A 'Truth' in Goble (2001), Chapter 5

Smullyan, R 'Gödel's Incompleteness Theorems', in Goble (2001), Chapter 4