|  |
| --- |
| When Do I Get My Money? |
| A Probabilistic Theory of Knowledge |
|  |
| **Jonathan Blamey** |

**01.11.10**

Ph.D. Thesis

Department of Philosophy

King’s College London

|  |
| --- |
|  |

### Abstract

The value of knowledge can vary in that knowledge of important facts is more valuable than knowledge of trivialities. This variation in the value of knowledge is mirrored by a variation in evidential standards. Matters of greater importance require greater evidential support. But all knowledge, however trivial, needs to be evidentially certain. So on one hand we have a variable evidential standard that depends on the value of the knowledge, and on the other, we have the invariant standard of evidential certainty. This paradox in the concept of knowledge runs deep in the history of philosophy.

We approach this paradox by proposing a bet settlement theory of knowledge. Degrees of belief can be measured by the expected value of a bet divided by stake size, with the highest degree of belief being probability 1, or certainty. Evidence sufficient to settle the bet makes the expectation equal to the stake size and therefore has evidential probability 1. This gives us the invariant evidential certainty standard for knowledge. The value of knowledge relative to a bet is given by the stake size. We propose that evidential probability can vary with stake size, so that evidential certainty at low stakes does not entail evidential certainty at high stakes. This solves the paradox by allowing that certainty is necessary for knowledge at any stakes, but that the evidential standards for knowledge vary according to what is at stake.

We give a Stake Size Variation Principle that calculates evidential probability from the value of evidence and the stakes. Stake size variant degrees of belief are probabilistically coherent and explain a greater range of preferences than orthodox expected utility theory, namely the Ellsberg and Allais preferences. The resulting theory of knowledge gives an empirically adequate, rationally grounded, unified account of evidence, value and probability.

# TABLE OF CONTENTS

[Abstract 2](#_Toc298826514)

[TABLE OF CONTENTS 4](#_Toc298826515)

[Chapter 1. 7](#_Toc298826516)

[THE BET SETTLEMENT THEORY 7](#_Toc298826517)

[1. A word on bet settlement 8](#_Toc298826518)

[2. Putting knowledge first 9](#_Toc298826519)

[3. Evidential certainty, externalism and truth 10](#_Toc298826520)

[4. Knowledge more basic than belief 13](#_Toc298826521)

[5. Belief is gradable 14](#_Toc298826522)

[6. Knowledge more basic than evidence 15](#_Toc298826523)

[7. Probability is a measure of both belief and evidential support 15](#_Toc298826524)

[8. The simple analysis 16](#_Toc298826525)

[9. Full belief 17](#_Toc298826526)

[10. Conclusive evidence 19](#_Toc298826527)

[11. Measuring degree of belief 27](#_Toc298826528)

[12. Bets model actions. 28](#_Toc298826529)

[13. Degree of belief in action 29](#_Toc298826530)

[14. Subjective degree of belief. 29](#_Toc298826531)

[15. Objective degree of belief 30](#_Toc298826532)

[16. Objective value of bets 31](#_Toc298826533)

[17. Stake Size Invariance 37](#_Toc298826534)

[18. Evidential Probability 44](#_Toc298826535)

[19. Stake Size Variation Principle (SSVP) 45](#_Toc298826536)

[19.1 The Stakes: B – C 46](#_Toc298826537)

[19.2 The Value of Evidence 48](#_Toc298826538)

[19.3. Value of conclusive evidence. 50](#_Toc298826539)

[19.4 Value of Evidence / Value of Knowledge 55](#_Toc298826540)

[19.5 Evidential Expectation 58](#_Toc298826541)

[19.6 Odds ratio 60](#_Toc298826542)

[19.7 Mutual Exclusivity, zero stakes and zero evidence. 61](#_Toc298826543)

[19.8 Kolmogorov’s axioms. 63](#_Toc298826544)

[19.9 Conditional Probability 70](#_Toc298826545)

[20. Subjective level of certainty 78](#_Toc298826546)

[21. Evidential level of certainty 80](#_Toc298826547)

[22. Scepticism 80](#_Toc298826548)

[23. Gettier counterexample to the tri partite analysis of knowledge. 83](#_Toc298826549)

[24. Conclusion 86](#_Toc298826550)

[Chapter 2. 89](#_Toc298826551)

[THINKING, KNOWING, AND BET SETTLEMENT 89](#_Toc298826552)

[1. Introduction 89](#_Toc298826553)

[1.1 Thinking and knowing. 91](#_Toc298826554)

[1.2 Socrates 93](#_Toc298826555)

[1.3 Descartes. 95](#_Toc298826556)

[1.4 Thinking the truth and knowing the truth. 97](#_Toc298826557)

[1.5 Mill on fallibility 98](#_Toc298826558)

[2. Measuring the difference in terms of probability 99](#_Toc298826559)

[2.1 Ramsey’s measure 101](#_Toc298826560)

[2.2 Certainty is when bets are settled 102](#_Toc298826561)

[3 Dutch Book Arguments 108](#_Toc298826562)

[3.1 Adding and subtracting unconditional bets and the law of disjunction 109](#_Toc298826563)

[3.2 Multiplying conditional and unconditional bets and the law of conjunction 113](#_Toc298826564)

[3.3 The Kelly Criterion 115](#_Toc298826565)

[3.3 Dutch books against knowledge when not certain 124](#_Toc298826566)

[3.4 Bet settlement requires justification and evidence. 125](#_Toc298826567)

[3.5 Bet settlement is intersubjective 126](#_Toc298826568)

[3.6 Settling bets when the proposition is false 129](#_Toc298826569)

[3.7 Settling bets with doubts, and knowing without settling bets. 130](#_Toc298826570)

[3.8 Jeffrey and the theorist’s terms. 133](#_Toc298826571)

[4. Logical closure and the certainty condition 134](#_Toc298826572)

[4.1 Challenges to logical closure 135](#_Toc298826573)

[4.2 Dutch Book argument for logical closure. 136](#_Toc298826574)

[5. Conclusion 137](#_Toc298826575)

[Chapter 3. 138](#_Toc298826576)

[STAKE SIZE VARIANCE AND THE PROBLEM OF RISK 138](#_Toc298826577)

[Introduction 138](#_Toc298826578)

[1 The problem of stake size variation for the certainty condition 143](#_Toc298826579)

[1.1 Rational degree of belief 1 necessary for knowledge 143](#_Toc298826580)

[1.2 But evidential standards for knowledge vary with stake size. 144](#_Toc298826581)

[1.3 Resolution of conflict: degree of belief 1 varies according to stake. 145](#_Toc298826582)

[1.4 General principle: evidential probability varies according to stake size. 146](#_Toc298826583)

[2. The Stake Size Variation Principle 149](#_Toc298826584)

[2.1 Descriptive and normative belief. 150](#_Toc298826585)

[2.2 Evidential externalism 152](#_Toc298826586)

[2.3 Propositional functions 153](#_Toc298826587)

[2.4 Reference class 155](#_Toc298826588)

[2.5 Certainty external to the evidence 156](#_Toc298826589)

[2.6 Stake size, wealth and utility 161](#_Toc298826590)

[2.7 Wealth is directional 166](#_Toc298826591)

[2.8 The Value of Knowledge 168](#_Toc298826592)

[2.9 The Stake Size Variation Principle 177](#_Toc298826593)

[2.10 Bid - Ask Spread. 180](#_Toc298826594)

[2.11 Probabilistic theory of knowledge 181](#_Toc298826595)

[3 Objection: Stake Size Invariance is a constraint on probabilistic coherence. 182](#_Toc298826596)

[3.1 The age old measure. 182](#_Toc298826597)

[3.2 Ramsey’s formulation 184](#_Toc298826598)

[3.3 Value odds not money odds 185](#_Toc298826599)

[3.4 Stake Size Invariance 186](#_Toc298826600)

[3.5 Dutch book arguments 190](#_Toc298826601)

[3.6 The conjunction rule, conditionalization and stake size escalation. 197](#_Toc298826602)

[3.7 Gambler’s Ruin 203](#_Toc298826603)

[3.8 Stake Size Variation Principle gives a better estimation of future frequency. 207](#_Toc298826604)

[3.9 Kelly Gambling 210](#_Toc298826605)

[4. Objection: Stake size variation adequately accounted for by non linear utility attached to risky outcomes. 212](#_Toc298826606)

[4.1 Bernoulli’s formula for calculating expectation. 213](#_Toc298826607)

[4.2 The Weber-Fechner Law 217](#_Toc298826608)

[4.3 Friedman and Savage and the wiggly line utility curve. 218](#_Toc298826609)

[Chapter 4 220](#_Toc298826610)

[SOLUTIONS TO PARDOXES FOR DECISION THEORY 220](#_Toc298826611)

[1. Introduction 220](#_Toc298826612)

[2. The Ellsberg Paradox 223](#_Toc298826613)

[3. The Allais Paradox 228](#_Toc298826614)

[4. The reflection effect 232](#_Toc298826615)

[5 The Lottery Paradox and the Preface Paradox 241](#_Toc298826616)

[5. 1 THE PREFACE PARADOX 250](#_Toc298826617)

[6 Decision weights 260](#_Toc298826618)

[7 Stake size variant degrees of belief and what is commonly meant by “probability” 261](#_Toc298826619)

[8 How much evidence justifies certainty? 264](#_Toc298826620)

[8.1 Level of certainty 268](#_Toc298826621)

[8.2 The greater the certainty, the less impact epistemic shock has on degree of belief. 270](#_Toc298826622)

[9. Conclusion 271](#_Toc298826623)

[References and Bibliography 277](#_Toc298826624)

# Chapter 1.

# THE BET SETTLEMENT THEORY

This thesis will be the proposal and defence of a probabilistic theory of knowledge. In its most simple form the probabilistic theory of knowledge is as follows:

S knows that *p* iff:

1. S is certain that *p*

2. S ought to be certain that *p*.

Certainty is equivalent to having no doubts and is measured as a probability 1 that *p* and a probability 0 that *~p*. Probability is given a betting interpretation. Under the betting interpretation of probability, probability 1 coincides with bet settlement. The two conditions can thus be rephrased.

1. S is disposed to settle bets that *p*.

2. Given the evidence and stake size, S ought to settle bets that *p*.

In probabilistic terms this becomes:

1. S has a degree of belief 1 that *p*.

2. The evidential probability that *p* given S’s evidence and interests is equal to 1.

Knowledge is factive: if S knows that *p* then *p* is true. This follows from the second condition. One ought never to be certain of a false proposition. If it is true that one ought to settle bets on *p*, then it is true that *p*. Evidential probability is the degree of belief one *ought to have* given one’s evidence and interests. If one *ought* to have a degree of belief 1 that *p*, then *p* is true. The probabilistic theory of knowledge takes the two conditions to be not only necessary, but also, given that truth follows from the second condition, jointly sufficient for knowledge. The two conditions can be called the belief condition and the evidence condition. The main focus of the thesis will be in defending the evidence condition, which involves a foundational examination of the relationship between evidential probability and stake size.

### 1. A word on bet settlement

This theory of knowledge can be called a *bet settlement theory of knowledge*. When there are no more doubts, and the evidence is conclusive, then the matter is *settled*. In a literal betting situation, the players begin by *placing* their bets on a clearly defined proposition. The players will be at this point ignorant of the truth value of the proposition. Then, at some point in the future, the players will discover, one way or the other, the truth value of the proposition. They are no longer ignorant, but now *know* the truth value. At this point, not only is the proposition *settled*, but also the bets on the proposition are settled. At this point the losing player releases the entire stake to the winner and the expected value of the bet becomes equal to the entire stake. So, given a literal betting interpretation of subjective probability, when bets are settled in favour of a proposition, the players’ subjective probability of the winning proposition is equal to 1.

Within the framework of a betting interpretation of probability, there is no higher degree of belief than 1, so bet settlement should be equated with certainty. Any legitimate doubts would be a legitimate reason to not settle the bet. So bet settlement is coincident with the highest degree of belief and the absence of legitimate doubt. When bets are settled, the players are both certain, since they accept the highest possible odds. They ought to be certain, since there are no legitimate doubts. Of course, it can often be the case that people literally settle bets (in that the stakes can be released to one player) while legitimate doubt remains. But in these cases, the bet is not legitimately settled, for if the doubts were legitimate and were raised, then the transaction itself is demonstrated to be illegitimate, and the bet unsettled.

It can also be objected that some bets are literally settled when the players are in some sense not subjectively certain. One way of dealing with this is to say that in many such cases, the proposition the bet was settled on is not the same as the proposition upon which doubts remain. This will be dealt with more thoroughly in chapter 2.

Another way of dealing with this objection is that the intuition that the players remain uncertain even when the bet is settled is down to the interest relativity of certainty. The players are certain *enough* to settle the bet at the stake size of the bet, but remain uncertain at higher stakes. It is the low stakes certainty that disposes them to settle the bet, and it is the higher stakes uncertainty that causes them to report private doubts. This stake size variation in certainty is the most substantive feature of the thesis, and will be argued for throughout.

### 2. Putting knowledge first

Following Timothy Williamson (2000) in *Knowledge and its Limits*, we will be “putting knowledge first”. Knowledge is conceptually prior to both belief and evidence, not constituted by them. The probabilistic theory of knowledge therefore does not pretend to be an analysis of knowledge in terms of its constituent parts. It is rather that the statement provides an axiom upon which a probabilistic theory of knowledge, evidence and belief can be built. Beliefs and evidence aim at knowledge, so it is against this standard that we should measure belief and evidence.

### 3. Evidential certainty, externalism and truth

Analyses of knowledge have often included truth as a necessary condition. The probabilistic theory of knowledge eliminates this condition. This is because, on this theory, to be evidentially certain is to be true. Therefore evidentially certain beliefs are a subset of true beliefs. If knowledge is evidentially certain belief, it follows that knowledge is true belief, making a truth condition redundant.

But is it really the case that to be evidentially certain is to be true? It can be the case that we gather a great deal of evidence for and against a proposition but reach no conclusion. In this case the evidence is inconclusive. The proposition is neither evidently true nor evidently false and therefore neither *p* nor ~*p* are evidentially certain. It can also be the case that we do reach a conclusion. This reaching of a conclusion is a proof. In this case the proposition has been proven by the evidence and as such is evidently true or evidently false. Evidence that culminates in proof is “conclusive evidence”. Evidence short of proof we can call “probable”. Evidence short of proof can be more or less probable. Probability comes in degrees. Using the most widely accepted conventions of probability, conclusive evidence for a proposition gives the proposition a probability 1 and conclusive evidence against gives it a probability 0. All *merely* probable evidence is therefore measured as a probability between 0 and 1. If the evidence for a proposition is merely probable, then the proposition is not evidentially certain. The hypothesis I am putting forward is then that:

If the evidence for p is conclusive, then p is true.

On this reading, conclusive evidence for a proposition is an external relation. This means that the conclusiveness of evidence is not fully determined by purely internal features of the evidence. The conclusiveness of evidence is partly dependent on the truth or otherwise of the conclusion.

An internalist notion of conclusive evidence, on the other hand, would assume that the conclusiveness of evidence must be internal to the evidence. A further common feature of evidential internalism is that any property of the evidence must be perceptible by a subject who possesses this evidence. If the conclusiveness of evidence was an internal property, then anyone who was in the same internal state as someone who had conclusive evidence that *p*, would also have conclusive evidence that *p*. This internalist notion collapses the appearance/reality distinction with regard to evidence. But a subject who is in possession of evidence E can mistakenly think that E is conclusive evidence that *p*. Therefore whether or not E is conclusive that *p* is at least partly external. (See T. Williamson, 2000, chapters 2, 4, 8 and 9).

For example: Let’s suppose that E consists in a clear view of an armchair. Suppose a subject S in possession of E comes to believe with certainty that there is a kitten sat on the armchair. So S takes E to be conclusive that there is a kitten. But, unbeknownst to S, he is suffering from a perceptual distortion brought on by lack of sleep. A second person, R, also in possession of E, does not find E conclusive that there is a kitten on the armchair, since R has no perceptual distortions, and there is observably no kitten on the armchair. In this thought experiment an externalist account of evidence would have it that S has mistaken E for conclusive evidence that there is a kitten. The internalist alternative to this interpretation would deny that S has the same evidence as R. This would be to say that the evidence that S has is not the view of the armchair, but the way the armchair appears to S. This internalist account of evidence would have it that S does in fact have conclusive evidence that there is a kitten on the armchair, because it appears to S as if there is a kitten on the armchair. So, on the internalist picture it is possible to have conclusive evidence for false propositions. This leaves internalism open to Gettier type cases and to scepticism. There is no logical problem with an internalist account of evidence, as long as it is made clear that the term “evidence” is being used in a technical way that is introduced by a definition. But, in a non technical context, calling the way things appear to people “evidence” is simply to misuse the term, or at least to use it in a non standard way. Evidence is something that is necessarily public, whereas appearances of this kind are private. A common rationalist slogan is “show me the evidence”. All S can show R is E, the view of the armchair. S cannot show R the way in which the armchair appears to S. There is nothing that S can show R that would oblige R to settle a bet with S that there is a kitten on the armchair. So in terms of bet settlement, S does not have conclusive evidence that there is a kitten. S is, however, *subjectively certain*. S is disposed to settle bets on there being a kitten. It is just that these bets would not actually *be* settled, because S is not evidentially certain, and his evidence is not conclusive.

The claim is that if the evidence for *p* is conclusive then *p* is true. It is uncontroversial that if we take the evidence for *p* to be conclusive, then we take *p* to be true and we take ourselves to know that *p*. The objection is that it is possible that we take the evidence for *p* to be conclusive when *p* is false. This is only an objection if one shares the internalist intuition that any evidence that we take to be conclusive is conclusive evidence. But all that follows from the possibility of taking evidence for *p* to be conclusive when *p* is false is that:

It is possible to take the evidence for *p* to be conclusive when the evidence for *p* is not in fact conclusive.

Here is an argument from the possibility of taking evidence E to be conclusive that *p* when *p* is falseto the possibility of taking evidence E to be conclusive that *p* when E is not conclusive evidence that *p*:

1. Suppose that a subject S takes evidence E to be conclusive that *p*.

2. Suppose that *~p*.

Given that *~p*, it is clear that we ought not to conclude that *p* on the basis of E. If we ought not to conclude that *p* on the basis of E, then E is not conclusive evidence that *p*. Therefore the fact that S finds E conclusive that *p* does not entail that E is conclusive that *p*.

### 4. Knowledge more basic than belief

The concept of knowledge is arguably more basic, more historically and cross culturally stable than the concept of belief. It is fairly uncontentious to suggest that the knowledge had by the Ancient Greeks was of the same kind as the knowledge we have today. When Socrates considered the question “what is knowledge?” he was considering the same question which forms the basis of this thesis.

Belief, on the other hand, is a very murky and changeable concept. In non philosophical contexts a person’s “beliefs” are their religious or moral principles. Since the twentieth century, the idea that beliefs are necessarily accessible to the believer has fallen by the wayside as the concept of subconscious belief became main stream in psychology. Advances in neuroscience and psychiatry have arguably transformed our very concept of belief. Many philosophers believe that beliefs are necessarily identical with brain states. Other philosophers believe that beliefs require language and that non linguistic creatures do not have beliefs at all. The question “what is a belief?” far from being the same question across millennia, is not even the same question across university corridors.

It is also arguable that belief is the more complex concept requiring a prior understanding of knowledge. Findings in developmental psychology suggest that the acquisition of the concept of a false belief comes relatively late, and marks an important milestone in the development of a theory of mind [See Chevalier and Blaye (2006) and Wellman, Cross and Watson (2001)][[1]](#footnote-1). Traditionally knowledge has been thought of as a kind of belief with the extra property of being true. It might be truer to the evolution of ideas to think of knowledge being the basic concept, and belief being a species of knowledge, but with the extra property of being possibly false. In the first person, the assertions that “I believe that *p*” and that “I know *p*” share the redundancy of truth. “I believe that *p*” is equivalent to “I believe that *p* is true” and the same goes for “I know that *p*” and “I know that *p* is true”. The difference is that asserting belief is to recognise at least the possibility of error, which requires a much richer conceptual frame work. This is not to say that the possibility of error does not exist in knowledge. It is perfectly possible for a subject to assert that he knows that *p* on good evidential grounds and yet *p* be false. “I know that *p*” and “*p*”, are similar in that they are both false if ~*p*. Whereas “I believe that *p*” stands alone in that its truth is independent of the truth of *p*.

### 5. Belief is gradable

As well as having an inherent possibility of error, belief is also gradable. Belief can come in degrees. With probability theory we can quantify degree of belief numerically. A reason to assert that “I believe that p” but to dissent from “I know that p” is that as well as believing that p, one also has some doubt that p. This merely probable partial belief state is clearly more representationally complex since it involves representing both *p* and *~p*; whereas a full belief, or the basic knowledge of plants and animals, only requires the representation of *p*.

### 6. Knowledge more basic than evidence

Evidence is an equally murky concept. What has been generally taken as evidence, conclusive or otherwise, has changed significantly with the emergence and development of scientific method. The most significant shift in the concept of evidence was the emergence of statistics and probability theory. With probability theory and statistical method, the degree to which a hypothesis is supported by the evidence can be quantified numerically. In this thesis we argue that the degree of evidential support necessary for knowledge is also equal to 1.

### 7. Probability is a measure of both belief and evidential support

Knowledge is species of belief that is justified by evidence. Knowledge is binary: either you know that p or you do not. Evidence and belief on the other hand are gradable: you can believe to more or less an extent and your belief can be more or less supported by the evidence. How justified does a justified true belief have to be to count as knowledge? How strong our degree of belief? To answer these questions we need a measure theory of evidence and belief. Probability theory is just such a theory. We can assign a probability measure to both the degree to which someone believes a proposition, and the degree to which that proposition is supported by the evidence. The degree to which it is rational to believe a proposition is the degree to which it is confirmed by the evidence. So degrees of belief and evidential support should be measured on the same scale. Probability theory provides this scale.

Knowledge can be analysed as a full belief that is justified by conclusive evidence. Using subjective probability theory a full belief is when the value of a bet is equal to the value of the prize. In this case the subject will consider it fair to settle outstanding bets. This is a subjective probability 1. Conclusive evidence gives an objective warrant for evaluating a bet on *p* as equal to the prize. Conclusive evidence is a reason to believe that all bets made in those conditions win. This evidence will then warrant settling a bet. This is an evidential probability 1.

### 8. The simple analysis

The simple analysis of knowledge can be given in two conditions:

S knows that *p* iff:

1. S does not doubt that *p* (belief relation).

2. S is right not to doubt that *p* (evidence relation).

That these two conditions are necessary is uncontroversial. To assert that one doubts that *p* is equivalent to denying that one knows that *p*. So not doubting is a necessary condition for knowledge. Likewise, to assert that the evidence for *p* is not conclusive and therefore one ought to doubt that *p* is equivalent to denying that one knows that *p*. So both not doubting and being right not to doubt are necessary for knowledge. Furthermore evidential certainty entails truth, (one is never right not to doubt a proposition which is false), and Gettier required that false beliefs can be justified. So the two conditions are jointly sufficient and not vulnerable to Gettier type counterexamples.

### 9. Full belief

Probability used as a measure of degree of belief sets total doubt at 0 and full belief at 1. Partial beliefs are where there is some doubt and some belief. A subject who does not doubt that *p* has a full belief that *p* which is a subjective degree of belief 1. We can call this *subjective certainty*.

Subjective certainty is relative to the level of interest a subject has in a proposition. The level of interest a subject has in a proposition is how much difference its truth makes to him. This is a non evidential feature of the context and can vary while the evidence remains fixed. For example suppose that a subject who has very little interest in whether hummus contains peanuts is subjectively certain that *hummus contains no peanuts* in that he will take it for granted and consider the matter settled without much thought. To borrow a phrase from Millikan, (1993) he will act as if *hummus contains no peanuts* without further ado. But if he subsequently discovers that if he eats some hummus he will live if it is true that *hummus contains no peanuts* but die if it is not true that *hummus contains no peanuts*, he will become highly interested in the proposition *hummus contains no peanuts*. At this new higher level of interest the subject is not necessarily subjectively certain that *hummus contains no peanuts*. Now that it is a matter of life and death, doubts may appear that simply did not exist before. In this new context, it is possible that the subject will no longer consider the matter settled but will seek more evidence to silence this newly elevated level of doubt before going ahead and eating the hummus. So by merely changing the level of interest from low to high, it is possible that a subject can go from being certain to being doubtful without any change in evidence. Or to put it another way, given the same evidence, it is possible that the same subject can be certain relative to low stakes bets, but less than certain relative to high stakes bets.

Subjective certainty then, is context sensitive and the feature of the context to which it is sensitive is what is at stake. The probabilistic analysis of knowledge can account for the context sensitivity or interest relativity of knowledge attributions that has been a feature of much recent epistemology. (Stanley 2006; DeRose 2004; Hawthorne 2004; Fantl and McGrath 2002; Shaffer 2006; Schiffer 2007). It accounts for the interest relativity of knowledge attributions by positing the interest relativity of certainty. Subjective certainty is interest relative just because it is the actual case that people can be more doubtful when the stakes are higher. Put like this the interest relativity of subjective certainty may seem uncontentious. But when subjective certainty is defined within decision theory as a subjective probability 1, the interest relativity of subjective certainty has far reaching consequences, since it spills over into the interest relativity of subjective probability. The degree of belief an agent has in *p* is not a unique value, but must be indexed to a stake size. This results in a two dimensional specification of probability based on what I call the “Stake Size Variation Principle”. These stake size variant degrees of belief turn out to have a lot in common with Tversky and Kahneman’s decision weights in their *Prospect Theory*, (1979) and consequently stake size variant degrees of belief explain the decisions people tend to make under risk in cases where standard decision theory fails. But, as opposed to Prospect Theory’s decision weights, the stake size variant degrees of belief conform to the basic axioms of probability calculus relative to any stake size and are thus proof against Dutch Books.

An absence of doubt at all levels of interest can be called “absolute certainty”. Absolute certainty is not necessary for knowledge, only subjective certainty at the relevant level of interest in the context of knowledge attribution. The first condition then is that:

1. S knows that *p* only if S’s subjective probability that *p* = 1, in the context of knowledge attribution.

### 10. Conclusive evidence

One is right not to doubt that *p* when one’s evidence is conclusive. Conclusive evidence then must be sufficient to rule out all doubt at the level of interest that is appropriate to the context. In betting terms, conclusive evidence is the evidence sufficient to settle the bet. If all reasonable doubt is ruled out by the available evidence, then the correct degree of belief at that level of interest is 1. To distinguish the correct degree of belief given the evidence from the subjective degree of belief we shall call the correct degree of belief to have relative to the evidence and interest the *evidential probability*.[[2]](#footnote-2)

Inspired by C. E. Peirce (1949) we think of evidential probability as attaching not to a proposition, but to an inference class. The agent’s evidence forms the premises of an inference and the probable belief forms the conclusion. For example, suppose our evidence includes the fact that Saladin is German, and that 44% of Germans have blue eyes. We can then infer from Saladin being German to Saladin having blue eyes. Without knowing anything more about Saladin, we know that this inference has a 44% success rate, when applied to all Germans. We can say that, at least relative to the whole reference class of Germans, the evidential probability of Saladin having blue eyes is 44%. On the other hand, were we to apply this inference to the class of people called Saladin, rather than to Germans, then our success rate is unlikely to be 44%. Whether we go on to use the inference on Germans, or on people called Saladin is not determined by our evidence, but by our interests. Evidential probability can thus be interest relative.

Evidential certainty is also interest relative because it is rational to demand more evidence for certainty when the stakes are higher. A person with a peanut allergy really ought to have higher evidential standards when considering whether an item of food that he is about to eat contains peanuts. This again spills over into evidential probability. This amounts to the more contentious claim that as well as being descriptive of human psychology, the stake size variation principle is a *normative principle*. This is the most interesting claim of this thesis and the bulk of the thesis will be dedicated to arguing for the rationality of the Stake Size Variation Principle.

The obvious objection to the stake size relativity of certainty stems from our previous claim that evidential certainty entails truth. The worry is that if evidential certainty at high stakes is that level of evidence necessary to guarantee truth at high stakes, then the lesser level of evidence required for evidential certainty at lower stakes will fail to guarantee truth. But the factivity claim is merely that evidential certainty is sufficient for truth. So at all stake sizes, *p* is only evidentially certain if *p* is true. It is just that, as the stakes go up, fewer true propositions are evidentially certain.

Another related worry is that it might be thought that only deductive proof entails truth. But this is to ignore the external element to evidential certainty. If evidential probability was a purely internal property of the evidence, then only metaphysically necessary inferences would lead to certainty, since only deductive proof can guarantee truth in all possible worlds. But since evidential probability is partly externally constituted, then evidence guarantees truth just so long as all the inferences in the *actual* world are true. So externalism about evidential certainty allows a more realistic, empirically achievable evidential certainty. This means we can be evidentially certain, for example, that Obama is President of America, while accepting that there is a possible world where, internally speaking, we have identical evidence, but where Obama is not the President of America. Since subjective certainty is not externally constituted in this way, it is possible to be quite reasonably subjectively certain of facts that aren’t evidentially certain because they are surprisingly false. Cases of error of this kind can be expressed by first person. It can be true to say: “I was certain that I would get the job, but I didn’t get the job.” But, due to the externality of conclusive evidence it is never true to say: “It was evidentially certain that I would get the job, but I did not get the job.” The Stake Size Variation Principle would thus predict that mismatches between subjective and evidential certainty of this kind would be more frequent at lower stake sizes.

However, the Stake Size Variation Principle is a normative principle and so also applies to evidential probability. Because only true propositions can be evidentially certain, it cannot be the case that there is more frequent error in evidential certainty at low stakes. But what can be the case is that higher stakes propositions will be more frequently true when they are *not* evidentially certain. We can see this with the *hummus contains no peanuts* example. This is a case where we have evidential certainty at low stakes, but we fail to have evidential certainty at high stakes. Given that the target proposition is true, it is the high stakes lack of certainty that is in error. Failing to be certain of a true proposition is a much weaker form of error than being certain of a false proposition. The relative merits of these two types of error grounds the Stake Size Variation Principle. The higher the stakes, the wiser it is to err on the side of caution.

There is another source of worry for the claim that evidential certainty entails truth. It is a well known issue in mathematical probability theory that probability 1 does not entail truth. In cases where there is a countably infinite partition it is standard to assume that each element in the partition has a probability 0 and therefore the negation of each element has a probability 1. If it is accepted that these assignments are uniformly correct, then we have a variety of counterexamples to the claim that a correct assignment of probability 1 to *p* entails that it is true that *p*.

For example, suppose we divide the area of a dart board into points and throw a dart at the dart board. Since each point has an area equal to zero, there is zero probability that the dart will hit any particular point. Therefore there appears to be a correct probability assignment of 1 to the proposition that the dart will not hit any specific point. However, the dart will hit one point, assuming the set up makes sense, and so a probability of 1 is correctly assigned to a false proposition.

Another type of counterexample is generated by taking limiting cases. For example, if you were to toss a coin N times, then the probability of getting at least 1 heads is equal to 1 - 1/2N. As N tends to infinity, then the probability of getting at least 1 heads tends to 1. Furthermore, the probability of the frequency of heads being equal to 1/2 increases with N. So as N tends towards infinity, the probability that the sequence will have equal numbers of heads and tails tends to 1.

These counterexamples fail to demonstrate that evidential probability 1 does not entail truth for two reasons. Firstly the Stake Size Variation Principle conforms to the axiom of normalisation. This means that the disjunction of every possibility in an infinite partition must sum to 1 given the conditions of identity of stakes and completeness of evidence. This is only possible if the probability of each member of the partition is infinitesimally greater than zero, and therefore the negation of each element is infinitesimally less than 1. For the identity of stakes to hold, every possibility must result in either winning or losing the bet at these stakes. Propositions that you are betting for are at positive stakes, propositions that you are betting against are at negative stakes. According to the Stake Size Variation Principle, (which we will go into fully in section 19 of this chapter), to bet on the dart hitting a specified point on the dart board (di), such that winning the bet (di) results in £2 and losing the bet (~di) results in £1 has an evidential probability (£1 + kdi) / (£2 + ∞£1 + K) where kdi is the evidence in favour of the dart hitting that particular point, and K is the total evidence for and against the dart hitting that particular point. This makes the negative stakes probability that (~di) = (£2 + ∞£1 + K – £1 – kdi) / (£2 + ∞£1 + K) which we take to be less than 1 by (£1 + kdi) / (£2 + ∞£1 + K), an infinitesimal amount. This is known as *almost certain*. In terms of expectation, the difference in the expectation of this bet and £1 is so small that it would not be possible to devise a currency in which to give the correct change from a £1 coin. The upshot of this is that, although it is permissible to say that the evidential probability of ~di is 1 for all practical purposes; technically it is not exactly 1, but infinitesimally less than 1, so that it is not *conclusive evidence* and therefore does not entail truth. The argument here is that conclusive evidence is necessary for knowledge and conclusive evidence is given evidential probability 1. Cases where there is a good argument for assigning a probability 1 to a false proposition are not cases of conclusive evidence and therefore not cases of knowledge.

The second reason the counterexamples fail is due to the externality of evidence. In order for there to be a stake size at all it must be specified which points on the dart board one is betting for and which points one is betting against. For reasons we will go into in section 19, one cannot bet for *all* the points, and one cannot bet against *all* the points, otherwise there would be no stake size and the outcome would be the same whatever happened. If we bet against a single point di, then we are assigning probability 1 to a proposition ~di that *might be* false. This specific bet can turn out two ways; it can win, if the dart hits the dart board at ~di; or it can lose, if the dart hits di. In the former case, one has assigned evidential probability 1 to a true proposition, in which case the counterexample fails. In the latter case, then the externality of evidence ensures that the evidential probability of hitting ~di cannot be conclusive, since di is true.

The externality of evidence flags up an interesting aspect of the example of the frequency in the limit. Let us suppose we toss a coin N times and count all the heads (H) and thus obtain the frequency H/N. The law of large numbers dictates that the larger N, the greater the probability that H/N will fall within ½ ± d for arbitrarily small d. Consequently as N → ∞ then the probability that H/N = ½ → 1. We can call this the “limit argument” and count it as evidence that H/N = ½ when N = ∞. One way of interpreting this is assigning a probability 1 to the inference class: “if we toss the coin an infinite amount of times, then we will get the same number of heads as tails.” But it is possible, for example, that we could toss only heads, or only one tail and the rest heads, or only 2 tails and the rest heads, or only N tails for any finite N and the rest heads, providing an infinite number of counterexamples. This looks like a probability 1 is being assigned to any number of falsehoods.

We will now provide an elimination argument that the limit argument doesn’t provide a counterexample to the proposal that evidential probability 1 entails truth.

Either the evidence of the limit argument is evidentially conclusive for the inference class or it is not.

If it is conclusive evidence then it entails that if we toss a coin infinitely then H/N = ½ is true. In this case, then there is no need to actually toss the coins, because we already have bet settling evidence in favour of H/N = ½. In this case it seems we have not generated a counterexample.

If on the other hand the limit argument is not conclusive, then it ought to be possible to have bet settling evidence that for a particular infinite sequence of coin tosses H/N ≠ 1/2. In this case evidential probability 1 would have been proven to have been assigned to a false proposition. For any finite N, then this is clear enough. The proposition H/N = ½ can be established on any actual sequence of coin tosses by counting the number that fall heads. Sequences in which H/N ≠ ½ will provide conclusive evidence against the proposition that H/N = ½ in that particular sequence. So if we tossed a coin an infinite amount of times and counted the heads, and H/∞ ≠ ½ then we have a counterexample to the claim that H/N = ½ when N = ∞, and therefore conclusive evidence that probability 1 does not entail truth.

I do not believe that such counterevidence is possible to obtain. This is because it is not possible to count the number of heads in an infinite sequence of coin tosses in such a way that there is an intelligible value for H/∞. Firstly, the infinite sequence would take an infinite amount of time and energy to complete, which is another way of saying it would never be completed. Secondly, there would be no point at which it was decidable that there were no more heads. Thirdly, it is not obvious how H/∞ can ever equal ½ since if H is finite than H/∞ is infinitesimal, and if H is infinite, then H/∞ is undefined, possibly 1. Fourthly, if someone were to attempt to derive a figure for H/∞ by extrapolating from the tendency as N increases, and if this is considered to be conclusive, then the original limit argument ought to be considered conclusive.

These objections pull in different directions, but the general point, and it is an epistemologically significant one, is that bets on the frequency in sequences of infinite length can never be settled. If they can never be settled, then they can never be lost and so there is no counterexample to the claim that evidential certainty entails truth. If, however, they could be settled, then this would provide external evidence that makes the evidence inconclusive. If in such cases we still deem it useful to assign probability 1 for reasons of mathematical orthodoxy, then we can merely say that this is a case of probability 1 that does not entail truth and is not sufficient for knowledge. Take for example, Jon Williamson’s (1999), guess the number game, where one person thinks of a number, and the other person guesses this number. Suppose we allow that the number could be any number out of the infinite series. In this case the probability that any number n is not the number could be plausibly assigned a probability 1. But it is realistic to imagine that this particular bet could easily be settled at low stakes by for example, asking the person who thought of the number whether or not it was n. In this case the probability 1 assignment is less plausible without detailed argument. If n actually was the number, then it is fairly plausible to suggest that the assignment of probability 1 to n not being the number was not correct.

The claim that the Stake Size Variation Principle is a normative principle amounts to the claim that just as subjective probability is stake size variant because the degree of belief can vary with the level of interest, evidential probability is also stake size variant, because having evidential standards that are sensitive to the level of interest is the wiser thing to do. As the level of interest increases, the possible scope of doubt increases, and the level of evidence required to dispel all doubt should increase proportionately. I will argue for the claim that the evidential probability of a proposition is sensitive to the level of interest. In light of this, the second condition is that:

2. S knows that p only if the evidential probability of p = 1, relative to S’s evidence and interest.

If both subjective and evidential certainties are necessary conditions of knowledge, then the interest relativity of knowledge attributions follows from the interest relativity of subjective and evidential certainty.

### 11. Measuring degree of belief

In order to give meaning to degrees of belief understood as probabilities we will turn to the seminal work of F. P .Ramsey (Mellor ed.1990; Galavotti ed.1991). Ramsey developed the old established idea that the degree of belief a subject has in a proposition can be measured in terms of the odds at which the subject would be prepared to bet on that proposition. A bet that results in B if *p* and C if *~p* can be thought of as a chance to win the stakes B – C if *p*. The odds at which the subject is prepared to bet can thus be indexed to a specific stake size by asking what value A – C would constitute the maximum price the agent would pay for the chance to win the stakes B – C. In other words, if a subject is indifferent between A for certain and B if *p* and C if *~p*, then the agent considers A – C a fair price for a chance to win the stakes B – C if *p*. So we define degree of belief as (A – C) /( B – C) which is the ratio:

*Price / Stakes*.

The value A is the *expected utility* of this particular bet, the value C is the subject’s holdings if he loses the bet, and the value B is the subject’s holdings if he wins the bet. The value (A – C) / (B – C) is the degree of belief the agent has with regard to this particular bet. Measuring degree of belief in this way means we can have an operational measure of degree of belief that allows the possibility of stake variant degrees of belief. A belief is stake size variant if the Price / Stakes ratio varies with the value of the stakes. For example, if a person was prepared to pay £10 for a bet that paid £20 if *p* and nothing otherwise, but only prepared to pay £80 for a bet that paid £200 if *p* and nothing otherwise, then her degree of belief measured in money odds would be 0.5 at stakes £20 and 0.4 at stakes £200.

### 1**2. Bets model actions**.

A bet can be used to model any action whose success depends on the truth of a proposition. Using the utility measure that Ramsey developed we can measure the interest a subject has in a proposition by the difference in the utility of success and the utility of failure. This utility difference we call the *stake size*. The degree of belief in the success of any action that can be modelled in this way is the difference in the expected utility and the utility of failure as a proportion of the difference in the utility of success and the utility of failure. In other words subjective probability is measured as the expected utility of an action as a proportion of the stake size.

Ramsey explicitly made the assumption that a subject’s degrees of belief remain constant across stake sizes. In other words however much you vary the value of the stakes, the price/stakes ratio will and should always remain the same. Ramsey does not provide positive arguments for this assumption, but rather the assumption simplifies the theory and is essential for his preference based utility scale. In the probabilistic theory of knowledge we explicitly reject the assumption of stake size invariance on the grounds that it is descriptively false and has no normative weight. Instead, we allow that the degree of belief can change with the stake size, and furthermore, in many instances it does and should vary according to stake size. This rejection of stake size invariance is the most far reaching and substantive claim of this dissertation.

### 13. Degree of belief in action

Any two possible courses of the world between which the subject is indifferent have the same subjective utility. We will represent the common utility of any set of possible courses of the world between which the subject is indifferent with capital letters A, B and C. If a subject is indifferent between options:

1.A

Or

2. B if p is true is and C if p is false

Then the subject’s degree of belief that p relative to this choice of action is equal to:

(Value of Action, A – Value of Failure, C) ÷

(Value of Success, B – Value of Failure, C)

So the degree of belief in *p* = (A – C) ÷ (B – C).

### 14. Subjective degree of belief.

On this theory there is a supervenience relation between choices over action and beliefs and desires. There can not be a different choice of action given the same beliefs and desires. (It is important to notice that, conversely, the same choice of action *can* result from different beliefs and desires. Intentional actions are fully determined by beliefs and desires, but beliefs and desires are underdetermined by actions). Beliefs thus conceived are measured by their causal efficacy in the process of fixing the expected utilities of actions. This is an entirely subjective measure of belief strength, since the measure is independent of any evidence the agent has. The subjective probability is purely a measure of an agent’s preference ordering over choices of action. It is the how the subject values an action as a proportion of how the subject values success.

(How S Values the Action) ÷ (How S Values the Success)

### 15. Objective degree of belief

The objective degree of belief in a proposition is the *actual* value of an action to the agent as a proportion of the actual value of success, rather than the value the agent puts on the action.

The objective degree of belief in an event is given by:

(Objective value of action) ÷ (Objective value of success)

The evidential probability need not equal the objective degree of belief, since the objective value of an action may not be determined by the available evidence. The objective degree of belief should not be confused with *objective probability*. Objective probability is a concept that is not directly relevant to the probabilistic theory of knowledge, and so we will try to say as little about it as possible. Objective probabilities are often conceived of as mind independent causal properties that generate patterns of relative frequency according to the axioms of probability. Even given that there are such causal properties, and that they do conform to the axioms of probability, it is not the case that the objective degrees of belief always match the objective probability.

### 16. Objective value of bets

One might ask at this point what we mean by the objective value of a bet? The objective value of a bet simply means the final value, as in the takings at the end of the day. For a single bet, this will be one of two values: win or lose. For a sequence of bets it will be the sum of the values of each of the individual bets.

Sequences of bets can be grouped into what I shall call a *reference class*. A reference class is simply a number of bets which all have the same evidential probability with respect to a particular store of knowledge. For example suppose a fisherman was selling oysters for a uniform price of say £1. If an oyster has a pearl in it the pearl can be sold for £50, with a net profit of £49. B = W + £49 where W is the starting wealth. If it has no pearl, then it cannot be sold, meaning that the buyer makes a net loss of £1. (C = W –£1). The action of buying an oyster from the fisherman can be modelled as a bet with stake size B – C = +£50 on there being a pearl. Because we have no evidence by which we can distinguish any oyster from any other in terms of its likelihood of bearing a pearl, then the evidential probability of any oyster having a pearl must be equal. This is just to say that, as far as our evidence goes, we have no more reason to think that any oyster is more likely to contain a pearl than any other, so they are all worth the same before we shuck them. If the value of an oyster is constituted by its chance of containing a pearl, then all oysters will have the same value. All such purchases of oysters from this fisherman thus form a reference class.

The reference class has two important properties: the relative frequency, and the size. The relative frequency is simply the proportion of oysters that have pearls. The size is the number of oysters that are bought. The objective value of any particular purchase depends entirely on whether or not there is a pearl in that particular oyster. But if a merchant were to buy many oysters on a regular basis, then the objective value of the reference class would equal the relative frequency multiplied by the reference class size multiplied by the stake minus the cost. So for example, if over a life time the merchant bought 10 000 oysters and 1 in 20 yielded pearls, then the objective value would be the relative frequency (1/20) times the stake size (£50), times the reference class size (10 000) minus the cost £10 000 = £15 000. This is the objective value to the merchant of his oyster speculation. If we divide through by the reference class size, we then get the per oyster value, which is £1.50. In this example, we can say that the objective degree of belief that a pearl contains an oyster for this merchant was (A – C)/(B – C) = (£1.50 + £1)/(£49 + £1) = 1/20. So in cases where the stake size is uniform, the objective degree of belief will always be equal to the relative frequency.

However, the relative frequency will depend on the specific reference class and should not be conflated with objective chance. We are not putting forward a frequentist interpretation of probability here. A famous problem for frequentist interpretations of probability is the “reference class problem”. Here is Hajek explaining the problem for actual frequentists:

“Actual frequentists such as Venn (1876) in at least some passages and, apparently, various scientists even today, identify the probability of an attribute or event A in a reference class B with the relative frequency of actual occurrences of A within B. Note well: *in a reference class B*. By changing the reference class we can typically change the relative frequency of A, and thus the probability of A. In Venn's example, the probability that John Smith, a consumptive Englishman aged fifty, will live to sixty-one, is the frequency of people like him who live to sixty-one, relative to the frequency of all such people. But who are the people "like him"? It seems there are indefinitely many ways of classifying him, and many of these ways will yield conflicting verdicts as to the relative frequency.” Hajek (2007)

According to actual frequentists, the probability of an event is equal to the relative frequency in the reference class to which the event belongs. But the event may belong to several different reference classes and thus have several different probabilities. This reference class problem is only a problem if we insist that any event must have a unique unconditional probability. Hajek goes on to identify this quest for a metaphysical account of unconditional probabilities as the source of the reference class problem. He suggests accepting that probabilities are essentially relative to reference classes, in other words probabilities are essentially *conditional*. This is a powerful and exciting idea since it involves favouring two place probability theories over one place probability theories. A two place theory will have conditional probability as primitive; whereas a one place theory, like Kolmogorov’s, will define conditional probability in terms of one place unconditional probabilities by the ratio definition.

(RATIO) P(A|B) = P(A∩B)/P(B), (provided P(B) > 0

Since “Kolmogorov's theory clearly reigns”, and since Kolmogorov “began by axiomatizing unconditional, or absolute, probability,” (Hajek, 2007 p.1) then Hajek’s conclusion is just too exciting for this thesis. We make the lesser claim that objective degrees of belief simply do not correspond to Kolmogorov’s “probabilities”, and so we prefer Ramsey’s system of degree of belief, where conditional degrees of belief are primitive, and simply deny that conditional degrees of belief ought to conform to the ratio definition.

There is an important parallel between the insistence on unique unconditional probabilities and Ramsey’s assumption of stake size invariance. Stake size and reference class size both form the denominator in probability calculations, while expected utility and relative frequency form the numerator. Changing the stake size and changing the reference class will have the same effect on the probability. In order to have a single unique probability for any event, the reference class and stake size must also be unique.

Uniqueness of probability can be recovered if we can specify on any particular occasion what the relevant reference class is. Since we are not trying to give an analysis of objective chance, we can make a move that was not allowable to frequentists. This is that the relevant reference class is determined by the interests of the subject of the belief. If this is allowable, then the metaphysical problem goes away, since every probability statement has a clear cut objective truth value once the reference class is determined. We are talking about action guiding degrees of belief so the relevant reference class is determined by the actions that the subject will perform to which the degree of belief is relevant.

In the example of the merchant and the oysters, the relevant reference class is the ten thousand oysters that the merchant actually ends up buying. The relative frequency is the proportion of pearl oysters *in these* ten thousand oysters. Had the merchant only bought 10 oysters, then the relative frequency could have been 1, 0 or any one place decimal in between, depending on how many, out of those ten oysters, had pearls. The objective degree of belief is not something necessarily knowable in advance to the merchant or anyone else. But it is related to knowledge and evidence in this way: it is what the evidential probability is *aiming at*.

Judgements of the objective value of bets in reference classes will guide behaviour. If the Merchant had a degree of belief 1/20 that oysters have pearls, he would have calculated the expected utility accurately. If, on the other hand he had over estimated the probability, and then borrowed on his expectation, he would have ended up in debt. Had he under estimated, he may not have bought the oysters at all, and the reference class wouldn’t have existed.

As we all know from the law of large numbers, a smaller size of reference class is going to make predicting the actual relative frequency in that reference class less reliable. Let’s suppose that the merchant has really good evidence that in general the relative frequency of pearls in oysters tends to 1/20. If the relevant reference class is small, then the relative frequency in the reference class is likely to deviate considerably from 1/20. Suppose for example the merchant only buys 10 oysters. He should expect the relative frequency of pearls in this reference class to be 0/10 with a negative deviation of 1/20; or, if he is lucky, 1/10 with a positive deviation of 1/20. If he is really, really lucky, all ten could contain pearls.

Whereas with a large reference class, the deviation is likely to dwindle down to next to nothing. This forms the basis of a normative argument for the Stake Size Variation Principle. If the stakes are high, then a large negative deviation may cause the agent to stop that type of activity altogether, making for a much smaller reference class.

For example, were the prices of Oysters £1 000 each and pearls £50 000, then the merchant may have only been able to afford ten oysters. If the deviation in this reference class was negative, and he found no pearls, he would have gone bankrupt and his pearl speculation would have come to an end. The reference class size would then have been only 10.

Whereas a positive deviation earlier on will enable the agent to continue that activity, making for a much larger reference class. Had the merchant found a pearl in the first ten, he would have gained £50 000 and would have been able to buy many more oysters. As the reference class gets larger, the deviations will tend to get smaller. If the merchant bought 1 000 oysters with his accumulative profits, then the likelihood of a deviation of 1/10 falls to close to zero. This has the over all effect that positive deviations are likely to be smaller than negative deviations. So even if the “objective probability” of an event of a particular repeatable type is known, and it can thus be predicted that the relative frequency will tend to converge on this figure, it could still be wise to have a degree of belief at high stakes lower than the objective probability.

Furthermore, there will be many reference classes where the concept of objective probability does not apply. Oysters and pearls may be a good example of this. There simply may be no stable relative frequency of pearls in oysters, and no mechanism that corresponds to the objective probability. This latter observation may address a kind of metaphysical disappointment with this denial of uniqueness. The objective degree of belief in the first oyster containing a pearl depends on how many oysters the merchant goes on to buy. If objective degree of belief is equated with objective probability, then this may seem disappointing to the point of being unacceptable. We want to know what is *the* probability of *this* oyster having a pearl; and that can’t be anything to do with the future frequency of oysters with pearls. I personally do not believe that there is any such thing as *the* objective probability of a specific oyster having a pearl, although the question is an interesting one. However, I do believe that there is an objective value of a set of bets, and this is simply the sum of all the wins and losses. If our evidence cannot distinguish between the bets in terms of their chance of winning then we can take an average, and call that the objective value of each bet. When considering evidence in deciding whether to bet, it is this objective value we should be aiming at, not the objective probability. This argument will be given again in more detail in chapter three.

To avoid confusion it is worth distinguishing what we mean by the *objective value of a bet* from the market price of a bet. The market price of a bet is its exchange value in the market. This has good cause to be called its objective value too, but this is simply not what we mean. There is no argument here, just a matter of terminology. The market price of a bet can easily be distinguished from its objective value in our sense. The market value of a bet will depend, among other things, on the evidence available to the market.

We should also distinguish what we mean by *objective value* from what is often called the *actuarial value* of a bet. The actuarial value of a bet is the sum of the objective probability of the outcomes multiplied by the value of the outcomes. On a one off bet, the actuarial value will only equal the objective value if the objective probability is 1 or 0. However, on a long run of bets, the actuarial value will usually be a good estimate of the objective value. Due to the law of large numbers, the longer the run, the better this estimate will be. Our thesis concerns not objective probability, but evidential probability. If the objective value predictably diverges from the actuarial value, then it is the objective value that matters to the subject. What governs what a subject’s degree of belief *ought* to be is not the actuarial value, but the objective value. So the evidential probability should aim at the objective degree of belief as defined in terms of objective value, not the objective probability. In other words, it is not necessarily the case that the evidential probability equals the objective probability if known.

### 17. Stake Size Invariance

There is an assumption common to interpretations of probability that I shall call “Stake Size Invariance”. Ramsey (1931) included this assumption when he stated his famous definition of subjective probability; I’ve italicized the stake size invariance assumption.

“If the option of A for certain is indifferent with that of B if p is true and C if *p* is false, we can define the subject's degree of belief in *p* as the ratio of the difference between A and C to that between B and C; *which we must suppose the same for all A, B's and C's that satisfy the conditions*.” [My italics]

“A”, “B” and “C” refer to utility values attached to possible states of the world. “A” refers to the expected utility of the gamble on *p*. The difference between B and C is what we are calling the stake size. The larger the difference between B and C, the bigger the gamble. The italicized supposition is that an agent will have the same degree of belief that *p* whatever the stake size. Ramsey says that we *must* suppose this. But he does not argue for this supposition. Nor is this supposition a part of the theory. In fact, it is a testable hypothesis that is, at least on a naïve understanding of utility, observably false. It is observably the case that people often prefer lower stakes bets to higher stakes bets at the same odds, meaning that the ratio of the difference between A and C to that between B and C will get smaller as the difference between B and C gets larger.

Ramsey was aware of the observable fact that people prefer lower stakes bets in relation to money odds, as were most probability theorists going back to Fermat, Pascal and the Chevalier De Mere. It is not entirely obvious to me why the supposition of stake size invariance could not simply be dropped, as I am suggesting in this thesis, since in practice, no one seems to conform to it. There is scarcely any explicit argument for stake size invariance, but I will offer one here just to clarify the debate. Degrees of belief understood in terms of betting odds are supposed to match our ordinary language concept of “probability”. Our ordinary language concept of probability has as a paradigm the probability of coin tosses and dice throws. So the degree of belief someone has in an event, for example, a six rolled on a fair die, should match the probability that that person ascribes to the event. Almost everybody will ascribe a probability of 1/6 to the roll of a six on a fair die. The explicit ascription of probability will not change as the stakes go higher. Even though people may prefer a small bet on six at 5 : 1 to a much larger one, no one would say that the *probability* of rolling a six is less than 1/6 just because the stakes have got higher. So we must assume that their degree of belief is stake size invariant.

My objection to this defence of stake size invariance is twofold. Firstly, it is permissible that our concepts can evolve. We are talking about a measure of degree of belief that is action guiding and a corresponding measure of evidence that is justifying. These measures conform to the laws of probability and fit quite well with our normal pronouncements of probability. But it is quite conceivable that peoples’ degree of belief and evidential justification can vary with the stake size while they pronounce that the probability remains invariant. We are not obliged to accept that such pronouncements are necessarily or analytically true. After all, people say all sorts of things that are scientifically naïve and are considered to be theoretically false. Our best theory of physics holds that time intervals vary relative to inertial frames of reference. To say that a time interval is invariant over shifts in velocity has just as much claim to ordinary language support as the claim that probability is invariant over shifts in stake size. The fact that people persist in saying that the sun rises in the morning should not commit us to the supposition that the earth does not spin. So nor should it constrain our theory of knowledge that people ascribe probabilities to events that do not match their degrees of belief in those events. Further to this, given our framework, we can make sense of these probabilistic attributions by the concept of “fair odds”, which is the odds at which the subject would be indifferent between which side of the bet he took. This will cover pronouncements attaching to dice and coins without assuming that anyone’s actual degree of belief will match the fair odds.

Secondly, “probability” can have an entirely objective meaning which is best captured by the word “propensity”. It is clear, in the case of dice at least, that there is some mind independent property of the dice related to its symmetry that governs the likely long run relative frequencies of dice rolls conducted in similar conditions. The relationship between these causal propensities and degree of belief and evidential probability are not clear. We will argue in this thesis that even under the assumption of a propensity for a long run frequency of 1/6, one’s evidential probability could be different depending on stake size, due to the different effect on the length of the run of negative and positive deviations. In other words, even granting the reality and coherence of objective propensities, it need not be the case that evidential probabilities must always equal the objective propensities if known.

However, for nearly a century now considerable theoretical effort has been made to defend stake size invariance against the charge that it is observably false. The first step is to change the units of measurement from money odds, to utility odds. The diminishing marginal utility of money then explains how people can maintain a stake size invariant degree of belief, while at the same time preferring lower stakes bets. This is one of the reasons why Ramsey used utility odds instead of money odds in his definition. Ramsey’s method for eliciting a subjective utility scale relies on the assumption of stake size invariance.[[3]](#footnote-3)

However this theoretical solution failed to save the supposition of stake size invariance against the charge of observable falsity. This is because even when money odds are replaced by utility odds, people still seem to prefer lower stakes bets, at least in some cases. At this point a measurement problem arises, since without the assumption of stake size invariance, it is hard to have a non circular measure theory of utility. But laying aside this problem, one can still preserve stake size invariance in the face of actual preferences by assuming an extra utility attached to risk itself. The theory is that people prefer lower stakes bets at equal utility odds because they attach a negative value to gambling, or in other words, are “risk averse”.

But even this solution doesn’t work, because, as Kahneman and Tversky discovered, it is possible to create decisions over risk where people *still* exhibit stake size variant degrees of belief which cannot be explained by risk averse utility functions. In their *Prospect Theory* they proposed “decision weights”, which behave like degrees of belief in terms of calculating expected utility, but that do not conform to the axioms of probability.

“In prospect theory, the value of each outcome is multiplied by a decision weight. Decision weights are inferred from choices between prospects much as subjective probabilities are inferred from preferences in the Ramsey-Savage approach. However, decision weights are not probabilities: they do not obey the probability axioms and they should not be interpreted as measures of degree or belief.” (Kahneman and Tversky 1979 p.280)

Ironically, the reason that Kahneman and Tversky’s decision weights don’t obey the axioms of probability is because they already assume S shaped utility functions in order to preserve stake size invariance. But once they drop stake size invariant degrees of belief and replace them with decision weights, the theoretical need for S shaped utility functions goes away. S shaped utility functions were introduced by Friedman and Savage, (1948) who admitted that they were implausible. Given a more natural utility function coupled with stake size variant degrees of belief, people’s preferences can be modelled much more simply assuming their degrees of belief are probabilistically coherent at a stake size, but can vary across stake sizes.

Perhaps the real reason why so much theoretical effort has been made to preserve stake size invariance is revealed in Kahneman and Tversky’s statement that decision weights don’t conform to the axioms of probability. Perhaps this centuries long rigorous defence of the observably false supposition of stake size invariance is down to the fear that stake size invariance is a necessary condition of probabilistic coherence. Kahneman and Tversky’s quick argument is that decision weights of *p* and *~p* tend to sum to less than 1. But this is easily explained by the fact that Kahneman and Tversky have already corrected their value function to account for what they call the “reflection effect”. This is roughly speaking the phenomena that people seem to be risk seeking over losses, but risk averse over gains. If *p* is at positive stakes, then *~p* will be a negative stakes, so a fully coherent decision weight would make the reflection effect disappear. I will go into more detail about this in chapter 3.

But why assume stake size invariance is a necessary condition for probabilistic coherence anyway? I think the reason for this is because a straightforward way of eliciting someone’s degree of belief is by asking them to give odds for a bet in ignorance of what side of the bet they will take. This method was made explicit by De Finetti and is such an easy concept to grasp that it has made it way into many text books on decision theory. But this method of eliciting degree of belief has the assumption of stake size invariance built into it as De Finetti was well aware. If one does not assume stake size invariance, then this method will give, not the degree of belief in *p*, but the midpoint between the degree of belief in *p* at positive stakes and the degree of belief in *p* at negative stakes, or the *fair odds*. If degrees of belief really do vary according to stake size, we would expect a bid-ask spread between these two degrees of belief, which gets wider as the stakes get higher. This would mean that, although the subject would be indifferent between either side of the bet, he would not be indifferent between being forced to take the bet or not, and would prefer to not bet at all than to take either side of the bet at the fair odds. This, according to Ramsey’s definition, would make his degree of belief at positive stakes lower than the fair odds and his degree of belief at negative stakes higher than the fair odds.

I have a quick argument to show that a subject can have different degrees of belief at different stake sizes that are not vulnerable to Dutch books.

Roughly speaking the argument is this: a strategy for avoiding Dutch books is to buy low and sell high. Buying a bet on *p* for positive stakes is structurally identical to selling a bet on ~*p* for negative stakes, so buying high and selling low is coherent. As the stakes on *p* get higher, the stakes on ~*p* get lower. So long as (your degree of belief in *p* at stake size U) = (1 – your degree of belief in ~*p* at stake size – U), your degrees of belief will be coherent and you will not be Dutch bookable. For example someone who is entirely ignorant of whether or not *p* could want to risk as little as possible on *p* and therefore refuse to buy any bets at any price on *p* or not *p*. This would make his degree of belief in *p* = 0 at all positive stakes and = 1 at all negative stakes, and the same for ~ *p*. It is clear that it is not possible to get someone into a Dutch Book who refuses to buy or sell any bets. Someone who has no evidence either way about *p* has no normative requirement to have an invariant degree of belief at different stake sizes just so long as the degree of belief at any one stake size is lower than it is at any lower stake size. Such an agent is admittedly risk averse to a perhaps unadvisable extreme, but he is none-the-less coherent.

### 18. Evidential Probability

Ramsey devised his subjective probability measure as a way of giving meaning to probability statements. This is often over looked, but it is clear from the text that Ramsey had something like the following principle in mind:

**Ramsey’s Principle**: The evidential probability of *p* relative to K is the correct degree of belief to have in *p* given that K was all that one knew.

Ramsey assumed stake size invariance, which we take to be false. So the principle must be amended to account for variation according to stake size.

**Stake Sensitive Principle**: The evidential probability of *p* relative to K and stake size U is the correct degree of belief to have in p given that K was all that one knew and one’s interest in *p* is equal to U.

The interest that S has in p is equal to the difference in value to S of the world if *p* and the world if not *p*.

So the same interests and evidence entails the same rational degree of belief.

### 19. Stake Size Variation Principle (SSVP)

In order to give formal substance to evidential probability as we have defined it, we will give a precise formula that defines evidential probability in terms of:

i) The stakes, B if *p* and C if *~p*.

ii) The value of evidence in favour of the proposition which we will call “k*p”*.

iii) The value of evidence against the proposition which we will call “k*~p”*.

iv) The total absolute value of evidence relevant to whether or not *p* which we will call “KP”, KP is the sum of the absolute values k*p* + k*~p*.

For the evidential probability *p* relative to stakes B – C and knowledge KP we will write:

PK(*p*)B - C

The Stake Size Variation Principle “SSVP” comes in two parts, one pertaining to cases where the evidence is one sided, and one pertaining to cases where the evidence is mixed.

If the evidence is one sided, so that k*p* = KP then:

PK(*p*)B – C = k*p* / (B – C) if k*p* < B – C

Or

PK(*p*)B – C = 1 if k*p* > B – C. In this case the value k*p /* (B – C) gives the level of certainty.

Consequently, *p* is certain whenever the value of evidence is greater than or equal to the stake size.

If the evidence is mixed so that k*p* ≠ 0 and k~*p* ≠ 0 then:

PK(*p*)B – C = (C + k*p*)/ (B + C + KP)

Consequently *p* is never certain while there remains any evidence against *p*.

In this section we will show how these two formulas can be derived from consideration of Ramsey’s definition of degree of belief and the way we have defined the value of evidence. We will then show that evidential probabilities derived from the SSVP conform to the axioms of probability.

### 19.1 The Stakes: B – C

Ramsey’s formula has it that if a subject is indifferent between A for certain and an opportunity which results in B if *p* and C if ~*p*, then his degree of belief in *p*

= (A – C) / (B – C).

Let us suppose that the subject’s evidence justifies indifference between A for certain, and B if *p*, C if ~*p*. In this case:

PK(*p*)B – C = (A – C) / (B – C)

The SSVP states that PK(*p*)B – C varies with the stakes B – C whilst maintaining probabilistic coherence. The values B – C are imposed on the subject from outside, whereas the value A is chosen by the subject in accordance with the degree of his belief in *p*. The stronger the belief in *p* the closer A will be to B. Whenever the stakes on *p* are B – C, the stakes on *~p* are C – B. Therefore, by the same definition, whenever

PK(*p*)B – C = (A – C) / (B – C)

Then

PK(~*p*)C­­ – B = (A – B) / (C – B) = 1 – PK(*p*)B – C

It is important to note that B and C refer to end states, not losses and gains. It is also important to note that combinations of bets may be at different stakes from their constituent parts. For this reason it is not possible to have a degree of belief in *p* and in *~p* both at positive stakes relative to the same action. Positive stakes means that B – C is positive. For example, suppose you had £100 to start with. You bet £10 on *p* for a chance to win £40 if *p* (B = £130, C = £90). You also bet £20 on ~*p* for a chance to win £25 if ~*p* (B = £105, C = £80). In this scenario, you have not in effect made two bets at positive stakes, but have only made one bet such that you pay £5 for a chance to win £10 if *p*. The positive stakes on this bet are B = £110 – C = £95 for *p* and the negative stakes are B = £95 – C = £110 for *~p*. If you are indifferent whether you take this bet or not, then your degree of belief in *p* is 1/3 and in *~p* is 2/3. So long as *p* is a proposition that satisfies the law of excluded middle, then no combination of bets on *p* and on *~p* can have more than two outcomes. Therefore, relative to any simultaneous set of bets: PK(*p*)B – C = 1 - PK(~*p*)C­­ – B. In other words, as the evidential probability *p* varies with the stakes, the evidential probability of *~p* varies in harmony and the probability of the partition remains 1.

It might be objected that it is often the case that a subject could weigh up whether to bet for or against a proposition *p* and will thereby be considering both *p* and *~p* at positive stakes in the same act of deliberation. In this situation, what is being compared is the expectation of each bet, and one chooses the bet with the highest expectation. In this comparison, both *p* and *~p* will have two values for evidential probability, one at positive stakes and one at negative stakes. The SSVP allows different evidential probabilities at different stake sizes. But taking the two positive stakes values, PK(*p*)B – C  + PK(~*p*)B – C < 1. At first glance this may seem incoherent, especially since we are talking about the same agent in the same act of deliberation. But this seeming incoherence vanishes as soon as a decision is made and the stakes collapse into one value for B – C. Therefore such an agent will never make a decision that leaves her vulnerable to a Dutch Book.

This is in fact a strength of the SSVP since it allows a bid-ask spread, which leaves it rationally open to decline both sides of a bet. Whereas any stake size invariant theory of evidential probability that was committed to the evidential expectation principle would make it a rational requirement to always prefer one side of a bet over the status quo as long as the odds were complimentary. For example, suppose you were offered a bet on either *p* or ~*p* at even odds where the stakes were to double your entire holdings (corrected for utility) if you win and renounce all your belongings if you lose. Suppose you had absolutely no information regarding whether or not *p*. If your degree of belief in *p* was less than ½ then your degree of belief in *~p* should be greater than ½ and the bet on *~p* would thereby have an expectation greater than declining both bets, whereas if your degree of belief was greater than ½ then you ought to prefer the bet on *p* to declining both bets. According to any stake size invariant theory of evidential probability, then to be coherent, if the degree of belief in one is less than half, then the degree of belief in the other must be greater than ½, and if you prefer to decline one bet, you must prefer to take the other bet. But surely it is not *irrational* to decline both sides of this bet!

A final word on the stakes, I often use the term “stake size” and refer to the value B – C. However, the SSVP is also sensitive to the ratio of (B – C)/C. In other words, the absolute values of B and C have an impact on evidential probability as well as the difference in values B - C. This means that PK(*p*)B­­ – C doesn’t necessarily equal PK(*p*)D – E even if B – C equals D – E. Hence stakes should always be expressed as the difference between two values, rather than expressing them as a simple magnitude.

### 19.2 The Value of Evidence

The value of evidence is a concept unique to this thesis and is a necessary element in the SSVP. To give meaning to the claim that evidential probability can vary with the stakes even when the evidence remains the same, we need a non probabilistic measure of evidence to make it clear what is meant by “the same evidence”. In order to make it commensurable with the stakes, we are using the *value* of evidence as a measure of evidence. The value of evidence can be measured by the change in evidential expectation the evidence brings. However, the effect of evidence on expectation is relative in most cases to the evidence one already has. If you already know that *p* then new evidence in favour of *p* is of little value. So to get an absolute value of evidence, we will take the value of evidence k*p* to be the increase in expectation that the evidence brings *from a state of sceptical ignorance.*

An ignorant sceptic is one who has no evidence for or against a proposition and therefore withholds belief. He neither believes *p* nor *~p*. This is not the same as having a degree of belief ½ in *p* and in *~p*. It is more like having a degree of belief 0 in anything *p* related. The ignorant sceptic prefers option 1, A for certain to option 2, B if *p*; C if *~p* for all values of A, B and C. In other words, the ignorant sceptic will always prefer not to bet. For example, suppose the ignorant sceptic got an email that told him that *p*:if he replied to the email with the words YES he would receive £20 000 000 but if *~p* then responding to the email will have no effect on his wealth, the stakes on *p* are therefore positive, B = W + £20 000 000, - C = W. (W is the current state of wealth). He gets another email that tells him that *q*: if he replies to the email with the words YES he will not be robbed of all his wealth, whereas if *q* is false, then if he replies to the email with the words YES he will be robbed of all his wealth. The stakes on *q* are B = 0 and C = - W. The ignorant sceptic is indifferent whether he responds to either email because he believes neither that *p* he will get £20 000 000, which makes him value this bet at its lowest level C if not *p*, nor does he believe that ~*q* that he will get robbed of all his wealth, which makes him value the bet at its highest level: B if *q*. The attitude of the ignorant sceptic is to believe nothing unless he has some kind of evidence. Consequently, the ignorant sceptic values bets on *p* at as low as C and bets on *~p* as high as B.

Now suppose the ignorant sceptic receives some evidence that changes the evidential expectation to a unique value A. If the evidence is one sided and in favour of *p*, then this will raise the expectation from C up to A. This makes the value of evidence k*p* = A – C. Therefore:

PK(*p*)B – C = k*p* / (B – C) = (A – C) / (B – C).

If the evidence is in favour of *~p*, then the evidence reduces the value from B down to A. This makes the value of evidence k*~p* = A - B. Presuming B – C to be positive, this means that the value of evidence k*~p* is negative, whereas the value of evidence k*p* is positive.

PK(~*p*)C – B = k*~p* / (C – B) = (A – B) / (C – B) = 1 - PK(*p*)B – C

However, the value of evidence we take to be absolute and not relative to any particular bet. Hence if there is evidence for and against, we take the total value of evidence to be the sum of the absolute values. The value of evidence is therefore the *change in expectation the evidence brings.* This adds value to your decision making whether the result is to decide for an action, or decide against an action.

It should be noted that we are only talking of cases where there is one sided evidence here. Because the evidence is one sided, when k*p* = A – C, k*~p* is still equal to 0. It is only if the all evidence is against *p* that the value of k*~p ­*= A – B.

### 19.3. Value of conclusive evidence.

Conclusive evidence is worth the entire stake size. Conclusive evidence often comes in the form of an observation that is sufficient to settle a bet. Imagine you have paid £5 for a bet 1 : 2 bet on Black Beauty winning a horse race. If she wins you will collect £15, £10 winnings plus your original £5. There is a sense in which the observation that Black Beauty crosses the line first is worth £15 in that it increases your fortune by £15. Independently of your wealth, the stake size on this bet will always be £15. So relative to this bet, the value of knowledge is £15, which is the value of knowing that Black Beauty wins at these stakes. You might think that the observation is only worth £10, because you already laid on £5 on your background knowledge. But the winning ticket is worth £15 whatever you paid for it, and the observation will settle this bet however much you disbelieved it before hand. Here we have the principle that is at the heart of this thesis:

The value of knowledge is equal to the stake size: KP = B – C.

When thinking in terms of betting on horses, this may seem too relative. The same observation seems to be worth different amounts depending on how much is at stake. However, there is a limit, we propose, to the stake size at which an observation will provide conclusive evidence. In horse racing, we would take the observation to settle almost any bet, but it is possible to imagine stakes so high that an observation wouldn’t settle the bet without some corroborating facts. Horse racing provides an artificial evaluation to an observation. Ordinarily the value of observing a horse passing a post would be of little interest. The unusually high value in this case is an artefact of the culture of horse racing and gambling. But we believe that knowledge often has genuine intrinsic value. Technical knowledge and scientific knowledge, knowledge of how things work and knowledge of the relationships between people, these things have a value that is hard to deny and this value is the stuff of human life. The value of evidence is therefore the *maximum* stake size at which the evidence is conclusive, and this is a genuine intrinsic feature of the information.

To get away from the betting framework where the value of knowledge is artificial, here is a more abstract example. Imagine a type of black ball that has an intrinsic use and value. This value is 1 Unit of value (U). We are to imagine that these black balls have a real value that is not dependent on any particular market. (Perhaps black balls operate as batteries and produce a certain amount of free clean energy.) There is a procedure E that may result in a black ball or not, no one knows. Let us say you reach into a sack and pull out what you find. If you pull out a black ball, then *b* | E is true, if you do not, then *~b* | E. If you do not try the sack then the issue of whether or not *b* | E does not arise. Now, the procedure E may vary in cost. If it goes above 1 Unit then there is no point in doing E. If it goes below 0, then you may as well do E whatever your belief in *b* | E. But whatever the cost, and whatever your starting wealth, the stake size on *b* | E will remain the same: B – C = U. You will always be one black ball richer if you do pull a black ball out, than if you fail to pull a black ball out. However, the cost of E and your background wealth will make a difference to the values of the stakes B, C and thus make a difference to the evidential probability.

We can see now that if you had observed 100 black balls being drawn from the sack, then you would be evidentially certain that if you do E, *b* will result. The value of this evidence would be k*b* = 100U, and according to the formula, this makes P (*b* | E)b = 1 with certainty level 100; whereas if you drew from the sack a hundred times and no black balls were ever observed, then you would be reasonably certain that if E then *~b*. In this case P (*~b* | E) = 1. However, add these two sets of evidence together, and we end up being neither certain of *b* nor of *~b*. But we are certainly no longer ignorant. The value of our knowledge pertaining to *b* is now 200U. We are no longer sceptically ignorant, but are now ready to accept bets for and against *b* | E at positive stakes depending on the odds. The question now is what bets are rational to accept given this evidence?

We can see that the ratio k*b*/KP will be equal to the observed frequency. Ignoring stakes, the value k*b*/KP serves as a good estimate of evidential probability. We notice how, as KP gets large relative to U, then the epistemic impact of each new observation of a black ball gets less. The 201st ball will swing k*b* / KP by less than 0.005, whereas the second ball may have swung k*b* / KP by as much as 0.5. Other measures of weight of evidence like I. J. Good’s which take the change in probability as a measure of evidential weight would have these two observations as having different weights, whereas the value of any two observations is the same according to the SSVP, regardless of what has already been observed. The value KP thus serves as a momentum that lessens the effect of new evidence of the same value. This preserves the absolute value of evidence whilst keeping its impact contextually relative.

Even in this abstract case, one may wonder whether a single observation of a black ball is enough to warrant certainty that the next ball will be black. Plausible rules of succession are notoriously hard. We say in defence of the SSVP that the reluctance to grant certainty after a single observation may be due to a reading of P(*b* | E) as the generalisation: “every E is followed by a *b*”. This latter generalisation is at stupendously high stakes, counted perhaps in infinite U, or at least U times all the black balls in the universe. Given infinite stakes, then it is a consequence of the SSVP that no number of observations will ever justify certainty. Whereas the simple statement that the next thing drawn from the sack will be a black ball is only of stakes 1U, and given you have just drawn a black ball, and have never drawn anything else, then it is reasonable to expect the same to happen again. The certainty meant here is merely practical certainty. It is merely when the best estimate of the practical value of a procedure is to expect success. The first pull at the sack has an expectation of plus zero, so the subject considered here has already received a windfall of U over his expectation. If his certainty should prove to have been a mistake, he at most loses the U that he gained through luck in first place.

There is an issue here with evidential externalism and our claim that PK(*b* | E)B – C = 1 entails that *b* is true or E is false. The issue is that it seems perfectly possible in a metaphysical sense that you could first draw a black ball, and then draw a white ball. In this case it appears that the SSVP gives evidential probability 1 to a false proposition. Of course, any reader of Hume will know that no amount of drawings of black balls will necessitate that the next ball is black. We can call the first occurrence of E & *~b* in such a sequence an “epistemic shocker”. A part of the way the SSVP works is to include epistemic shockers as an external part of the evidence, but not to include non epistemic shockers. In general the SSVP assumes that in cases of mixed evidence the next instance will be a loss, and weights the negative evidence accordingly. It shares this feature with Laplace’s rule of succession such that the inductive probability is (*p* + 1) / (*p* *+ ~p* + 2). However, the SSVP only takes this as the evidential probability when the unobserved instance actually is a case of *~p*. Transposed into Laplace’s rule then this would be the proviso that if the unobserved instance is a case of *p*, and there have been no cases of *~p* then the evidential probability is simply (*p* + 1)/(*p* + *~p* + 1) = (*p* + 1)/ (*p* + 1) = 1. Because of the externality of evidence, then evidential certainty only occurs when the object proposition is true.

For example, if you had observed 100 black balls and no other results, then the evidential probability would be 1 with certainty level 100 / 1 if the next ball was actually black, but an evidential probability (C + 100) / (B + C + 101) if the next ball was not black. This shows starkly how the evidential probability is not necessarily internally accessible. However we do not defend evidential internalism, nor do we see it as a requirement for any theory of evidential probability. It might help to make a distinction between “rational certainty” where certainty is justified by your evidence but does not entail truth; and “evidential certainty” where rational certainty is a necessary (but not sufficient) internal condition for evidential certainty, and the non falsity of the object proposition is a necessary external condition for evidential certainty. It might be worth mentioning here that “non – falsity” covers cases of conditional probability where the condition is not met. So P(*b* | E)B – C = 1 is satisfied in cases where E is false. In this case *b* given E is not false, but perhaps not true either. In other words, as long as you do not ever pull any balls from the sack after the first one, then the evidential probability can remain certain whatever non black balls are lurking in the unopened sack. Epistemic shock can only come through experiment. In this way the SSVP *encourages experimentation* at positive stakes certainty, since it raises the expectation of untried low stakes bets, while counselling caution at high stakes.

### 19.4 Value of Evidence / Value of Knowledge

We can measure the value of evidence in terms of the increased value in expectation that evidence will bring. The increase in expectation has to be from a position of ignorance of the evidence. The hypothetical value of knowing a proposition on receiving conclusive evidence depends on the plausibility of the proposition prior to the evidence. So this measure of value of evidence has to be against a backdrop of shared presuppositions K. We can assume that generally the value of knowing an item of common knowledge that forms the shared presuppositions of the language community is zero, (for example, you won’t make much money betting that the sun will rise tomorrow, nor do you need much evidence to justify this belief). Whereas the value of knowing something that is generally considered to be highly unlikely can be much higher (for example, if you knew that Goldman Sachs had defrauded the British government, this information could be very valuable indeed, but you would need a great deal more evidence to justify this belief). We can see then that the value of evidence e which is not conclusive that *p* is a function of PK(*p* | e) – PK(*p*), and consequently the value of conclusive evidence that *p* is a function of 1 – PK(*p*).

Because the SSVP allows for different degrees of belief at different stake sizes, there is a phenomena that we will call “Stake size escalation”. To put it simply, because of the possibility of escalator bets and the relationship between conditional probability and the probability of conjunctions then for Dutch books to apply to conjunctions, the stakes increase by a factor of 1/PK(hi | hi-1)U for each additional conjunct. For example, were you to know that a dice roll was to come up six, then you could potentially make a profit of five times your entire holdings W, since the background probability of throwing a six is 1/6. This would make the market odds 1 : 5, and you would place W on six, and win 5W plus your original W. But if you were to know that two dice rolls were to land six, you could potentially make a profit of 35 times your holdings by betting your entire holdings W on the first roll, and then your new increased holdings of 6W on the second roll. So though the value of evidence of each die roll is 6W, the value of knowledge of the conjunction of both is not 12W, but 36W.

We can think of this in terms of the number of alternatives ruled out by the knowledge of a proposition. Each conjunct *hi* rules out only one alternative, namely *~hi*. So the sum of conjuncts only rules out N alternatives, whereas the conjunction of a binary function rules out 2N alternatives. So while the value of the evidence only increases by a factor of N, the value of knowledge increases by the power of N.

Ramsey, and Alan Turing, both came up with solutions to the problem of measuring the value of knowledge. Turing’s solution was to take a measure of the change in odds of the proposition the evidence supported. This means that the value of evidence can only be measured relative to a proposition for which one already has assigned a prior probability. We then have F(h, e) meaning the factor in favour of h given by e. This is simply:

F(h, e) = Od (h, e) / Od (h)

Where Od (h, e) is the odds on h given e expressed as a fraction.

To scale the ratio a log to base 10 was taken. Good (1950) then transformed the odds ratios into probabilities and created the weight of evidence measure:

W(h, e) = log P(e, h)/P(e, ~h).

The problem with both these measures is that they fail to define conclusive evidence. This is because the odds on h given e when e is conclusive evidence for h is 1 : 0. As an Odds ratio this is 1/0 which is undefined. So F(h, e) is undefined whenever e is conclusive evidence for h. Likewise, if e is conclusive evidence for h, then  *p* (e, ~h) = 0, and therefore W(h, e) is undefined.

Gillies (1990) compared the above measure, which he called the “Turing-Good weight of evidence” with Popper’s “severity of test” measure. (Good, 1950; Popper, 1974). The severity of test is given by “Q” in the following formula:

Q = P(e, h & k) – P(e, k), (“k” here refers to the background stock of knowledge.)

In contrast to weight of evidence, it is possible under Popper’s measure for e to be conclusive evidence that h. Though Popper himself was against betting interpretations of probability, his measure Q can be expressed in betting terms as the advantage that adding h to one’s stock of knowledge gives in terms of expectation on a bet on e divided by the stake. In this Popper was preceded by Ramsey, who measured the value of knowledge in much the same way (Ramsey, 1991 p.285-287). All that is missing is the stake. Q is the difference in expectation that the new evidence creates as a proportion of the stake. To know h is to have an expectation equal to the entire stake. Given that the value of knowing h is U, in other words, the stakes on a bet on h is U, then the value of knowing e relative to h is QU. QU is the increase in expectation of a bet on e at stakes U that is given by h.

As we have said, we do not assume stake size invariance, an assumption that both Good (1950) and Popper (1974) implicitly made. Consequently we do not assume that Q is the same for all stake sizes. We instead propose that the value QU is the constant and that Q varies with stake size. In other words, the change of expectation granted by a new piece of evidence remains the same, while the increase in probability varies according to the ratio of the value of the evidence and of the value of knowing the proposition which the evidence confirms.

### 19.5 Evidential Expectation

The evidential expectation of a bet is the value of a bet considered in the light of the evidence. The thinking behind the SSVP and the probabilistic theory of knowledge is that knowledge and value are inextricably linked through the evidential expectation principle. The evidential expectation principle is that it is a rational requirement to be weakly indifferent between opportunities with the same evidential expectation. It is this principle that motivates the SSVP. We take it that considerations of value do provide reasons for action. So it is often rational to prefer a lower stakes bet to a higher stakes bet even if the odds are the same. If this is right, then the evidential expectation principle *only holds* if the evidential probability varies with the stakes. If evidential probability did not vary with stake size, then all bets at the same odds would have the same evidential expectation. Therefore, given that it is sometimes rational to prefer low stakes bets to high stakes bets at the same odds, then it would follow from stake size invariance that the evidential expectation principle does not always hold. In this case evidential probability would not provide a sufficient rational constraint on action and would not function as an input into decision making. In other words, unless evidential probability varies with the stakes, then there is no rational requirement to match degrees of belief to evidential probability, and evidential probability loses its prescriptive meaning.

As we have discussed, if the evidential probability is (A – C) / (B – C) then the evidential expectation is equal to “A”. We can thus demonstrate that the one sided SSVP holds.

PK(*p*)B – C = (A – C) / (B – C)

A = C + PK(*p*)B – C (B – C) = B + PK(~*p*) C – B (C – B)

When the evidence is one sided in favour of *p*:

k*p* = A – C = PK(*p*)(B – C)

Therefore: PK(*p*)B – C  = k*p* / (B – C)

And PK(~*p*)c – B  = (B – C – k*p*) / (B – C) = 1 – (k*p*) / (B – C) = 1 - PK(*p*)B – C

When the evidence is one sided in favour of ~*p*:

k*~p* = A – B = PK(~*p*) (C – B)

Therefore: PK(~*p*)B – C  = k~*p* / (C – B).

And PK(*p*)B – C  = (C – B – k*p*) / (C - B) = 1 – (k~*p*) / (C – B) = 1 - PK(~*p*)C - B

Put into words, the evidential expectation is obtained by adding a proportion of the stake size to the value of failure. In this case the proportion of the stake size added to failure is the evidential probability. So the greater the probability of success, the greater the proportion of the stake size is added to the value of failure. From this it is clear that the evidential probability must be greater than or equal to 0, or the evidential expectation would be lower than C, the value of failure, which is absurd. It is also clear that the evidential probability must be no greater than 1, or the expectation would be greater than B, the value of success, which would likewise be absurd.

For this reason we must add the proviso that

PK(*p*)B – C = k*p* / (B – C), unless k*p* /(B – C) > 1 in which case PK(*p*)B – C = 1 with certainty level k*p* / (B – C).

In cases of one sided evidence in favour of *p*, where the value of evidence is less than or equal to the stake size, then the expectation is equal to:

C + k*p*

But if the evidence is one sided in favour of *~p* then the expectation is equal to

B + k~*p*

### 19.6 Odds ratio

The SSVP is better suited in some ways to odds ratios rather than probabilities. The SSVP for mixed evidence is formed by taking the odds ratio between the expectations calculated as if the evidence was one sided.

C + k*p* : B + k*~p*

This gives PK(*p*)B – C = (C + k*p*) / (B + C + KP)

Thinking about the SSVP in this light may help to see how it is constructed and why it works so well in describing and justifying decisions in the face of varying stakes.

First we consider the ratio between the value of evidence k*p* : k*~p*. Thisgives an unweighted odds ratio. Converted into probabilities this give P(*p*) = k*p* / KP and P(~*p*) = k*~p* / Kp where KP = k*p* + k*~p*. It is important to note that when converting into a probability k*p* and k*~p* all become positive. We can see here that using k*p* / K would give a stake size invariant evidential probability that would satisfy the axioms and match the observed frequency.

The value k*p* can be looked at as a weight of reason that pushes the expectation up from C towards B. In this case, even if k*p* is zero, and there is no reason to suppose that *p*, then the expectation is still as high as C. Therefore the value C is a given when arguing in favour of *p*. So we add C to k*p* on the *p* side of the scales.

On the other side of the scales, we have the inverted situation. Any evidence k*~p* can be seen as reason to lower the expectation down from the value B towards the value C. Since the value of k*~p* is negative, this is to *add* the value k~*p* to B. As in the case of k*p*, the ~*p* side of the scales can assume that the expectation will be at most B, and therefore the value B is a given when arguing for ~*p*. We can therefore add B to the *~p* side of the scales. This gives us the odds ratio:

C + k*p* : B + k*~p*

Which is the ratio between the expectations, though in this case, k*~p* has its absolute value.

This gives PK(*p*)B – C = (C + k*p*) / (B + C + K)

### **19.7 Mutual Exclusivity, zero stakes and zero evidence**.

The SSVP only operates relative to a bet with a well defined stake size. This means it does not apply to cases where there is no stake size, including cases where B – C = 0. One may wonder what is to be said in these cases. If an answer is wanted for the purposes of mental exercise and intellectual curiosity, then k*p* / K seems a sensible zero stakes evidential probability. But our opinion is that if there are no stakes, then there is no rational requirement to proportion your belief to your evidence.

As we have discussed, the SSVP is silent in cases where there is no evidence whatsoever. However, for formal reasons it maybe useful to give a value for PK(*p*)B – C  when there is no evidence at all. This would be: C/(B + C) in cases where there is one proposition and its negation forming a partition of two possibilities. We can generalise this to C / (B + NC) for N mutually exclusive propositions and N + 1 possibilities (the extra possibility is formed by the negation of the disjunction of all N propositions). If there is positive stakes on a disjunction of M mutually exclusive propositions in a partition of N mutually exclusive propositions, (and therefore N + 1 possibilities) then the zero evidence probability is MC/(B + NC). This ensures that the zero evidence evidential probability at positive stakes where all that is known is the number of elements in the partition is always less than or equal to the logical probability derived from the principle of indifference.

The thinking here is that knowledge of mutual exclusivity has value. In the case of a single bivalent proposition this value is reflected by adding C to k*p* and B to k*~p* to form the odds ratio. This reflects the a priori knowledge that the law of excluded middle holds with regard to *p*. Each additional mutually exclusive proposition adds an additional possible outcome which is of epistemic value C. The negations of mutually exclusive propositions are not themselves mutually exclusive, so there is always one more possible outcome than there are propositions. This is extra possibility is given the value B. For each additional proposition the added knowledge is that the law of excluded middle holds with regard to the disjunction of all the propositions and the negation of this disjunction. Since the negation of the disjunction is consistent with *~p,* it is only C that gets added for each new proposition. For there to be a stake size at all, there must be at least one possibility in the partition that is being bet against, so there will always be one possibility assigned value B.

So for example, if there are three colours, red, blue and green that are mutually exclusive, and the negation of all these colours, forming a fourth element in the partition, then to bet on red or blue at stakes B – C with evidence K would have the evidential odds ratio:

2C + kred + kblue : B + C + kgreen + k~(red OR blue OR green).

From this we form the evidential probability:

PK(red OR blue)B – C = (2C + kred + kblue) / (B + 3C + K)

In general, when betting on a disjunction *dm* of M mutually exclusive propositions in a partition of N mutually exclusive propositions , which together with the negation of the disjunction of all propositions makes a partition of N + 1 outcomes, then:

PK(*dm*)B – C = (MC + k*dm*) / (B + NC + K)

Of course it is possible to bet at positive stakes on the negation of any disjunction, including the negation of the disjunction of all the propositions. In this case, the value B is given to the negation of the negation, since B is always assumed as the value added to *~p* where *p* is the proposition bet upon. Therefore

PK(~*dm*)B – C = (C + NC – MC + k~*dm*) / (B + NC + K).

At negative stakes, the B stays with proposition bet upon, so:

PK(~*dm*)C - B = (B + NC – MC + k~*dm*) / (B + NC + K).

This only deals with cases where bets have only two possible value outcomes, B and C. We have yet to extend the SSVP to cases where there are multiple outcomes with multiple values. I can see no obstacle to extending the SSVP in this way, other than time and complexity.

### 19.8 Kolmogorov’s axioms.

We will now show that relative to any particular bet with well defined stakes, Kolmogorov’s axioms hold. Kolmogorov’s axioms were created out of measure theory and are therefore not particularly apt for evidential probability. Furthermore they were created in an intellectual environment that was uniquely hostile to value judgements. For these reasons we prefer Ramsey’s “Laws of probability” and in fact it is much simpler to show that the SSVP fits with Ramsey’s laws. However, there may be some doubt as to whether Ramsey’s laws are sufficiently intact after denying the assumption of stake size invariance, and Kolmogorov’s axioms are widely held to give the essential formal properties of any probabilistic theory.

The axioms are:

1*.* Positivity.

*P*(*a*) ≥ 0 for all *a*  *F*

2. Normalization.

P(Ω) = 1

3. Additivity.

*P*(*a*  *b*) = *P*(*a*) + *P*(*b*) for all *a, b*  *F* such that *a* ∩ *b* = .

We can establish that these axioms holds under two conditions:

1. *Identity of Stakes*. The stakes are identical. The axioms only hold when considering bets on the same partition where there are two value outcomes: B if *p* and C if *~p* where “*p*” is the belief that is bet upon.

2. *Completeness of evidence.* The evidence is *complete*. The evidence is complete when it is all taken from the same sample and closed under logic. For example, when considering evidence for whether or not it will rain on a particular day and whether or not it will snow on a particular day, then the evidence is only complete when, on all the days in the evidence sample it was recorded whether or not it had rained and whether or not it had snowed. In this case, the numbers of days when it did rain plus the number of days when it did not rain ought to equal the number of days when it snowed plus the number of days when it did not snow. When the evidence is complete then:

i) Ka = Kb for all propositions, disjunctions and conjunctions in the partition.

ii) kaORb = Ka + kb – ka+b

The completeness condition is only in regard to *inductive* evidence. As we mentioned above, for each extra proposition added to the partition, the value C is added to the value of evidence. However, for convenience of expression, we will not count this knowledge as being a part of KP when writing the SSVP, so that k*~p* refers only to empirical evidence for *~p* and the conceptual knowledge given by the partition will be simply added to the denominator and numerator as appropriate in the form of NC.

The first axiom is that any probability measure is positive.

1*.* Positivity: *P*(*a*) ≥ 0 for all *a*  *F*

In general it is necessary that this axiom holds because evidential probability is evidential expectation based. The evidential expectation is given by:

A = C + PK(*p*)B – C(B – C)

It is necessary that if B > C then A ≥ C; whereas if B < C then A ≤ C

Assume B – C > 0. If PK(*p*)B – C < 0 then the evidential expectation would be less than C, which is absurd.

Assume B – C < 0. If PK(*p*)B – C < 0 then the evidential expectation would be more than C, which is absurd.

Since the SSVP is expectation based, then for all values k*p,* K, B and C:

PK(*p*)B – C ≥ 0

In the case of one sided evidence:

PK(*p*)B – C = k*p* / (B – C) whenever k*p* < (B – C)

Values B and C are always positive. k*p* is positive whenever B – C is positive and negative whenever B – C is negative. Therefore k*p* / (B – C) ≥ 0 as long as B – C does not equal 0.

In the case of mixed evidence:

PK(*p*)B – C = (C + K*p*) / (B + C + K)

This value is always positive because B, C, K and k*p* are always positive.

It might be objected that k*p* is negative whenever B – C is negative. Although it is true that k*p* is considered negative when B – C is negative when forming the one sided probability, because the mixed evidence SSVP is formed from the odds ratio, then all values become positive. This may seem odd, but when transforming odds ratios into probabilities, both numerator and denominator become positive even though they pull the expectation in opposite directions. The SSVP can be understood as taking a naïve frequentist intuition that the evidential odds ratio would be k*p* : k*~p*, and then weighting this value by adding C to the for side and B to the against side, thereby forming C + k*p* : B + k*~p*. Although any bookie will interpret the left column as negative and the right column as positive when it comes to settling bets; when forming the probability all these values are taken to be positive.

2. Normalisation: P(Ω) = 1

This axiom can be interpreted as the requirement that the entire set of possibilities must sum to 1. Since the entire set of possibilities relative to a bet that results in B if *p* and C if *~p* is covered by *p* at positive stakes and *~p* at negative stakes then we take it that the first axiom is satisfied just in case:

PK(*p*)B – C + PK(~*p*)C - B = 1.

This necessarily follows from the SSVP for mixed evidence.

PK(*p*)B – C = (C + k*p*)/(B + C + KP)

PK(~*p*)C – B = (B + k*~p*)/(B + C + KP)

Therefore:

PK(*p*)B – C + PK(~*p*)C – B = (C + k*p*)/(B + C + KP) + (B + k*~p*)/(B + C + KP) = 1

It also follows when the evidence is one sided and k*p* < B - C

PK(*p*)B – C = k*p*/(B – C), whenever k*p* = KP

PK(*~p*)C – B = (C – B + k*p*)/(C – B), where it is presumed that k*p* is positive if C – B is negative and vice versa.

Hypothesis: k*p*/(B – C) + (C – B + k*p*)/(C – B) = 1

1. (C – B + *kp*)/(C – B) = (B – C – k*p*)/(B – C)

2. (k*p*/(B – C) + (B – C – k*p*)/(B – C) = 1, except when B – C = 0.

When the evidence is one sided and k*p* > B – C then PK(*p*)B – C = 1 and PK(*~p*)C – B = 0 by definition.

This covers all cases where the stakes on *p* are B – C and the stake size on *~p* = C – B. Because of the way that stakes are defined, then necessarily whenever the stakes on *p* are B – C then the stakes on *~p* are C – B. Therefore the axiom holds whenever there is any non zero stake size.

We have thus shown that it is a consequence of the SSVP that PK(*p*)B – C + PK(*~p*)C - B = 1

When there is a partition made up of N mutually exclusive propositions, then for any *dp* and *~dp* where *dp* is a disjunction made up of M propositions in a partition made up of N propositions, then:

PK(*dp*)B – C= (MC + k*dp*) / (B + NC + K)

PK(*~dp*)C – B = (B + NC – MC + k*~dp*) / (B + NC + K)

Therefore, PK(*dp*)B – C + PK(*~dp*)C – B = 1.

We have thus shown that the normalization axiom holds for any partition where each element in the partition is either bet against or bet for at the same stakes.

3. Additivity: *P*(*a*  *b*) = *P*(*a*) + *P*(*b*) for all *a, b*  *F* such that *a* ∩ *b* = .

This axiom is called the axiom of additivity and is considered slightly more controversial than the other two axioms. The axiom is important because it allows the introduction of further divisions to the partition. It does this via the introduction of the concept of mutual exclusivity. Mutual exclusivity is introduced via the stipulation that *a* ∩ *b* = . This stipulation has the effect of adding an extra possibility to the partition. The general rule for mutual exclusivity is that for N mutually exclusive propositions there are N + 1 possible outcomes. The extra possibility is the negation of the disjunction of all the propositions.

Kolmogorov’s axiom can be shown to hold for partitions with any number of mutually exclusive elements:

Proof: First take a partition of 2 mutually exclusive elements: *a* and *b.* Assume that the evidence is complete, so that Ka = Kb = KaORb = K and k*aORb* = K*a* + K*b* when *a* and *b* are mutually exclusive.

Prove that PK(*a*)B – C + PK(*b*)B – C = PK(*a*OR*b*)B – C

1. PK(*a*)B – C = (C + k*a*) / (B + 2C + K)

2. PK(*b*)B – C = (C + k*b*) / (B + 2C + K)

3. Therefore PK(*a*)B – C + PK(*b*)B – C = (2C + K*a* + K*b*) / (B + 2C + K)

4. PK(*aORb*)B – C = (2C + k*a* + k*b*) / (B + 2C + K)

5. Therefore PK(*a*)B – C + PK(*b*)B – C = PK(*aORb*)B – C

We can now generalise to disjunctions containing any number of elements. Suppose that *a* is a disjunction containing L elements and *b* is a disjunction containing M mutually exclusive elements in a partition of N elements. The disjunction *a*OR*b* therefore contains L + M mutually exclusive elements.

1. PK(*a*)B – C = (LC + k*a*) / (B + NC + K)

2. PK(*b*)B – C = (MC + k*a*) / (B + NC + K)

3. Therefore PK(*a*)B – C + PK(*b*)B – C = (LC + MC + K*a* + K*b*) / (B + NC + K)

4. PK(*aORb*)B – C = (LC + MC + k*a* + k*b*) / (B + NC + K)

5. Therefore PK(*a*)B – C + PK(*b*)B – C = PK(*aORb*)B – C

We have thus shown that additivity holds for a partition of any size.

A further question of interest is whether additivity holds in the case of a countable infinity of mutually exclusive propositions. There is no special problem for the SSVP here. We merely replace N with ∞ and obtain:

PK(*a*)B – C = (LC + k*a*) / (B + ∞C + K)

PK(*b*)B – C = (MC + k*a*) / (B + ∞C + K)

PK(*aORb*)B – C = (LC + MC + k*a* + k*b*) / (B + ∞C + K)

Whether these values have any meaning is up to enthusiasts about infinity to decide. It should be noticed that knowledge of the mutual exclusivity of the infinite partition is itself of infinite value. The problem often posed for countable additivity is that if the probability of each proposition is zero, then the probability of the disjunction of the entire partition should also, according to additivity, be zero; whereas if the probability of each member is not zero, then the sum of the probabilities of the disjunction of all the partition will sum to greater than 1. There are two things to say about this with regards to the SSVP.

Firstly it is not immediately obvious to me that (C + k*p*)/(B + ∞C + K) = 0. To my mind this is an arbitrary stipulation and one may as well call this value undefined, or *almost zero*, or infinitesimal. Let *dp* be the disjunction of every proposition in the partition except for the negation of the disjunction of all the propositions in the partition. In this case it is plausible that the sum of the probability of every proposition in the partition excluding the negation of all the propositions sums to neither infinity nor 0 but rather 1 – (B + k*~dp*) / (B + ∞C + K). We can call this *almost evidentially certain*. Almost evidential certainty is not sufficient to entail truth.

Secondly, as we have already mentioned, the Kolmogorov axioms only hold when there is a non zero stake size. In order for there to be a non zero stake size, then there must be at least one proposition not included in the disjunction. Therefore the problem does not arise. It can be demonstrated that the normalization axiom still holds in the infinite case:

Let *f* be a proposition in a countably infinite partition.

PK(*f*)B – C = (C + k*f*) / (B + ∞C + K)

PK(~*f*)C – B = (B + ∞C – C + k*~f*) / (B + ∞C + K) = 1 – (C + k*f*) / ((B + ∞C + K)

Therefore PK(*f*)B – C +PK(~*f*)C – B = (B + ∞C + K) / (B + ∞C + K) = 1

If there is any problem here, it is due to the concept of infinity and not due to the SSVP. So long as values like (C + k*f*) / (B + ∞C + K) are intelligible, then it is possible to have intelligible degrees of belief in an infinite partition which conform to the axioms of probability as laid out by Kolmogorov.

### 19.9 Conditional Probability

We are now left with the definition of conditional probability which completes Kolmogorov’s treatment of probability. Conditional probability *p* | *q* is introduced by what has come to be known as the “ratio definition”. The ratio definition for conditional probability is:

P(*p* | *q*) = P(*p*&*q*) / P(*q*)

This is not an axiom, but a definition. It defines conditional probability in terms of unconditional probability. However a great many probabilists and philosophers of probability are unhappy with this aspect of Kolmogorov’s treatment of probability. The main complaint is that the definition leaves conditional probability undefined when the probability of the condition is zero.

Rival axioms of probability treat conditional probability as basic. Ramsey’s theory of probability has this feature. Conditional probability is introduced in terms of the expectation on a conditional bet. This is a fundamentally different approach from Kolmogorov’s.

This is how Ramsey defines a degree of belief in *p* | *q* in terms of the expectation on a conditional bet.

If a subject is indifferent between:

1. B if *q*, C if ~*q*.

2. D if (*p & q*), E if (~*p & q*), C if ~*q*.

Then S’s degree of belief that p given q = (B – E)/(D – E).

This ought to mark the expectation on a bet that results in D if *p* and E if *~p*, where the bet is only valid if *q*. For reasons of completeness, Ramsey assumes that there is already a bet on *q* that pays B if *q* and C if *~q*. It then follows that, given stake size invariance, the subject should only be indifferent between 1 and 2, if his degree of belief bel(*p&q*) = bel(*p* | *q*)bel(*q*).

However, this is where the SSVP runs into problems. There is not one, but two stake sizes on the conditional belief. There is the stake size on *q* which is B – C, and there is the stake size on *p* | *q*, which is D – E. This means that the identity of stake size does not hold, which is one of the conditions for the axioms of probability to hold.

Furthermore, the completeness of evidence does not hold for conditional probability, since evidence for *~q* is relevant for P(*q*), but is not relevant for *p* | *q*, so the value of evidence Kq ≥ Kp|q.

Because these two conditions are not met, then the ratio definition of conditional probability does not need to hold for the SSVP for coherence in that a mere failure of the ratio definition will not necessarily lead to vulnerability to a Dutch book.

The failure of conditional probability to meet the two conditions of completeness of evidence and identity of stakes are substantive features of the essence of conditional probability, and this is why the difficulty of showing that SSVP conforms to the ratio definition of conditional probability is a strength of the SSVP rather than a weakness.

Firstly, the stake size: Let *q* be the proposition that a six is thrown on the first roll of a fair die. Let *p* be the proposition that a six is thrown on the second roll of the same die. Suppose you started with £100. You take an unconditional bet on *q* that pays £5 if *q* and costs £1 if *~q*. Therefore B = £105 and C = £99. You then take out another conditional bet on *p* | *q* that pays £30 if *p* and costs £6 if not *p*. This conditional bet is only valid if *q*. So on this bet D = £135, E = £99 and C = £99. It should be clear that these two bets taken together are equivalent to a bet that pays £35 if *p*&*q* and cost £1 if *~p&q*, making P(*p* & *q*)D – C = P(*p* | *q*)D – C | B – C P(*q*)B – C.

Here we have a phenomena called *stake size escalation*. The stake size on *p* & *q* must be six times higher than the stake size on *q* for the identity to hold. It is not clear what the stake size on the conditional probability taken in isolation is since there are three possible value outcomes rather than two: *p&q*, *~p&q* and *~q*.

Secondly, if the value of evidence for *p* and *q* is complete, then the value of evidence relevant to *p* | *q* will be less that the value of evidence relevant to *q.* For example, suppose a cohort of 10 000 infants were studied and it was recorded how many (k*b*)were put to bed on their tummies (*b*) and how many (k*c*)died of cot death (*c*). Because of the completeness of the evidence sample Kb = Kc. To clarify the example, let us suppose the evidence pertaining to one baby dying of cot death is at stakes 1U so that Kb = 10 000U. However, if we look at the evidence relevant to whether a baby died of cot death given that it was put to bed on its tummy (*c* | *b*), then only k*b* is relevant. Therefore Kc | b= k*b* ≠ Kb.

We can see the epistemic significance of this if we fill in some values. Let us pretend that k*b* = 1000U and k*~b* = 9000U, so that each baby has a stake size of 1U. In other words, 1000 out the 10 000 babies were put to sleep on the tummies. Now let us pretend that k*c* = 5, and k*~c* = 9995, in other words, 5 out of the 10 000 died of cot death. Let us pretend that k*c* | *b* = 1. In other words, 1 out of the 1 000 babies who were put to bed on their tummies died of cot death, whereas 4 out of the 9 000 babies who were not put on their tummies died of cot death.

If we ignore stake sensitivity and use P(*a*) = k*a* / Ka, it looks like the ratio definition holds:

P(*b*) = .1, P(*c&b*) = .0001, therefore P(*c* | *b*) = 1/1000 = (k*c* | *b*) /( Kc | b).

From a similar calculation we obtain P(*c* | *~b*) = 0.0004/.9 = 1/2250.

However, anyone seriously interested in avoiding cot death should not conclude from this pretend study that you can reduce the probability of cot death by over 50% by laying your baby to rest on its stomach. Such comparisons should be treated with caution because the two conditional probability judgements are based on different sets of evidence. The belief in (c | ~b) is supported by 9 times the amount of evidence in support of the belief in (c | b). This difference in evidential support ought to make (c | ~b) more resilient in the face of new evidence and more resilient in the face of differing stake sizes. These two forms of resilience are more closely related than they may first appear to people used to deliberately ignoring value judgements when doing statistics.

Let me explain through an example. Suppose we were to use the evidence from this study to form a policy that would effect 10 000 future babies. The greater resilience of c | ~b in the face of new evidence means that we can expect that our belief in (c | ~b) is likely to *change less* after the results from the next 10 000 babies comes in, than is our belief in (c | b). If we accept this, then it must also be true that our evidential expectation based on (c | ~b) is likely to deviate less from the actual result than an expectation based on (c | b).

We can see with the example that these differences are reflected in the SSVP.

Let us suppose the stake size on c | b was B = 0 and C = U. This is to reflect the fact that one baby dies if it gets cot death.

Pkb(*c* | *b*)0 - U = (U + U)/(U + 0 + 1000U) = 2/1001.

Pk~b(c | *~b*)0 ­– U  = (U + 4U) / (U + 9000U) = 5/9001.

We can see that at negative stakes, the difference between the two beliefs gets exaggerated. This is because (c | b) has less evidential support, and therefore increases more due to the negative stakes. This reflects the fact that we should be cautious of a negative deviation on our expectation due to the lower resilience of the less evidenced belief.

At positive stakes the difference goes the other way.

Pkb(*c* | *b*)U - 0 = (0 + U) / (1001U) = 1/1001.

Pk~b(*c* | *~b*) U - 0  = (0 + 4U) / (9001U) =4/9001.

In both cases (*c* | *b*) is more sensitive to stake size. A measure of resilience is the *bid –ask spread*. This is the difference between the probability at positive stakes and at negative stakes. The bid ask spread is 1/1001 on (*c* | *b*) but is only 1/9001 on (*c* | *~b*). The general rule is that the bid ask spread is inversely proportional to the value of evidence relevant to a claim. To be precise the bid ask spread is given by:

(B – C) / (B + C + K)

In other words, the bid ask spread decreases with the value of evidence and increases with stake size as a proportion of B + C. This is a strength of the SSVP as studies of markets show that bid-ask spreads do vary with stake size and evidence in this way. It also means that the SSVP cannot be shown to satisfy the Kolmogorov definition of conditional probability. This is because Kolmogorov’s definition is blind to the value of evidence and the stake size, and has the assumption of stake size invariance built in to the definition. This is a defect we can bear because the ratio definition of conditional probability faces some formidable objections. (See Hajek 2003).

There is something further we can say about conditional probability and this is if we take the one sided SSVP. Let us suppose that there had been no evidence in favour of a proposition in an evidence sample. For example, suppose there had been no instances of two siblings both dying of cot death. In this case we can be certain up to stakes 10 000 that two siblings will not both die of cot death (supposing we have no other evidence). Of course this certainty depends on the external evidence of the projected set of babies who define the stake size. However, applied to the populations larger than 10 000 infants, this certainty vanishes. This is because the stake size exceeds 10 000U so that k*p* < B – C and PK(~*s*)B – C = k*p* / B – C. What this means in practical terms is that a parent of two can be certain with a very high level of certainty that it won’t be the case that both his children die of cot death. Whereas a policy maker who is responsible for a population of 20 000 should have a degree of belief of only ½ that out of all the children in his care, there will be no cases where two siblings both die of cot death.

Now given the ratio definition of conditional probability and given that the chance of at least one sibling dying at stake size U – 2U = (2U + 5U)/(10 000U + 3U) = 7/10 003, and the probability of both siblings dying at stake size U – 2U = 0, then the conditional probability of the second sibling dying given the first dying ought to be:

PK(1st & 2nd)U – 2U/PK(1st)U – 2U = 0/7/10 003 = 0.

However, this ought not to give much reassurance to a parent whose first baby has died of cot death. The evidence sample from which this judgement is made is only 5U making the value of evidence much lower. Furthermore, we have a new reference class set up of the parent’s children dying of cot death which includes a surprising new bit of evidence. We also have a change of stakes. The stakes U – 2U were based on the parent having 2U, i.e. two children if neither died of cot death, but just U, i.e. one child, if one died of cot death. So the stakes on the second dying given the first dying are 0 – U. Given this, it would be foolish to assign probability 0 to the second dying given that the first died. From the perspective of a parent who has had one baby die, then the probability of the second dying could be calculated thus: k*(p*) = 6U, KP = 10 001U, B = 0, C = U. Therefore PK(*p*)B – C = 7/10 002. This is very slightly higher than the probability of the first dying. This calculation ignores the fact that the two babies are in the same family which ought to be considered relevant. This could be dealt with if we took the parents relatives to form a separate evidence sample. Using this reference class, the fact that the first baby is in the same family as the second baby would make the death of the first baby much more epistemically significant when considering the probability of the death of the second baby. However, we should be wary of attempting to combine these different sources of evidence, since the object proposition falls under a different description. All that is being said here is that one should be very careful when considering conditional probabilities, and a blind use of the ratio definition is likely to lead to misleading conclusions. The important factors are:

1. The stake size on a conditional is the ratio of two stake sizes and therefore is not identical with the stake size of the conjunction or the condition.

2. The value of evidence relative to the conditional is likely to be less than the value of evidence relative to the conjunction and the condition.

3. The condition itself ought to provide additional evidence relevant to the conjunction and therefore to the conditional. When considering the conditional, we must hypothetically add the evidence of the condition to the value of knowledge relevant to the conjunction.

Finally, we will show that supposing the stakes to be identical then the ratio definition in fact provides an upper limit to conditional probability:

PK(*p*&*q*)B – C/PK(*q*)B – C = (C + k*p&q*)/(B + C + K) ÷ (C + k*q*)/(B + C + K) = (C + k*p&q*)/(C + k*q*)

Now, if we were to take a case where the conditional probability PK(*p&q*) was calculated on the basis of the same evidence at the same stake size, then k*p* | *q* = k*p*&*q*; and Kp | q = Kq. Therefore:

PK(*p* | *q*)B – C = (C + k*p&q*) / (B + C + k*q*) ≤ (C + k*p&q*)/(C + k*q*).

From this we get the constraint on conditional probability:

PK(*p* | *q*)B - C ≤ PK(*p*&*q*)B – C/PK(*q*)B – C.

This is because the conditional probability must support higher stakes when combined in a multiplier bet, and the conditional probability is less supported by the evidence.

In conclusion we hope to have shown that the ratio definition does provide a constraint on conditional probability, but is not a condition on coherence, since the conjunction is at higher stakes than the condition when the definition must hold, and the value of evidence for the conditional is lower than the value of evidence for the conjunction and the condition.

### 20. Subjective level of certainty

Our analysis of knowledge sets the necessary degree of belief for knowledge at 1, or subjective certainty. Degrees of belief that conform to the Stake Size Variation principle can be evidentially certain when the value of knowledge K*p* is greater than the stake size, and less than certain when the value of knowledge K*p* is less than the stake size. This means that it is possible for a subject to be certain at low stakes, and yet less than certain at high stakes. We can name the threshold stake size above which the subject is no longer certain, the *subjective level of certainty*. If the subject conforms to the Stake Size Variation Principle, then the subject’s subjective level of certainty will be equal to K*p,* the value of knowledge in favour of *p*.

The Stake Size Variation Principle has the more general consequence that people whose degrees of belief conform to the Stake Size Variation Principle can have higher degrees of belief at lower stakes than they do at higher stakes while remaining invulnerable to Dutch books and therefore arguably probabilistically coherent.

Our theory here presented thus supports both contextualist and interest relative invariant accounts of knowledge. It also explains the findings of Tversky and Kahneman (1979) that people are risk averse over gains but risk seeking over losses. In particular it interprets Tversky and Kahneman’s *decision weights* as stake size variant degree of beliefs, and shows that they do, counter to their own admission, conform to the axioms of probability as laid out by Ramsey. It also explains people’s (including Savage’s) reactions to the Allais problem.

The idea that people can be certain at low stakes but less than certain at high stakes is not so controversial as it may first appear. Both Ruth Millikan and Timothy Williamson have observed that knowledge or full belief can be good up to a maximum stake size, and fail beyond that threshold:

"Outright belief still comes in degrees, for one may be willing to use *p* as a premise in practical reasoning only when the stakes are sufficiently low".

(Williamson, T. 2000, p. 99)

"Unless the stakes are high, we do not require of the person who "knows that *p*" because he has been told by so-and-so that he be ready to place extremely high stakes on the truth of *p*,"

(Millikan, Ruth.1993, p. 253)

Popper has made similar observations.

“There exists a subjectivist theory of probability which assumes that we can measure the degree of our belief in a proposition by the odds we should be prepared to accept in betting. This theory is incredibly naïve. If I like to bet, and if the stakes are not high, I might accept any odds. If the stakes are very high, I might not accept a bet at all.” (Popper, 1974 p.79n)

Kant from the *Critique of Pure Reason* practically lays out the theory that certainty is dependent on stakes size, and that a subject who is certain at low stakes, will not necessarily be certain at high stakes. (Kant 1961 A825 B853 p 648)

“It often happens that someone propounds his views with such positive and uncompromising assurance that he seems to have set aside all possibility of error. A bet disconcerts him. Sometimes it turns out that he has a conviction which can be estimated at one ducat but not at ten.”

Replace ducats with units of utility and what Kant is talking about here is a measure of the subjective level of certainty.

### 21. Evidential level of certainty

The evidential level of certainty is the maximum stake size at which a body of evidence is sufficient to settle a bet. This is given by the value of knowledge K*p*. This is the evidence required to dispel all reasonable doubt at the specified stake size.

The evidential level of certainty is set by the size, relevance and breadth of the reference class. The evidence that warrants certainty that *p* will consist of a series of bets in the past on propositions in the same reference class that have all been successful. In any such case, the hypothesis that a future bet on *p* will be successful has the most likelihood given the data. However, as the sheer number of bets in the reference class goes up, the likelihood of any possible competing hypothesis goes down. There is no space for a full discussion here, but there exist various statistical tools that could be pressed into service in extracting the level of certainty from the data, including likelihood ratios, significance testing, and measures of variance.

A feature of our analysis is that the reference class must include the bet or bets in question. This means that a belief in *p* is only certain if *p* is true. Therefore knowledge entails truth as a consequence of the second condition.

The level of certainty is a utility measure and does not need to conform to the axioms of probability.

### 22. Scepticism

An advantage of this analysis of knowledge is it gives a solution to the problem of scepticism. Sceptical arguments are problematic because they seem to force us into the conclusion that most if not all our everyday knowledge attributions are false. Perhaps the most famous sceptical argument is that of dream scepticism. Suppose I deal a card face up on the table and it appears to me to be the ace of spades. Call this visual appearance E. This visual appearance by ordinary standards suffices as conclusive evidence that:

ACE: There is an ace of spades face up on the table.

Consequently I claim to know that ACE since:

1. I do not doubt ACE

2. Given E, I am right not to doubt that ACE.

The inference from E to ACE we can think of as being in a class of inferences from visual appearances of type E to propositions of type *p*. We then claim that 2 holds because all inferences in this class are successful.

The sceptic points out that if I was dreaming, then E could be true and ACE false. In our terms this would be an instance of a failed inference in the relevant class of inferences. The sceptic then claims that:

2. Given E, I am not right not to doubt that ACE.

Therefore I do not know that ACE.

The sceptical argument is powerful because no one would claim that the visual appearance of a playing card constitutes conclusive evidence that I am not dreaming, since such visual appearances are the stuff of dreams.

In our analysis, knowledge ascriptions are relative to a set of actions modelled by bets. Accordingly, one is right not to doubt ACE given *p* iff all bets made on the basis of the class of inferences from E to *p* are successful. Our answer to the sceptic is that the dream hypothesis does not provide an adequate counterexample, since even if the dream hypothesis is true, then bets made on the basis of inferences from E to *p* will still be successful. This is because if the dream hypothesis is true, then any bets on *p* are also made in the dream, in which case dream evidence E will be sufficient to settle them. If I dream that I bet a million pounds that the card is an ace and go on to dream that I see that the next card is an ace, then in the dream I win the bet. It is simply not the case that I owe anyone a million pounds when I wake up.

One may object that this is a little too quick. The doubter might claim that the visual appearance of the Ace was in a one second dream, while the bet took place in the real world. In this case, the sceptic still can’t win unless the card appears not to be an Ace to the sceptic. The point of the example is that the bet settling evidence must satisfy both parties to the bet. Either it does, or it does not. If it does not, then it is not bet settling evidence and we have a case where the sceptic is correct, and the object proposition is not known. Whereas if it does, then the sceptic has been proven wrong. So for the sceptical scenario to work it must apply to both parties, in which case, the bet itself is undermined. Of course, a sceptic outside the bet could doubt whether the bet should have been settled. But it would be hard for such a sceptic to explain who exactly he thought he was talking to, and what exactly he was talking about. (Is the bet in his dream? Is he correctly witnessing two dreamers betting? Is he correctly witnessing a one second dreamer betting with a non dreamer?)

Hence we get the conclusion that most of what we believe on the testimony of our senses is knowledge.

But what about the dream hypothesis itself and other sceptical hypotheses like the brain in the vat hypothesis and the painted mule hypothesis? Do we know that these things are false? This is a distinct question. In the above example, the doubter was using the uneliminated possibility of the dream hypothesis to undermine knowledge of card being an ace. But what if we were to consider a bet on whether I am now dreaming? Or whether I am now a brain in a vat?

In our analysis, knowledge ascriptions must be relative to a set of actions that can be modelled by bets. The thing in common with sceptical hypotheses is that it is very hard to come up with actions that depend for their success on the sceptical hypothesis being false. This is because the hypothesis, if true, will in nearly all conceivable cases, not only affect the success of any proposed action, but affect its failure too. If an event happens that affects both the success and the failure of a bet, then the stakes are ordinarily returned making the stake size zero. When the stake size is necessarily zero, then the denominator is zero and the rational degree of belief is undefined. The truth of the brain in a vat hypothesis, if proven, would have the effect of invalidating any bets made whilst in the vat. Most of us will never perform an action in our lives that depends for its success on our not being a brain in a vat. The painted mule hypothesis may be an exception to this, but even in this case, there are few actions that any of us are likely to perform that will succeed if a particular animal is a zebra, but fail if it is a painted mule. For those actions that do fall into this category, then the evidential standards for whether the animal is a zebra will be higher, and include proof that it is not a painted mule.

In short, the sceptical hypothesis does not constitute a legitimate doubt, so does not undermine evidential certainty.

### 23. Gettier counterexample to the tri partite analysis of knowledge.

Our analysis is likewise not vulnerable to Gettier counterexamples. The straightforward reason is that we do not accept that one can be justified in having a full belief in a false proposition. From this it follows that one can never be evidentially certain of proposition if any member of the relevant reference class is false. Gettier’s first counterexample is as follows:

Smith has an assurance from the president of the company that Jones has got the job, and has counted ten coins in Jones’s pocket. Call this evidence E. This he takes to be conclusive evidence that:

(p) the man who gets the job has ten coins in his pocket.

So Smith:

1. Does not doubt that p.

2. Given E, is right not to doubt that p.

But it is in fact Smith himself who gets the job, and Smith has ten coins in his pocket. Most, (though not all)[[4]](#footnote-4) people accept that in this case, Smith does not know that *p* even though he has a justified true belief that *p*.

We contend that Smith fails to know that *p*, since his evidence is not in fact conclusive that *p* in the context that we as the readers inhabit. This is because the inference class to which *p* given E belongs does not always lead to successful bets. This is because E is not sufficient evidence to settle bets on *p* with anyone who is aware that Smith gets the job. Since this includes the readers of Gettier’s paper we simply do not accept that counting ten coins in Jones’s pocket is conclusive evidence that the man who gets the job has ten coins in his pocket in the context of knowledge attribution. In fact it is not evidence at all.

The reason why the Gettier counterexamples had such an impact is that it is very easy to generate “Gettier-style” examples. As ad hoc repairs to the tri partite analysis proliferated, so did new Gettier-style counterexamples for which the ad hoc repair failed. This generative mechanism produced a vast literature that is impossible to survey.

Our solution is not an ad hoc repair. It follows from the probabilistic analysis of knowledge that we argue for on independent grounds. The culprit in this first counterexample is the inference from the director’s assurance that *p* to *p*. It is easy to imagine cases where the director’s assurance that Jones will get a job is sufficient evidence to settle bets that Jones will get the job. We think then that, at least internally, Smith is justified in believing that Jones will get the job. But we are not attending very well to the information we have been given. We don’t know this mysterious director or the reliability of his assurances. The one assurance of his that we know about has been false. So from our perspective there is good reason that Smith should doubt this director’s assurances. This does not mean that there is good reason to doubt all directors’ assurances everywhere. Nor does it mean that we should always doubt even this director’s assurances. But the reference class of the director’s assurances that include his assurance that Jones got the job contains at least one member that is not successful, and this is the assurance that Jones got the job itself.

In short Jones does not know, because Jones is not evidentially certain. Jones is not evidentially certain because the inference from the director’s assurance that *p* to *p* is not always successful, so the evidential probability based on the director’s assurance is less than certain.

### 24. Conclusion

In this chapter we have laid out the raw bones of the probabilistic theory of knowledge. This is that the two conditions for knowledge are

1. Subjective certainty.

2. Evidential certainty.

These conditions are necessary and jointly sufficient. The factivity of knowledge follows from the second condition since it is impossible to be evidentially certain of a false proposition.

Given the interpretation of certainty in terms of bet settlement, this theory of knowledge is fairly conservative. It merely claims that knowledge entails full belief supported by conclusive evidence. It deals with traditional problems by using probability theory to give a precise account of the level of belief and evidential justification required for knowledge. The main area of controversy is in the claim that evidential certainty varies according to interest.

In chapter two we will make the case that certainty, both subjective and evidential, is necessary for knowledge, countering the objection that certainty is too great a demand.

In chapter three we will argue for the interest relativity of evidential probability. In doing so we will formally introduce the Stake Size Variation Principle which states precisely how evidential probability varies according to stake size.

In spite of the probabilistic theory of knowledge’s conservatism and its success in dealing with traditional epistemic problems and solving many long standing problems for decision theory, there is one recurring objection which is that the Stake Size Variation Principle is over ambitious and at odds with the decision theoretic literature. I am frankly puzzled by this type of objection, which seems to me another way of saying that the Stake Size Variation Principle is an interesting idea that no one has thought of before. The only counter to this charge of originality and ambitiousness is to claim that, in spite of appearances, the Stake Size Variation Principle has been thought of before and is in any case not a particularly interesting idea.

In chapter three I attempt to show that in fact the financial markets already behave as if their beliefs were subject to the Stake Size Variation Principle and that many paradoxes for decision theory are caused by the fact that people’s degrees of belief in general conform to the Stake Size Variation Principle, instead of to stake size invariant objective probabilities as is often assumed by text book decision theory. This might mitigate the charge of being original somewhat, since if the majority of people already conform to the principle then it is not so much that I have thought of a new idea as that I am merely reporting an already well studied phenomena. Further mitigation from the charge of wanton originality can come from the equation of stake size variant degrees of belief with Kahneman and Tversky’s *decision weights*. Kahneman and Tversky won the Nobel prize for their contribution to economics and their descriptive model of human economic behaviour includes the hypothesis that people do not use probabilities to calculate action guiding expectation, but instead use *decision weights* which can vary according to stake size. My Stake Size Variation Principle merely incorporates this idea into Ramsey’s theory in which degrees of belief are simply defined as whatever calculates action guiding expectation.

I hope to dissuade the reader from thinking the Stake Size Variation Principle too interesting by demonstrating mathematically that it follows from the rather mundane principles that one should buy low and sell high, and that one should proportion one’s degree of belief to actual frequencies rather than hypothetical frequencies. I hope to dispel any remaining interest by emphasizing the homely prudence of avoiding staking too much of value when you don’t know what you are doing. These prudent platitudes should not cause too much alarm.

# Chapter 2.

# THINKING, KNOWING, AND BET SETTLEMENT

## 1. Introduction

Let us suppose you think for good reason that *p*. Let us suppose further that *p* is true. What more do you need to know that *p*? This is the question that this chapter addresses. We answer by giving a bet settlement theory of knowledge, which says you need two further things:

1. You need to become certain that *p*. This means that you need to have no doubt that *p* and that you need to consider the matter *settled.* In betting terms this means that you will consider any bets on *p* settled in favour of *p*. You will thereby consider any bets for *p* at any odds other that 1 : 0 advantageous, and bets at any odds other than 0 : 1 against *p* disadvantageous. Using standard subjective probability theory this means your degree of belief in *p* must rise to probability 1.
2. You also need to have evidence that makes it certain that *p*. This means that you must have access to evidence that settles the matter. In betting terms, your evidence must establish to the relevant community that bets on *p* should be settled in favour of *p*. The relevant community are those that share a set of presuppositions K which specify conclusive evidence conditions for *p*. Using the same method of measuring degrees of belief, evidence E is sufficient to settle bets that *p* iff evidential probability PK (*p* | E) = 1.

In the course of this thesis we will sometimes indicate any shared presuppositions and background knowledge against which E is conclusive evidence with a subscript K, so PK (*p* | E) = 1 means that E is conclusive evidence for *p* when K is presupposed. It may be objected that there is often considerable disagreement within a language community whether or not E provides sufficient evidence to settle bets that *p*. This is not a problem because in cases of disagreement whether E is sufficient to settle bets on *p* there will be equal disagreement as to whether someone in possession of E knows that *p*.

For example, Tom, who has a photograph of Bill standing in front of the Eiffel Tower, has conclusive evidence that Bill has been to Paris. But this is only because it is generally presupposed that the Eiffel Tower is in Paris. If we did not presuppose that the Eiffel Tower was in Paris then Tom’s possession of the photograph would not be conclusive evidence that Bill has been to Paris. Further evidence would be required to settle whether the Eiffel Tower was in Paris before we attributed knowledge and settled bets that Bill was in Paris on the strength of the photo. But ordinarily, Tom would count as knowing that Bill had been to Paris without having any idea how to prove that the Eiffel Tower is in Paris. So the truth of “Tom knows that Bill has been to Paris” depends on facts about what is generally presupposed, facts which are external to Tom. It is therefore possible that Tom, without undergoing any internal change, could go from knowing to not knowing because of a change in the back ground presuppositions. There is a relationship between presuppositions and stake size. A greater stake size means that the burden of proof is higher and less is presupposed.

We will be arguing that both the belief condition and the evidence condition are necessary, since there is plenty enough disagreement about this. But we will also leave it open that, given sufficient attention to the evidential condition, the two conditions are jointly sufficient. The first barrier to joint sufficiency is that it is widely agreed that knowing entails truth, whereas it is less obvious that bet settlement entails truth. The second barrier to sufficiency is that even if we add truth as a third condition, then there still may be situations where the two conditions are fulfilled, the proposition is true, but there is no knowledge. These will be Gettier type counterexamples where the bet settling evidence only accidentally settles the matter. In response to this, we claim that the two conditions could be sufficient if we accept that:

1. No evidence is sufficient to settle bets on a false proposition, and
2. Someone in a Gettier situation does not have sufficient evidence to settle bets.

The argument is this: supposing an agent is certain that *p* and has bet settling evidence that *p*. It would seem that the burden of proof would then be on a third party to claim that they didn’t know that *p*. If the third party claims that the agent does not know that *p* because *p* is false; then it is also the case that the agent’s evidence isn’t bet settling relative to that third party. If the third party claims that the agent does not know because the agent is in a Gettier situation, then it is also the case that the agent’s evidence isn’t bet settling relative to that third party.

However, definitive arguments for the joint sufficiency of these two conditions are beyond the scope of this chapter.

### 1.1 Thinking and knowing.

There is a phenomenon with which we are all intimately familiar and this is the difference between thinking a thing and knowing it. This distinction is of great importance both epistemologically and psychologically. It is on this distinction we will focus our attention, and in order to best do this we will be attending to cases where there is a transition from thinking a proposition to knowing that it is true.I will illustrate with three examples from three main categories of knowledge acquisition.

***Observation***

I see a shape bobbing about in the sea and I think it is a seal. After watching it for several minutes it swims away and I get a good view of its whole body. Now I know it is a seal.

***A priori reasoning***

I think 53 is a prime number since it is odd and not divisible by 5. Mentally I check whether it is divisible by 3, 7 or 11 and conclude that it is only divisible by itself and 1. Now I know that 53 is a prime number.

***Testimony***

Boris Johnson was ahead in the polls for the London Mayoral election. On the day after the election I think he has won. I go to the newsagent and buy a paper. On the front page is a picture of Boris Johnson, the new Mayor of London. Now I know that he has won.

In the examples the transition from thinking to knowing involves an increase in both evidential standing and an increase in conviction. In all three cases, before I came to know that proposition was true, I already believed it to *some* degree, and I already had *some* justification. The transition from thinking to knowing is not one from having no justification to having some justification, nor from disbelieving to believing, nor from believing something false to believing something true. In a minimal sense I already believed with justification that 53 is prime, Boris Johnson is Mayor and the creature in the sea was a seal. The transition is rather that my tentative belief became certain, and my evidence became conclusive. Whilst still only thinking, my justification would have only been to bet at some odds, whereas when I came to know, my justification was sufficient to demand settlement of the bet.

### 1.2 Socrates

The aim of this chapter is not to elucidate the meaning and normal usage of these two English words, “thinking” and “knowing”, but rather to capture the qualitative difference between these two epistemic and psychological categories. This latter project has been of interest to some of the greatest philosophers, though the terms used have not always been “thinking” and “knowing”. A great many canonical texts were written in Ancient Greek, Latin and French. Relying on accepted translations we can see that this psychological and epistemic distinction between thinking and knowing transcends any particular language, culture or historical period.

For example, Socrates[[5]](#footnote-5) in the Meno distinguishes between “correct opinion” and “knowledge” pointing out that correct opinion will lead to successful action in the same cases as knowledge. Socrates uses “opinion” as a state of mind that can fall short of knowledge even though “correct” or “true”. (Meno, 97 – 100).

In the Theaetetus, Socrates distinguishes between thinking and judgement:

“ – when doubt is over and the two voices affirm the same thing, then we call that its ‘judgement.’ So I should describe thinking as discourse, and judgement as a statement pronounced, not aloud to someone else, but silently to oneself.” (Theaetetus, 190).

Of course it is clear that such judgments can be false and therefore not knowledge. One cannot make a statement true just by pronouncing it silently to oneself. But Socrates does describe the *phenomenology* of the transition from thinking to knowing, though it may be no different internally from the transition from thinking to being deluded. The transition necessarily involves the quieting of doubt. While the mind still retains doubt, it does not know, it merely thinks.

In Book VI of the Republic we find Socrates talking again about the difference between knowledge and opinion. In a most remarkable passage where Socrates is at his most speculative, he draws a line with ignorance at one end and knowledge at the other, saying that opinion is expressible as the ratio between either side of a line that bisects this line. He compares this to a similar bisected line with total clarity and light at one end and total obscurity and darkness at the other. He then says “the division in respect of reality and truth or the opposite is expressed by the proportion – as is the opinable to the knowable, so is the likeness to that of which it is a likeness.” Socrates gestures at a yet not fully worked out idea that it is only though the form of the good that opinion approaches knowledge, in the same way that it is only though the light of the sun that we can see things with complete clarity.

Taking a charitable interpretation of Plato’s Socrates, we can see this as an early attempt to measure degree of opinion as a ratio between distance from knowledge and distance from ignorance. In modern terms this is an odds ratio. If we were to understand the good, then we could turn this into a probabilistic measure by setting the value of knowledge at 1. We can express Socrates’ idea for a measure of opinion as a proportion of knowledge in terms of value. This would mean that the measure of opinion is a measure that falls inclusively between 0 and 1 and could be used to calculate expectation given you knew what the value of knowledge was. (I hasten to add that this interpretation is in no way forced by the text, and the usual interpretation is quite different.)

### 1.3 Descartes.

Descartes, in *Rules for the Direction of the Mind*, elected to only assent to what was certain, not what was merely probable.[[6]](#footnote-6) In one of many re-phrasings he says “we reject all such merely probable cognition and resolve to believe only what is completely known and incapable of being doubted.” (Descartes 1984/85, vol. 1p. 10). Using the words “thinking” and “knowing”, Descartes could here be translated as resolving to believe only what he knows and not what he merely thinks. “Believing” should here be thought of in the way that Socrates described judgement in the Theaetetus as a kind of silent assent. This tallies with Descartes’ own words since he makes the same point in *The Meditations* thus: “hold back my assent from opinions which are not completely certain” (Descartes 1984/85, vol.2 p.12). Descartes is using belief and assent interchangeably.

Descartes himself pointed out that it is difficult in practice to elect what to believe. Sometimes through force of habit one continues to assent to a statement even though in a philosophical moment one has elected to withhold assent. If we take “assenting” as indicating a willingness to claim knowledge and act accordingly, then by refusing to assent to what is merely probable, Descartes is raising his own epistemic standards. Thus he could be thought of as resolving only to claim to know what he does in fact know. This would have the effect of both reducing the number of propositions to which he would give assent, and increasing the proportion of true beliefs to false, making Descartes a more reliable witness, but a less productive researcher. We can think of the epistemic project naturalistically (Millikan 1993, p91) as finding the right balance. The higher we raise the assent bar, the less false beliefs we will assent to, but there will be a corresponding lessening of true beliefs. If the bar is already so high that all our beliefs are true, then raising our epistemic standards higher still will have an entirely negative impact, reducing our true beliefs without corresponding increase in the proportion of true to false.

Did Descartes set the bar too high? Let’s suppose that a Cartesian doubter was standing at my elbow at the moment when I came to know that the object in the sea was a seal. Recall that this was when the seal swum in such a way that I got a good clear view of its entire body. The Cartesian doubter and I both share the same evidence, but let us suppose the Cartesian doubter does not assent to the proposition that the object in the sea is a seal. He merely *thinks* it is true. He does not assent to it fully since he reckons it a merely probable opinion quite capable of being doubted. But he is still of the opinion that it is a seal; one can have an opinion without fully assenting to it. We are not in disagreement about whether the object of our attention is a seal: the Cartesian doubter thinks it is, and I am certain that it is. What we are in disagreement about is whether I *know* that it is a seal. In Socrates terms, I have said to myself with a single voice “that is a seal”, whereas the Cartesian doubter is still wondering whether he is dreaming, or whether perceptual experience can be relied upon. He is still thinking about it and he still has doubts. This is clearly a conceptual matter since the same evidence is before us. There is no further evidence that can resolve the matter. Even if we gathered further evidence of a kind that would fully convince the Cartesian doubter that the object was a seal, he could still withhold assent from the proposition that I knew that it was a seal before he did. He could claim I merely thought it was a seal, since I didn’t have good enough reason to *know* that it was a seal. My evidence was not sufficient to silence all his doubts. In this case, though the same evidence was before us, I would think it sufficient to settle a bet whereas the Cartesian doubter would not. We would be in agreement about the seal, but in disagreement about the evidential standards sufficient for knowledge.

### 1.4 Thinking the truth and knowing the truth.

It should be clear from these examples that the difference between thinking and knowing does not reside solely in the truth or otherwise of the proposition. When a person claims to know a true statement, her knowledge claim can be false in one of two ways.

Firstly, as discussed above, although she judges the proposition to be true, her evidence isn’t sufficient for her judgement to count as knowledge.

Secondly, although she claims to know that the proposition is true, she may in fact have insufficient conviction. This second kind of error is flagged up by Hume when criticising religious believers.

"The usual course of men's conduct belies their words, and shows that their assent in these matters is some unaccountable operation of the mind between disbelief and conviction, but approaching much nearer to the former than to the latter." (Hume, 1825 p. 434). Here Hume is claiming that, though people can strongly assent to a proposition, their actions show that they do not have the conviction consistent with their words. For knowledge, one not only needs conclusive evidence, but one also requires a sufficient degree of confidence. Hume recommends that we proportion our belief to our evidence. When our evidence is sufficient for knowledge, then we should have a proportionate degree of conviction, and this should be demonstrated not just through our words, but also through our actions.

To get the contrast clear between thinking that *p* and knowing that *p*, we are holding the truth of *p* constant. We have found that there are two differences between thinking and knowing:

Firstly, knowledge requires a sufficient level of evidence, which we can call conclusive evidence. The evidence sufficient for knowledge must be good enough to dispel all doubt. How much evidence is sufficient to dispel all doubt is unclear and a matter for philosophical enquiry.

Secondly, knowledge requires a sufficient degree of confidence. The degree of confidence necessary for knowledge can be calledfull conviction. This is not equivalent to mere verbal assent, but must also involve a disposition to act with certainty.

### 1.5 Mill on fallibility

It may be observed that it is all very well assuming the proposition to be true, but surely the very point in holding back from assent when the evidence is insufficient is because the proposition might well be false. When assessing how much evidence is needed for knowledge, it seems natural to think that the evidence should guarantee that the belief is true. This leads to the unattractive proposition that if we have enough evidence to claim that an opinion is knowledge, then that opinion is infallible.

J.S. Mill in *On Liberty* said that we have no business assuming that our most cherished opinions are infallible since many of the most well evidenced and widely held opinions in the past are now held to be false. Yet assent and conviction are necessary for action.

If it is true, as history shows us, that our most warranted beliefs are fallible, then infallibility can’t be a necessary condition for knowledge. The argument I am putting forward here is subtle. If we acknowledge, as we should, that we are at our very best fallible beings, but at the same time recognize that we do have a great deal of knowledge, then we have no business assuming that everything that we know is infallible. Mill makes a similar point against the suppression of free speech.

“There is the greatest difference between presuming an opinion to be true, because, with every opportunity for contesting it, it has not been refuted, and assuming its truth for the purpose of not permitting its refutation. Complete liberty of contradicting and disproving our opinion, is the very condition which justifies us in assuming its truth for purposes of action; and on no other terms can a being with human faculties have any rational assurance of being right. […]This is the amount of certainty attainable by a fallible being, and this the sole way of attaining it.” [Mill (1859) p.21]

These great words from Mill I hope illustrate a subtle distinction I want to make between being incapable of being reasonably doubted, and being undoubtable in a logical sense. Of course one can, and may, doubt anything, if one is bloody minded enough and insensitive enough to reasoned argument and the testimony of the senses. If we want to silence *those*kinds of doubts before claiming knowledge, we have a hopeless task. But doubt can be silenced in a reasonable person in shallower ground. All that is necessary is that there is no lack of trying in the subject, no suppression of uncomfortable counterarguments or inconvenient facts. If, with this lack of suppression, still no doubts arise, then we have reached the highest amount of certainty attainable, and surely that is good enough to be called knowledge. In betting terms, evidence sufficient to demand settlement is not necessarily sufficient to reject any new counter evidence. In legal terms, evidence sufficient to condemn a man is not sufficient to rule out the possibility of appeal. This is the difference between certainty and infallibility. Only infallible evidence is unfalsifiable. Evidence sufficient for knowledge must merely be unfalsified.

## 2. Measuring the difference in terms of probability

We can put a measure to the difference in evidence and belief that distinguishes knowing from merely thinking. Both evidence and belief can be measured using probability. The thought is that we, like Socrates in *The Republic*, draw a line with ignorance at one end and certainty at the other. We call certainty 1 and ignorance 0. As the belief approaches certainty, it is represented by a number closer and closer to 1.

We can give a clear sense to degree of belief in terms of probabilities measuring dispositional betting odds, with certainty being measured by 1, or the disposition to settle bets. Given a measure of degree of belief we can have a measure of the degree of evidence. If a degree of belief is appropriate given the evidence to any rational person, then the evidence can be described as warranting that degree of belief. We can then simply quantify the evidence on the same scale as we quantify degree of belief. So, for example, the degree of evidence that warrants a degree of belief 0.7, is a degree of evidence 0.7. The degree of evidence that warrants certainty is 1.

We now have two questions with a straightforward probabilistic answer.

1. What degree of belief, measured as a probability, is necessary for knowledge?
2. How much evidence, in terms of probability, is necessary for knowledge?

We can assume that a degree of belief is warranted by the same degree of evidence. In other words, the evidential probability is equal to the degree of belief the evidence justifies[[7]](#footnote-7). So we measure evidential probabilities in terms of the degrees of belief they justify. This means that, since knowledge should be justified by the evidence, the answer to 2 will be greater than or equal to the answer to 1. We can rewrite the two questions:

1. What minimum degree of belief, measured in terms of betting odds, is necessary for knowledge?
2. What evidence, measured as the degree of belief it justifies, justifies betting at these odds?

We can now see that given that the answer to the first question is degree of belief 1 (certainty), then the answer to the second question must also be 1, since anything less, would mean that the knower’s belief wouldn’t be justified by the evidence.

### 2.1 Ramsey’s measure

In order to measure degree of belief I will make use of Frank Ramsey’s (1926) measure of degree of belief. Ramsey devised a conceptual measure of degree of belief in terms of the agent’s expectation. “I suggest that we introduce as a law of psychology that his behaviour is governed by what is called the mathematical expectation; that is to say that, if *p* is a proposition about which he is doubtful, any goods or bads for whose realisation *p* is in his view a necessary and sufficient condition enter into his calculations multiplied by the same fraction, which is called ‘the degree of his belief in *p*’. We thus define degree of belief in a way that presupposes the use of mathematical expectation.” (Ramsey 1990, p. 70).

Let capital letters A, B and C refer to possible courses of the world ordered according to merit on an interval scale. For simplicity we can think of them as referring to utility measures of outcomes.

If S is indifferent between options

1. A for certain,

2. B if *p* is true or C if *p* is false,

Then S’s degree of belief that *p* is equal to:

(A – C) / (B – C)

This then is the fraction by which he multiplies the goods and bads whose realisation depend on *p*.

(We must here mention that Ramsey supposed that the degree of belief an agent has would be stake size invariant. However, we take this to be false, and assume nothing other than the subject’s degree of belief is measured here relative to the stake size given by B – C.)

If we can model an action by a bet on *p* where the subject is indifferent whether he takes the bet or leaves it, then we can measure his degree of belief. This measure only measures the degree of belief the subject has, not the degree of evidence. The degree of belief could be completely inappropriate or irrational given the evidence that S has that *p*. But a fundamental feature of rationality is that your degree of belief *should*, if you are rational, be proportional to your evidence. Philosophical attempts to justify this normative link are interesting and difficult, perhaps the most famous being Lewis’s (1994) “Principal principle.” But the strategy here is to simply define the degree of evidence in terms of the degree of belief it justifies. So if we assume that S is rational and that S has evidence E and no other relevant evidence, then we can assume that the evidential probability of *p* given evidence E relative to stake size B - C is also equal to:

(A – C) / (B – C)

### 2.2 Certainty is when bets are settled

Given the basic idea of measuring degree of belief in terms of bets, it is easy to show that certainty occurs when bets are settled, in other words when the loser pays the winner the amount agreed. The way the options are worded in Ramsey’s formula makes this analytic. If S is certain that *p* is true, then we can substitute the two options with the following:

1. A for certain, and

2. B for certain or C if *p* is false.

If one is certain of *p* it is rational to be indifferent between these options only when A = B. If A = B then S’s degree of belief that *p* will equal:

(A – C) / (A – C),

Which is equal to 1 in all cases where A ≠ C.

If on the other hand S was certain that *p* was false, then S would only be indifferent when A = C, which would make S’s degree of belief that *p* will equal:

(A – A)/(B – A),

which is equal to 0 in all cases where A ≠ B.

If one is certain that *p* then it does not matter how good or bad the effects of an action if *~p* is the case. The value of C if *~p* makes no difference in terms of preferability between the options. But this need not mean it is meaningless or redundant. Counterfactuals can still be important even if the antecedent of the counterfactual is certain to be false. For example, it is still important to know that if I fall out of the window I will seriously injure myself even though I am certain that I will not fall out of the window. In fact it is my knowledge of the counterfactual that informs my certainty of the negation of the antecedent.

The value of C does make a difference in one important case, and this is where C = B. In this case the stake size, defined by (B – C), is zero. If C = B, then the agent’s evidence and belief for *p* is irrelevant to his choice between the options. Indifference between options 1 (A for certain), and 2 (B if *p*, C if *~p*)when B = C, fails to define the agents degree of belief that *p* and so does not entail certainty that *p*. We can also see that when B = C, the denominator (B – C) = 0, and the value for the degree of belief is undefined. The coincidence of the mathematical result that whenever B = C the degree of belief is undefined, and the practical intuition that whenever B = C indifference doesn’t not entail certainty lends weight to the adequacy of the measure. We can generalise the intuition by saying that the magnitude (B – C) measures the interest in *p*, or the value of knowledge that *p*. When B – C = 0, then *p* is of no interest and there is no value to knowing that *p*. If B – C is very large, then there will be a proportionate interest in *p* and consequently knowledge that *p* will be highly valuable. (B – C), then, is the *stake size,* and measures the value of knowledge that *p* relative to the bet.

It can be demonstrated that bet settlement is measured by a degree of belief 1. Suppose you have a bet on *p* that results in B if *p* and C if *~p*. Suppose you won the bet, in other words, suppose the bet was settled in your favour. At this point it is clear that the value of the bet is equal to B, since this is how the bet was defined. Therefore it is also true that you should be indifferent between A for certain and the bet on *p*, iff A = B, since the bet on *p* is now worth B. Therefore, if the bet on *p* is settled in favour of *p*, your degree of belief in *p* is necessarily equal to 1.

For example, suppose you had a bet that cost £10 if ~*p* and paid £10 if *p*. Suppose your fortune before taking the bet was £100. In terms of final values, if the bet wins you will have £110 and if the bet loses you will have £90. Now suppose that you have bet settling evidence that *p*. When you collect your winnings you will have £110. This means you ought to be indifferent between winning the bet and having £110 for certain, since winning the bet *is* having £110 for certain. In this case you are indifferent between:

1. £110 for certain

And

2. £110 if *p* and £90 if *~p.*

Making your degree of belief in *p* = (£110 - £90)/(£110 - £90) = 1.

It is important to note that all that has been demonstrated here is that *if* we use betting odds to measure degrees of belief, then bet *settlement* measures a degree of belief 1. We have not demonstrated that this is how certainty should be measured, or that bet settlement always coincides with subjective certainty conceived of in some other way. Bets are often settled on propositions by a pre agreed criterion, for example a boxing match might be settled on the judge’s verdict even though it is widely known that the judges’ verdicts are often controversial and wrong. Given this, punters may well agree to settle bets on the judges’ verdict even though they do not personally agree with the judges’ verdict. However, in this case, the odds at which they originally bet correctly measured their degree of belief in the judges’ verdict, and not in the “true” result of the boxing match. Likewise, the bet settlement measures certainty in the judges’ verdict, not in the “true” winner of the match.

We can see an important epistemological point herein. All successful human cultures have played games and have developed sophisticated criteria for settling bets on those games. This may seem trivial in itself, but with it comes the conceptual apparatus that underlies science, engineering, medicine, politics and commerce. To take a case from medicine, there is often a simple diagnostic procedure to establish whether or not someone has a particular disease. A positive diagnosis, on the other hand, is not identical with the disease. It is always *possible* that a person without the disease could have a positive diagnosis. Given a large enough data set we may be able to give frequency rates for the disease in the population, the rate of disease among people with a negative diagnosis and the rate of positive diagnosis among people without the disease. In this case we could give a probability of disease for someone with a positive diagnosis, and a positive diagnosis would therefore not constitute bet settling evidence for the disease. But this data set would require that there was some further criteria for establishing that someone had the disease. Without the further criteria, there is no way to distinguish any positive diagnosis from any other, so we may as well believe with certainty that anyone with a positive diagnosis has the disease. This is not to say that our current criteria cannot be improved upon. It is rather to say that any improvement in our criteria would give us a finer grained understanding of the disease, rather than a decrease in diagnostic error probabilities. This would have the effect of making further probabilities conditional on the diagnosis sharper. If for example on the first criterion, 55% of people with the disease died, and on a second criterion, 100% died, the second criterion would be a lot more informative with regards to mortality, and so would be a better diagnostic criterion. In the same way, a change of the rules of boxing might lead to fairer, less controversial results. But this would not be by increasing the probability of correctly judging who won any particular game. It would be by improving the game itself. The new improved rules would tend to make the match better reflect the relative merits of the two boxers and therefore make the pre match odds better predictors of the final outcome by making the end results less of a matter of luck and corruption.

So if we buy in to the notion of measuring degrees of belief in terms of Ramsey’s formula, then it deductively follows that bet settlement occurs at probability 1. Since this is defined as certainty then we can give our degree of belief and level of evidence necessary for knowledge as probability 1. This will be equivalent to the degree of belief and level of evidence necessary to settle a bet. “Thinking”, on this model, is measurable in terms of the odds at which you are indifferent between taking the bet or leaving it, whereas knowledge is the level of conviction and evidence required to settle the bet. The question then of how much evidence is necessary for knowledge is thus equivalent to the question of how much evidence do I need to show you before you pay up. We can measure variation in strength of knowledge in terms of the maximum stake size at which the evidence is sufficient for certainty, based on the simple idea that more evidence is required to settle bigger bets, with the limiting case that when the stake size is zero, no evidence at all is needed to settle a bet.

So, let us return to the conceptual dispute with the Cartesian doubter as to the evidential requirements for knowledge. Let us suppose that we share a set of presuppositions K and we both have the same evidence E before us. In the example (see this chapter section 1.3) the evidence was a view of a seal in the sea. The question is whether E gives us grounds sufficient for knowledge that *p*, in the example this was that there was a seal in the sea. We can now formalise the dispute into a dispute as to the evidential probability PK(*p* | E). My judgement is that PK(*p* | E) = 1, whereas the doubter judges PK(*p* | E) < 1. The way of formalising the disagreement does not decide the matter. But it does at least gives us a way of quantifying the dispute. We can ask how important it is whether or not there is a seal in the sea. This will set the stake size, let us say U. I can then propose a bet with the Doubter at those stakes U. The Doubter then sets a price *h* so that I pay him *h* on the promise that he pay me U if I can provide him with conclusive evidence that *p*. This framework does not settle the matter of whether E is conclusive evidence that *p*, I think it is, and the Doubter thinks it isn’t. But it at least allows me to ask the question: “when do I get my money?” This would be to ask the question what further evidence do I need to provide to raise the Doubter’s degree of belief from h/U to 1. We can then establish that the value of this further evidence is equal to U - h to the Doubter. The hope is than when the Doubter is confronted with the practical cost of his doubt, he may relent and agree that after all we do know that there is a seal in the sea from the perspective of our evidence.

## 3 Dutch Book Arguments

The beauty of using bets to measure degrees of belief is that it follows thatan agent’s degrees of belief conform to the laws of probability. Ramsey pointed out that “If anyone’s mental condition violated these laws, his choice would depend on the precise form in which the options were offered him, which would be absurd. He could have a book made against him by a cunning better and would then stand to lose in any event.” (Ramsey 1926 pp. 78).

However, Ramsey assumed that a subject’s degrees of belief would remain the same regardless of the magnitude of the stake, a supposition we called ***stake size invariance***. As we discussed in chapter 1, we do not suppose stake size invariance, but take it to be both descriptively and normatively false. Without supposing stake size invariance, it could be argued that Dutch book arguments can’t establish that degrees of belief must conform to the laws of probability. Objections to Dutch book arguments are often levelled on the packaging principle, which says that the value of two items of value held together has the same value as the sum of the values of the items held separately. Armendt, when considering arguments against stake invariant beliefs (Armendt, 2008, p. 16), argues that the packaging principle on the face of it is not true. However, even without assuming the packaging principle, as long as we have an objective, additive measure of value, which is exchangeable, like money, we can establish some probabilistic constraints on degrees of belief across stake sizes and the law of disjunction for degrees of belief at the same size.

The measure of value must have two features: i) If the subject strongly prefers B to A then the subject will be indifferent between B and A + C for some value C such that B = A + C. ii) The subject has a current state of wealth “A”, and that no price that exceeds A can be paid for any bet.

Given such a measure of value, Dutch book arguments can establish that there are probabilistic constraints on degrees of belief across stake sizes. In what follows we will establish through Dutch book arguments that i) degrees of belief in the same proposition are constrained by upper and lower limits across stake sizes ii) degrees of belief at the same stake size must conform to the law of disjunction; iii) degrees of belief with stakes related logarithmically must conform to the law of conjunction.

### 3.1 Adding and subtracting unconditional bets and the law of disjunction

First we will assume that the subject’s current state of wealth is A. An unconditional bet on *p* is given by two values, the price (A – C) and the stake (B – C). A bet on *p* is a bet where the subject gains the stake and loses the price if *p,* whereas the subject loses the price if *~p*. The probability of *p* expressed by this bet is equal to (the price)/(the stake). We will indicate the stake size by a subscript U, so P(*p*)U= x means the subject’s degree of belief in *p* at stake size U is equal to x. In Ramsey’s theory, if a subject is indifferent between taking the bet and leaving it, then the value of the bet to him will equal his holdings (A). It what follows we will assume that a subject cannot pay a greater price for a bet than his total holdings. This means that for all positive stakes, holdings (A) – price (A – C) ≥ 0.

Now we can now define the addition and subtraction of bets. Adding two bets together is simply to accept the terms of both bets. This is to pay the sum of the prices, and to receive the sum of the stakes on every winning bet.

Subtracting a bet is to convert prices into gains and stakes into losses. So to subtract a bet on *p* is to gain the price in all events and pay the stakes if *p*. When subtracting a bet, the stakes minus the price cannot exceed the subject’s holdings. We can see straight away that a bet on *p* subtracted from a bet on *p* at the same stakes is equal to no bet at all. From this it follows that P(*p*)U = 1 – P(*~p*)-U

We can now give a Dutch Book argument for maximum and minimum cross stakes constraints.

MINIMUM. If S has P(*p*)U = x, then S has P (*p*)V ≥ Ux/V whenever V > U.

This is the rule that you can’t prefer a bet with a smaller prize to a bet with a larger prize for the same price. This interpretation takes the price to be a fixed value rather than a proportion of the stakes. For example, if you were prepared to pay £1 for a bet that paid £10 if *p* and nothing if *~p*, then you should be prepared to pay at least £1 for a bet that pays £20 if *p* and nothing if *~p*. This would make your £10 degree of belief in *p* = 0.1 and your £20 degree of belief in *p* = 0.05, which is 0.1\*£10/£20.

Proof: Suppose S had P(*p*)U = x and P(*p*)V = y such that y < Ux/V. The Dutch bookie would simply give S a bet on *p* at stakes U in exchange for a bet on *p* at stakes V for a small fee of xU - yV, and, if *p* was true, pay the subject U out of his winnings V, leaving the bookie with a profit of xU – yV if *~p* and V – U if *p*.

MAXIMUM. If S has P(*p*)U = x, then S has P (*p*)V ≤ x whenever V > U.

This is the rule that you cannot have a higher degree of belief at a higher stake size than at a lower stake size.

Proof: Suppose the subject had P(*p*)U = x and P(*p*)NU = y such that y > x. The bookie would simply sell a bet at stakes NU for yNU, then buy back N bets on *p* for xU each, and make a profit of yNU – xNU. The bets will cancel each other out, so the Dutch bookie will make this profit whether or not *p*.[[8]](#footnote-8)

We can show that a bet on *p* at positive stakes is equivalent to 1 – a bet on ~*p* at negative stakes.

If S is indifferent between:

1. A for certain

And

2. B if *p* and C if *~p*.

Then S is indifferent between

1. A for certain

And

2. C if *~p* and B if *p*

In which case S’s degree of belief in *p* = (A – C)/ (B – C) at stakes (B – C)

And S’s degree of belief in *~p* = (A – B)/ (C – B) at stakes (C – B)

So

Hypothesis: (A – C)/(B – C) = 1 – ( A – B)/(C – B), unless B – C = 0

1. (A – C)/(B – C) + (A – B)/C – B) = 1

2. (A – C)/ (B – C) + (B – A)/(B – C) = 1

3. (B – C)/(B – C) = 1, unless B – C = 0.

Therefore P(*p*)U = 1 – P(*~p*)-U except when U = 0.

We can now prove the law of disjunction:

DISJUNCTION P(*p* OR *q*)U = P(*p*)U + P(*q*)U – P(*p*&*q*)U

Here is a betting pay off table which shows that at stakes U, this equivalence must hold.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Price/stakes | *p & q* | *p & ~q* | *~p & q* | *~p & ~q* |
| Bet on *p* at a/U | U – a | U – a | - a | - a |
| Bet on *q* at b/U | U – b | - b | U – b | - b |
| Minus bet on *p & q*  At c/U | - U + c | + c | + c | + c |
| Total  Bet on *p OR q* | U - a – b + c  U - price | U – a – b + c  U – price | U – a – b + c  U – price | - a – b + c  - price |

We can see from the last line that a bet on *p* OR *q* at (a + b – c) / U is identical to (a bet on p)+ (a bet on q)– (a bet on p & q). So necessarily P(*p*)U + P(*q*)U – P(*p* &*q*)U = P(*p* OR *q*)U

This demonstrates that if a subject were to prefer a bet on *p* OR *q* at stake size U to a bet on *p* at stakes U plus a bet on *q* at stakes U minus a bet on *p* &*q* at stakes U, then the subject would have inconsistent preferences and would be vulnerable to a Dutch Book. The Dutch bookie could simply swap the same bet presented in different forms for a small fee.

### 3.2 Multiplying conditional and unconditional bets and the law of conjunction

A second type of bet is a conditional bet. This is a bet on *p given q.* A conditional bet is a bet on *p* that is only valid if *q.* Ramsey defined conditional degree of belief in *p given q* as follows:

If a subject is indifferent between:

1. B if *q*, C if ~*q*.

2. D if (*p & q*), E if (~*p & q*), C if ~*q*.

Then S’s degree of belief that p given q = (B – E)/(D – E).

We then have the value of a conditional bet defined in terms of the value of a non conditional bet. It follows from this that if the unconditional degree of belief in *q* varies with the stake size B – C, then the conditional belief *p* | *q* will vary proportionately. The stake size of the conditional bet is D – E. So the conditional belief will vary according to D – E as well.

To multiply unconditional and conditional bets, if the unconditional bet wins then the entire stakes B - C are placed on the conditional bet as the price, which is B – E. The multiplied stakes are thus formed by dividing the unconditional stakes by the conditional probability.

MULTIPLICATION RULE FOR STAKE SIZES. If the stake size on the unconditional bet is U, then the stake size on the conditional bet will be U/P(*p* | *q*).

We can now demonstrate the law of conjunction:

CONJUNCTION: P(*p & q*)U = P(*p*)P(*q* given *p*)U P(*q* given *p*)U

Suppose P(*p*)U = a/U and P(*q given p*)U/b/U = b/U

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Price/stakes | *p & q* | *p & ~q* | *~p & q* | *~p & ~q* |
| *p* at a / U | U – a | U – a | - a | - a |
| *q given p* at U / U2÷b | U2/b– U | - U |  |  |
| Total  Bet on *p*&*q* | U2/b – a  U2/b – price | -a  - price | - a  - price | -a  - price |

So P(*p & q*) at stakes U2/b = a/U2÷b = P(*p*)UP(*q given p*)U/b÷U

So an agent who had a degree of belief in *p* of a/U at stake size U and a degree of belief in *q given p* of b/U at stake size U2/b, but did not have a degree of belief in *p & q* of ab/U2 at stake size U2/b, would have inconsistent preferences and would be vulnerable to a Dutch Book. The Bookie could simply swap the bet on the conjunction for a bet on *p* multiplied by a bet on *q given p* and swap back under a different description for a fee.

This escalation of stake sizes for conjunctions is an important phenomenon which is intimately related to risk and frequency. As the stake size escalates exponentially through iterated conditional bets, the partition of possibilities increases in size and the probability of each possibility decreases proportionally.

### 3.3 The Kelly Criterion

To illustrate the importance of this we refer to a paper by Kelly JR (1956) in which he wished to give a general cost function meaning to Shannon’s theorem for transmission rate. Kelly showed that there is an optimal stake size at which one should bet at an advantage given that there is a periodical event on which one can gamble at any stakes you like at fixed odds. To illustrate with Kelly’s example, suppose that you had special information on a daily horse race where the market odds were 1 : 1. Suppose the special information told you which horse would win with a probability *e* such that *e* > ½. Naïve stake size invariant decision theory would recommend that you bet your entire bank roll W on the first day, and then bet your entire accumulated bank roll each day after that because this way you maximise your expectation. Your expectation would then equal *e*2W, with a daily profit of (*e* – ½) of your accumulated wealth. But it is obvious that you have quite a high probability of going bust after a few days using this strategy, since as soon as you lost a single race, you would lose your entire capital, and subsequently you would gain only (e – ½) of nothing. The probability of going bust after N days is equal to 1 – *e*N. If you played for long enough the probability of going bust would rise to probability 1. This means that, in view of this long run, the initial expectation of *e*2W is extremely over optimistic.

A better strategy is to bet only a fixed proportion of your bank roll L and allow your bankroll to grow logarithmically. This would ensure that you would never go bust. The question remains, given odds *o* and evidence *e* what is the optimal value for L? The assumptions in this model are that i) money is infinitely divisible ii) the game never terminates iii) *e* gives the correct probability distribution of wins and losses in frequency terms. For example, if *e = x* then the probability of winning N bets consecutively given that N bets are made is equal to xN.

According to Kelly, the optimum proportion of your bank roll “L” to pay as a price for a bet when betting at odds ½ when the probability of winning is *e* on the special information is given by:

L = 2*e* – 1.

This is a special case of a more general formula that can calculate the optimum stake size for any advantage at any odds. In gambler friendly terms, the general formula is: L = edge/odds.

The significance for a betting odds measure of degree of belief is that if we assume that the evidential probability is the optimum odds at which you should be indifferent, then Stake Size Invariance about evidential probability is false, since if your evidence consists of *e* and the optimum stake size at which you should bet varies according to the odds, then it is conversely true that the minimum odds at which you should bet varies according to stake size. Further more it means that the ideal degree of belief that *p* measured by betting odds is not always equal to the objective “probability” *e* even assuming that there is a fact that the objective probability that *p* = *e* and that you know the value for *e*. In fact, the evidential probability *p* and the objective probability *e* only converge when the stake size is 0.

Kelly’s criterion tells us that the ideal degree of belief to have relative to a bet varies with stake size, getting lower as the stakes get higher. For example, suppose *e* = 0.6. Then L = 0.2. Since the bet is at even odds, the stake size is 2L. This means that you should only accept bets on *p* at even odds whenever the price of the bet is less than 0.2W and the total stake is less than 0.4W. Given the betting measure of degree of belief this means your degree of belief is more than ½ for stakes less than 0.4W and less than ½ at stakes of more than 0.4W.

This is a *normative claim* based on a careful consideration of the relationship between the rate of growth and the law of large numbers. It is not a descriptive claim about the psychology of bettors. Kelly gamblers have been shown in simulations and in real life gambling situations to out perform classical gamblers. The history is that Kelly’s paper was almost completely ignored until a friend of Shannon’s named Ed Thorp used the Kelly criterion to compute optimal stakes for Black Jack and then the securities markets. It was Thorp’s success as a hedge fund manager that made people take notice of Kelly’s otherwise obscure paper. (Poundstone, 2005).

The Kelly criterion does not see each in a sequence of N bets on a repeatable, causally independent trial as truly independent. Although to naïve decision theory, each sequence of W wins and L losses is equiprobable, they are *not* equivalent in terms of the expectation on conditional bets. This is because the expectation is conditional on your state of wealth at the beginning of each trial, and this state of wealth is entirely dependent on the outcomes of the previous trials. So instead, the sequence should be seen as the averaging of two long conditional bets with a logarithmically increasing or decreasing stake size. As we saw in the Dutch book argument for the conjunction rule, for a conjunction of N conjuncts, the stake size increases by U/P(*p*)UN. Given MINIMUM, this allows for a considerably lower degree of belief in the conjunction than in the independent conjuncts. Given a strategy of two way conditional bets, so that the stake size remains at a constant proportion of your wealth, then the de-escalation of stakes conditional on losing bets will be U/P(~*p*)-U-N.

It is important to note that though there are superficial similarities between Kelly’s expectation calculations and the Von Neumann/Morgenstern utility theory and subsequent expected utility theories; Kelly explicitly denies that his calculation has anything to do with the utility theory of Von Neumann since preference based utility measures would “depend on things external to the system and not on the probabilities which describe the system, so that its average value could not be identified with the rate as defined by Shannon.” (1956 p. 918).

Finally, Kelly’s criterion does not apply to certainty. This is because odds are 0 : 1 at certainty, making edge/odds undefined since the denominator is zero. Furthermore, if the odds are 0 : 1, then the edge cannot be greater than zero, making the numerator of edge/odds equal to zero. This means that the value for L when the odds are certain and *e* = 1, is 0 / 0, which is undefined.

The Stake Size Variation Principle, on the other hand, allows certainty up to a maximum stake size, and less than certainty above this threshold. The argumentative use we are putting the Kelly criterion to is this: We allow the Stake Size Invariantist the maximum possible in terms of assumptions. We allow that evidence can come in the form of absolutely certain knowledge of a stake size invariant objective probability that is invariant over a hypothetically infinite run. We then show that, using Kelly’s criterion, even if we grant these assumptions to the Stake Size Invariantist, we can still demonstrate that the best degree of belief given evidence of this kind is sensitive to stake size. We are assuming that the best degree of belief is the one that maximises your long term growth. Since we are equating evidential probability with the probability that you ought to have given the evidence, then this argument shows that evidential probability is stake size sensitive.

The Stake Size Variation Principle differs significantly from Kelly’s formula and it might be worth explaining the differences as a way of making it clearer how the SSVP works. The major interesting difference is that Kelly’s formula assumes that evidence comes in the form of a single point value “e” which represents the probability of the propositional function in question from the perspective of the special information of the agent; whereas the SSVP has the complimentary values k*p* and k*~p* which are measures of the value of evidence for and against *p* respectively. Adding the two values together gives k*p* + k*~p* = KP which is the total value of knowledge available to the agent relevant to *p*. SSVP therefore gives a two dimensional representation of the evidence available to the agent. It gives the ratio k*p* : k*~p* from which we can form the stake neutral probability k*p* / KP; and it also gives the magnitude of knowledge represented by KP. This affects the resilience of the degree of belief in the face of stake size variation, as well as resilience in terms of the degree of change in degree of belief given new evidence.

Frequency information tends to have these two dimensions, which can most easily be thought of as Relative Frequency and Sample Size. For example, supposing you had a statistic such that a given team had won 80 matches and not won 20 matches. The two dimensions in this statistic are the relative frequency, (0.8) and the sample size (100). In terms of the SSVP we need to give a value to the evidence, which is to say how valuable is it to know the results of a particular match? To make the example easy, let us just stipulate that the value of knowledge of each match is £1. In this case K*p* = £80 and K*~p* = £20, so that KP = £100.The relative frequency would then be K*p*/Kp = 0.8 and the sample size would be KP = £100.

If one was a Kelly gambler and had this statistic in the private wire, then one may as well use the value 0.8 to obtain “e”, although the statistic doesn’t warrant certainty that e = 0.8 if e is thought of as the objective probability under a frequency interpretation. However, the sample size Kp= £100 is surely also relevant. Intuitively speaking, the larger Kp, the more confident that the objective probability is equal to 0.8 you should be, and the greater the stakes you should bet. However, Kelly’s formula is blind to this kind of information, and assumes that “e” always represents the true frequency without specifying how one discovers “e” or how to respond to degrees of certainty in the value for “e”. (I am told that real Kelly gamblers use a divider, a real number n which is a function of the gambler’s confidence in e, by which the gambler divides her wealth. This will result in lower levels for L given the same odds. Such a divider would make Kelly Gamblers behave more like SSVP gamblers.)

The SSVP on the other hand, incorporates the magnitude Kp into the equation. Because k*p* is added to C to form the numerator and KP is added to B + C to form the denominator; the greater KP relative to B + C, the less the stake size B – C will affect the probability.

For example, suppose C = £49 and B = £50. This would be the case if your total wealth was £49.50 and you were considering a bet at even odds priced at 50p that paid £1 if *p* | E at even odds. If k*p* = £80 and KP = £100 then the SSVP PK£100(*p* | E)£1 = (C + k*p*)/ (B + C + KP) = 0.648. But suppose we were to keep the stakes fixed and lower the magnitude KP to £10, keeping the ratio the same, so: C = £49, B = £50, K*p* = £8 and Kp = £10. In this case PK£10(*p* | E)£1 = 0.522. If on the other hand we raised KP by a factor of 10 so that k*p* = £800 and KP = £1000, still keeping the stakes the same, then PK£1000(*p* | E)£1 = 0.772. In general the evidential probability will get closer to 0.8 as the ratio K/B+C tends to infinity. Consequently, if K is infinite, as perhaps it would be in an abstract mathematical case, then P(*p*) = 0.8, but for any finite value of K, then P(*p*) < 0.8 at positive stakes. (This is only universally true when k*p*/K > ½ . For some stakes with a binary partition, then when k*p*/K < ½, the evidential probability can be higher than k*p*/K as there is a pull towards the logical probability ½).

We have used a purely quantitative example to make it clear how the value of knowledge functions. But a good feature of the SSVP is that it can deal with qualitative evidence as well. For example, if your special information was not in enumerative form, but did change the evidential probability of the result, then a value could still be given for k*p* and KP. Suppose you had heard that one of the players in the opposing team had been partying hard the night before and was likely to under perform. How much is the information worth? Of course such questions are always going to be a matter of judgement and no simple formula is going to give the answer. But at least the SSVP gives a unified system in which to express the answer. The value of this knowledge is the change in expected value that the knowledge brings to a bet from a state of ignorance. If a subject already has a great deal of evidence relevant to the match, this information may not have such a big effect on their degree of belief. For example, suppose, as above: k*p* = £80; k*~p* = £20; and KP = £100. You then find out that the star player of the opposing team is sick with a hang over. You rate this at £20 against. The new value of k*~p* will thus be £40 and the new value for KP = £120, whereas there is no new information for *p* so k*p* remains the same at £80. Because there is evidence both ways, the addition of £20 for won’t have the impact of increasing the expectation by £20, since it is partially cancelled out by the evidence already on the table. How to assign value to evidence is admittedly left in want of explanation, but at least the framework is there to derive a stake size variable probability from mixed evidence. If you didn’t know whether or not the player had partied hard the night before, but did judge that this information would swing your ignorant expectation by £20, then there is a clear sense in which this information is *worth* £20.

The other difference between Kelly’s formula and the SSVP is that Kelly’s formula pays no attention to the absolute magnitude of wealth, so that the recommendation to stake 20% of your wealth on *p* | E, is entirely a function of e and o and remains the same whether your wealth is piggy bank of coppers, or whether it is the entire GDP of the UK. The SSVP on the other hand measures wealth and knowledge in the same units, so that the absolute stake size is important as well as its value proportional to the wealth of the individual. Again, the SSVP is two dimensional, giving both a magnitude and a ratio value for stake size; whereas the Kelly formula is one dimensional in that it is blind to absolute differences in wealth.

The SSVP has two wealth values: C and B. C is the wealth if the bet on *p* | EB – C loses and B is the wealth if the bet on the *p* | EB - C wins. The SSVP does not include a value for the given odds or the wealth of the agent prior to the bet and so gives the same expectation for bets regardless of what price was paid or the agent’s state before entering into the bet. The subject’s starting wealth is relevant to decision making in that one should only take the bet if W < P(*p* | E)B – C (B – C).

For example, if Charlie, a poor homeless tramp, is offered heads or tails by an eccentric millionaire whereby the tramp will receive £1 000 000 if heads, and £2 000 000 if tails, then his expectation will be equal to (C + k*t*) /(B + C + Kt). We can assume that k*t* / Kt = ½, and for the purposes of the example, we will set Kt = £2 000 000. In this case the tramp’s expectation on this bet would be £1 000 000 + £1 000 000\*£2 000 000/£5 000 000 = £1 400 000. This expectation would be no different for little Lord Fauntelroy, who already had a fortune of £1 500 000 and bought a bet for £500 000 at even odds on a coin toss. However, the difference would be that, while Little Lord Fauntelroy would be a fool to take this bet, Charlie would be crazy to decline it. This is not a difference in evidential probability or expectation, it is merely a difference in opportunity.

What *is* important in calculating the evidential probability is the relative value of the wealth, the stake size, and the value of knowledge. If the value of wealth far exceeds the value of knowledge, then the evidence ratio k*p* / Kp has very little effect, and the stakes ratio C/(C + B) has a much greater influence; whereas, if the value of knowledge is much greater than the value of wealth, then the evidential probability will tend to be much closer to the evidence ratio.

We can see the SSVP in terms of these two attractor values: C/(C + B) and K*p*/KP. If we consider C/(C + B) in isolation we can see that if C and B are equal, then C/(C + B) is equal to ½. We can also see that B exceeds C by the stake size, so the proportion of the stake size relative to wealth is given by (B – C)/W. For a degree of belief relative to a stake size to be of practical interest, then W must fall between C and B. It is a consequence of the SSVP that the greater the stake size relative to C, the greater the bid ask spread, and the lower the evidential probability at positive stakes. So although the SSVP starts from an entirely different place, the effect on behaviour goes in the same direction. The greater (B – C)/C, the lower the evidential probability and therefore the lower the minimum odds you would accept. Therefore, for any fixed odds “o”, there will be a maximum stake size as a proportion of your wealth that you will accept bets at o. For example, in a case where you had zero evidence and o = 1/4, then the maximum stake size at which you would bet would be where C/(B + C) = ¼, which is whenever B = 3C. W = C + o(B – C) so W = C + 3C/4 – C/4 = 1 ½ C. This would make the maximum price of the bet ½ C, which is 1/3W. In other words, L = 1/3. However if we take a case where there is a value for k*p* and KP, the value for L will depend on the exact values for B, C, k*p* and KP. There will not be a general rule, but in any specific case the value for L can be determined.

It may be objected at this point that there are two distinct arguments for the SSVP here, the argument that an increase in stakes will broaden the range of possibilities you need to eliminate with your evidence, and the Kelly argument from the logarithmic nature of growth. The worry is that these arguments are unrelated and therefore do not support each other. But I hope to persuade the gentle reader that stake size escalation is at the heart of both arguments, they are both about the growth of possibilities. Given k*p* / KP is a best estimate of frequency, then the greater KP, the better the estimate. If k*p* / KP = 1, then our best estimate is that *p* | E will always be true. But if we expand the possibilities by increasing the stake size, then the estimate will not be so robust. The same applies if k*p* / KP is less than 1. In the probabilistic case, we can say, in the style of David Lewis in Elusive Knowledge (1996) “Our evidence predicts that average frequency will tend to k*p*/KP, psst! As long as we ignore the possibility that we stop playing after four consecutive heads but continue playing after four consecutive tails.” At low stakes we can ignore this possibility, but when L > ¼ we can’t ignore this possibility and so must adjust our estimate downwards.

3.3 Dutch books against knowledge when not certain.

Now suppose I claimed to know that *p* but that my degree of belief was less than 1. This would mean that I would be indifferent between:

1. A for certain

2. B if *p* and C if ~p

Such that A < B.

This should mean that I would be happy to exchange option 1 for option 2 since I am indifferent between them. But given that I know that *p*, then I know that option 2 is equivalent in value to B. Therefore I should be indifferent between A and B. But by hypothesis, my degree of belief that *p* is less than 1, so I must prefer B to A. So claiming to know something when my degree of belief is less than 1 leaves me open to a Dutch book and to the charge of inconsistency. It is equivalent to being indifferent between A and B but preferring A to B. The cunning bettor would simply swap A for B plus a small sum, and then swap B back for A at no charge. So the thesis that knowledge claims should only be made on a degree of belief 1 is supported by a Dutch book argument.

Given the minimum and maximum constraints, this means that certainty at stake size U entails certainty at any stake size lower than U, but only entails a degree of belief equal to at least U/V for any higher stake size V.

### 3.4 Bet settlement requires justification and evidence.

What is good about equating knowledge with bet settlement is that just like knowledge, bet settlement requires conclusive evidence. The evidence must satisfy not only the subject of the belief, the bettor; but also relevant third parties, namely the bookie. This has two important consequences. Firstly it highlights the social character of knowledge, since in order to have knowledge, you must not merely have some internal reasons for your conviction, but you must also be able, at least in principle, to justify it to a third party. Secondly it allows for contextual and interest relative variation in knowledge ascriptions and probability assignments. What suffices for bet settlement in one context may not suffice for another, and big bets will tend to require more evidence than small bets. We can then simply give the maximum stake size at which evidence justifies certainty and settles bets as the *value of evidence*.

### 3.5 Bet settlement is intersubjective

We can call evidence that is bet settling “conclusive evidence”. What is meant by “conclusive evidence” in this context is clearly explicated by P. F.Strawson:

“It is important to notice that where we may speak of conclusive or overwhelming evidence for *q*, of *p*1 – *p*n making it certain that *q*, there we may also speak of *proving, establishing,* or *putting it beyond all doubt* that *q*, although, since the arguments are inductive, we do not produce as grounds anything that entails *q*. For it is not a question of support falling generally short of entailment; of entailment being the perfection of support. They are not related as winner to the runner-up and the rest in the same race. The perfection of support is proof, but not deductive proof; it is conclusive evidence.”

(Strawson, 1971, pp. 237-238)

In order for two agents to enter into a bet on a proposition, they must tacitly agree on what counts as conclusive evidence of the proposition bet upon, and what counts as conclusive evidence of its negation. There would be no point in agent R entering into a bet on *p* with agent Q unless R was sure that Q would pay up when it was evident that *p* was the case.[[9]](#footnote-9) Although when entering the bet their degrees of belief can be different and therefore entirely subjective, yet when they settle the bet their degrees of belief must converge and both rise to 1. This means they tacitly share a set of dispositions which govern the kinds of evidence they will accept as bet settling. These dispositions will be intersubjective and will be constitutive of the shared meaning of their terms. We can call the state of knowing what constitutes conclusive evidence for a proposition: “understanding” the proposition. This is akin to grasping the truth conditions. We can also name the curiosity born out of such understanding “wondering if”. If one wonders if *p* then one must have at least some grasp of what constitutes conclusive evidence that *p*.

Of course it is quite possible that two agents could bet on some statement and yet one of them come to know it given evidence E whereas the other doesn’t consider E to be conclusive at all. However this would constitute a serious conceptual disagreement and, if both parties were honest, the stakes should be returned. If a statement “S” seemed obviously true to R on evidence E, but did not seem true at all to Q on the same evidence E, it is clear that R and Q do not share an understanding of “S”. This means that there is no proposition expressed by “S” that they were betting upon and so the bet was invalid. A bet is settled when both parties agree that a proposition is true. But as soon as the bet is entered into, both parties must already tacitly agree about what kind of evidence will justify claiming to know that *p* is true. This agreement will be in the intersubjective dispositions that make communication possible.

A further complication is that in a competitive bet, the stakes relative to the bettors will not generally be the same, with one bettor having negative stakes while the other has positive stakes. This might lead to a good deal of disagreement about whether a certain evidence proposition is conclusive evidence for a proposition, with the disagreements increasing in likelihood as the stakes get higher. This shouldn’t trouble us, since such disagreements are commonplace when stakes are high.

The SSVP ascribes a lower value for conclusive evidence to the person who loses a bet, than to the person who wins. The self interest of the parties will work in the opposite direction. It is the person who loses a bet who needs to be persuaded by the conclusive evidence, since it is the person who loses the bet who must part with their money. There is an interesting problem here, but in the end I think that it does not constitute an objection to the SSVP. In fact there will be evolutionary advantages to a group where those with the most to gain by believing *p* require the greatest level of evidence for *p* before becoming certain, while those who lose out if *p*, will require less evidence. Such a group are more likely to co-operate. According to the SSVP it is the person who must pay up who becomes convinced first. The winner of this bet is unlikely to object, and can accept the money without being fully convinced. If we imagine an inverted SSVP agent, then she may be able to make short term betting gains by refusing to pay up at negative stakes at levels of evidence that she accepts as conclusive at positive stakes. But in the long run she would be run out of town. A whole community of inverted SSVP players would fall a part, with no one ever honouring their agreements unless it was to their advantage to do so, and one party always feeling aggrieved at any settlement.

For the greater part of our beliefs, we as a species are on the same side, so we can expect wide inter subjective agreement. In a community of SSVP agents, we would expect a greater level of certainty among people who are not particularly involved, where the decision makers will tend to be more epistemically reserved. I think this picture maps on to reality pretty well. Both these points will make fruitful areas for further research, since there are some testable predictions, at least in principle, here.

### 3.6 Settling bets when the proposition is false

It is of course true that two or more people could settle a bet on a false proposition. The mere act of settling the bet would not transform their beliefs into knowledge. Convergent conviction is consistent with error. Bet settlement is a measure of a degree of belief 1 in all parties to the bet, not a measure of invariant truth. If we assume that the betters are honest and rational then we can assume that the probability conditional on their evidence is also 1. When honest and rational people are convinced on their evidence of the truth of a false proposition, then they will accord an evidential probability 1 to that proposition and take themselves to know that the proposition is true. I will quote Levi:

“Propositions accorded probability of one are liable to be false…The ramifications of this approach do admittedly stand in need of further examination. But the position is frankly fallibilistic. Empirical propositions can justifiably be believed and, indeed, admitted into evidence even though it is possible that they are false.” (Levi, 1967 p.209)

However, a bet that is universally considered settled on the evidence, can, on the appearance of unexpected new evidence, be re-opened. Such epistemic shocks only show that it is possible to consider evidence conclusive when it is in fact not. They do not show that bets are ever genuinely settled on false propositions, they only show that bets sometimes *appear* to be settled when they are *in fact* not settled. A general agreement that a proposition should be accorded a probability of 1 on the evidence is not sufficient for evidential certainty. This observation is no more problematic than the observation that a general agreement that a proposition is known is not sufficient for it to be knowledge.

A problem for reliabilist theories of knowledge is that whether a belief is reliably formed depends on the background environment in which it is formed. In Kornblith’s words: “There is nothing in the world, Brandom argues, that serves to specify the reference class against which reliability is to be assessed.” (Kornblith, 2002 p.64) Reliability here can be directly couched in terms of frequency interpretations of probability. A belief is reliably true when beliefs formed “in those conditions” are always true. The problem is to specify “those conditions”. In other words, the problem is in specifying the reference class. Naturalists like Kornblith (1999, 2002) and Millikan (1989, 1993) will fix the reference class in terms of the adaptive environment of the organism. The proper function of beliefs is to be true, a naturalist could claim, and in the natural cognitive environment of the believer, their beliefs will tend to be true. My approach is broadly this: the reference class is set by the presuppositions and dispositions common to the relevant language community. If the environment is freakishly different from that in which the presuppositions were formed, then beliefs formed on the basis of those presuppositions will lose their reliability. Bettors unaware of this loss of reliability will tend to continue to settle bets in these adverse conditions. However, if the change of environment is made explicit, then the evidence will no longer be considered reliable enough for bet settlement.

### 3.7 Settling bets with doubts, and knowing without settling bets.

***Settling bets with doubts***

One line of objection to this thesis is that in some actual situations people can be obliged to settle bets whilst still harbouring doubts. Suppose Beth bet with Alf that Cherico would win the Grand National. Cherico had very bad form and had never won a race before, nor had ever completed a race in anything approaching the time necessary to win the Grand National. Surprisingly, Cherico wins the Grand National. Alf must pay off his bet to Beth. But Alf is suspicious. The grounds of his suspicion are simply that it was extremely unlikely that Cherico would win, unless there was some kind of cheating. But Alf recognizes that unless he can come up with some proof that Cherico was cheating, he must pay up. Given this situation, I am obliged to say that Alf *knows* that Cherico won the Grand national, but this seems counterintuitive since Alf doesn’t even *believe* that Cherico won the Grand National. A similar case would be where a Jury finds a defendant guilty because they accept that the evidence against him is conclusive, but yet one juryman may still wonder privately whether the defendant really was guilty. It might seem that this juryman should in all honesty settle bets on the guilt of the defendant, whilst it again would seem incorrect to say that he knew that the defendant was guilty. A third case would be a scientist who must accept that his theory has been falsified by a particular experiment, whilst still privately believing that his theory is true. Such a scientist should not be described as knowing that his theory is false even though he really ought to settle any bets on his theory given the experimental results.

These are difficult cases and demand a complex treatment. The approach I shall take is to challenge the subject’s denial of knowledge in these cases, and deny the acceptability of their doubt. To take Alf’s case, if Alf’s doubt was sincere, then he could demand further evidence before he paid up. If, for example, he thought that the horse that won the Grand National was not Cherico, he could ask that a blood test be conducted. Beth may or may not accept this request as reasonable given the value of the bet. The very fact that Alf is prepared to pay up is evidence that he does not really consider his doubts to be reasonable. It is possible that someone who thinks that they have doubts, does not really have doubts. If the scientist really had doubts about the experiment, he could repeat the experiment. If the Juryman really had doubts about the defendant’s guilt then he would have delivered a not guilty verdict.

Another approach is to say that these cases are cases where the public acceptance is of a slightly different proposition from the private doubt. Alf did not doubt that Cherico won. He doubted that Cherico won *fair and square*. But the bet was not on this more complex proposition. The juryman was not settled in his mind that the defendant was guilty; but he was settled that it was beyond reasonable doubt given the evidence before the court that he was guilty. Again these two propositions aren’t necessarily co-extensive. They *ought* to be; but this is a problem for the justice system, not for the bet settlement theory of knowledge. The scientist was certain of the results of the scientific experiment. What he had doubts about was whether the scientific experiment was conducted properly; again a different proposition.

***Not settling bets through anxiety***

The other kind of case is where, through anxiety, or epistemic modesty, bets aren’t settled on propositions that are in fact known. For example, a school boy in an examination is uncertain that the answer to the question: “what was the year of the Battle of Hastings?” is “1066”. If he were asked to bet on it, he may bet at quite low odds, but he would not settle a bet on it, since he is so anxious. According to this thesis, the school boy does not know the answer. However I recognise that many philosophers have used this kind of case as an argument for fallibilism on the grounds that intuitively the school boy does know the answer (given that the answer is correct and he was justified in believing it). I think that the intuitions are based on the fact that the case is under described, allowing the reader to imagine both cases where the school boy has genuine doubts, in which case he doesn’t know, and cases where the school boy in fact does not have any (reasonable) doubts, and therefore really does know. This may trick the reader into thinking there can be cases where someone knows something whilst retaining reasonable doubts and therefore not being certain.

A simple test case would be where there was a penalty for writing down the wrong answer, but no reward for writing down the correct answer. In this case, only someone who knew the right answer would write it down. A nervous school boy would prefer the option of writing nothing down. One might wonder why anyone would write down any answers in such a test since it is safer to write down nothing. But it is only safer to write nothing if you do not know the answer. If you do know the answer it is equally safe to write down the answer. We can see that by this test, the sheer size of the penalty is likely to make a difference. If the penalty for writing down a wrong answer was death, then few people would “know” the answer, whereas if the penalty was just a point in an arbitrary score, then many more people would take the risk. If the same school boy prefers the option of writing nothing, then to my mind it is clear that he doesn’t know the answer at those stakes. This is not to say that he doesn’t have a good idea what the answer is. It might be that in a lower stakes situation he would be certain. He clearly has the correct opinion that it is 1066; but correct opinions are not necessarily knowledge, and this philosophical result is over 2000 years old, so there is no need to labour it here.

### 3.8 Jeffrey and the theorist’s terms.

Measuring degree of belief in terms of preferences and betting odds gives a value for degree of belief that, in Jeffrey’s words, is in the “theorist’s terms” (Jeffrey 1970). The focus of this thesis is not on partial belief, but on knowledge. “Certainty” is defined as a degree of belief 1 measured in terms of betting odds. This then is the theorist’s term. It is this theoretical notion of certainty that is necessary for knowledge, not the content of any sincere avowal of certainty.

Also, since Ramsey’s formula is couched in terms of preferences over options, the same measure can apply to non linguistic animals and people who cannot, or will not, participate in bets. Literal betting, and bet settlement, should only be thought of as ways of measuring the difference between thinking and knowing, it would be absurd to suggest that they are constitutive of these mental states. But the underlying effect of evidence and conviction on decisions over choice *is* constitutive of degree of belief, and in this sense, any intentional choice of action can be modelled by bets specified in terms of utility.

All that is being claimed here is that if a subject is indifferent between A and B, but prefers A in any case to B if *p* is true and C if *p* is false, then that subject is less than certain that *p* and therefore does not know that *p*. How it is established that a subject has these preferences is a practical problem which is irrelevant to the definition of the theoretical terms. Between honest speakers of a common language, then knowledge that *p* is measurable in terms of the conditions under which bets would be settled.

## 4. Logical closure and the certainty condition

One strong argument for the certainty condition is that it seems to be a normative requirement that knowledge claims are closed under logic. It is always irrational to have inconsistent knowledge claims. But without the certainty condition, it is possible to rationally have inconsistent beliefs. The chief example of this is the Lottery Paradox (Kyburg, 1961). If the probabilistic threshold for rational belief is less than one, then it is rational to believe that every ticket of a lottery will lose and to at the same time believe that one ticket will win. This clearly offends against logical closure.

### 4.1 Challenges to logical closure

Logical closure does not go unchallenged though. Nozick’s (2001) highly plausible truth tracking theory of knowledge led him to challenge logical closure. The central cases are when a subject seems intuitively to know that p, but does not seem to know that ~q where q entails ~p, and therefore p entails ~q. Nozick is principally concerned with sceptical arguments such as the brain in the vat argument. Take a proposition p that we want to say that S knows. For example, S knows that S has two hands. In Nozick’s truth tracking theory this is because, in nearby worlds where S does not have two hands, S does not believe that S has two hands. But S does not know that it is not true that S is a brain in a vat. This, on Nozick’s account, is because in nearby worlds where S is a brain in a vat, S still believes that S is not a brain in a vat. This solution to scepticism comes at the heavy price of abandoning logical closure. This is because Nozick’s account permits S to claim to know that S has two hands, to claim to know that S having two hands logically entails that S is not a brain in a vat, but to withhold a claim to know that S is not a brain in a vat.

Dretske’s (1970, 1981), account of epistemic operators also contains a challenge to closure. Dretske claims that epistemic operators such as “knows that” and “have reason to believe that” do not penetrate to all the entailments of the object proposition. More specifically, they do not penetrate to shared presuppositions associated with the knowledge statement. For example, if S knows that the animal in the enclosure is a zebra, and knows that if the animal in the enclosure is a zebra then it is not a painted mule, then if “knows that” were fully penetrating, S should know that the zebra in the enclosure is not a painted mule. But we can recognise that, for all S knows, the zebra could be a painted mule. So knowledge is not a fully penetrating epistemic operator.

### 4.2 Dutch Book argument for logical closure.

Logical closure for bet settlement can be argued for using coherence arguments since if one settled bets on *p* given evidence E, and settled bets on *q* given evidence *p*, but did not settle bets on *q* given evidence E, then one would have inconsistent preferences and would be exposed to Dutch books. The bookie could simply exchange bets on *q* for bets on *p* and then exchange them back plus a small fee. This assumes that the bets are all at the same stakes. However, as we saw by the conjunction rule, depending on the odds on *p* and on *q* prior to E, the stakes on the conjunction *q* & *p* could be a lot higher than the stakes on *p*. This would allow for someone to consider E sufficient to settle bets on *p* but not to settle bets on *p* and *q* without being vulnerable to a Dutch book.

According to the Stake Size Variation Principle, coupled with the stake escalation of conjunctions, then it could be possible to know that *p* and know that *q*, at low stakes but to merely think that *p* & *q* at the elevated stake size attributed to the conjunction.

So according to the Stake Size Variation Principle, logical closure is limited to propositions at the same stake size.

Furthermore, logical closure does not demand that one should assent to everything that follows logically from what one claims to know. Dutch book arguments only demand that one should settle bets on all propositions that are demonstrated to follow logically from what you know, and withdraw knowledge claims from anything that is demonstrated to be inconsistent. In the zebra case, the intuition is supposed to be that ordinarily S will claim to know that the animal in the cage is a zebra, but not claim to know that it is not a cleverly painted mule. In bet settlement terms this means that S will settle bets on “X is a zebra”, but will not settle bets on “X is not a painted mule”. To get S into a Dutch book, one would have to get S to simultaneously settle bets on X being a zebra, whilst betting at some odds on X being a painted mule. However, if this second proposition was open, i.e. its negation wasn’t presupposed, then S would withdraw settlement on X being a zebra. This is what intuitively happens. We would expect a person to claim to know that the animal was a zebra, but when asked whether she knew that it was not a painted mule, to not only withhold assent from “X is not a painted mule,” but to withdraw the previous knowledge claim that “X is a zebra”.

## 5. Conclusion

Since ancient times philosophers have distinguished between thought and knowledge. A thought is worth a part of the truth, whereas knowledge is worth the whole of the truth. Following Ramsey, we can measure degree of belief and the evidence that justifies it by taking the value of a bet as a proportion of the stakes. This proportion is the probability. Knowledge comes when the evidence and degree of belief are sufficient to settle bets. The probability of knowledge then is 1, since the value of a bet you know you will win is equal to the total value of the stakes. We introduced the concept of the *value of evidence*, which allowed us to explain how certainty can be stake size relative, and therefore how evidential certainty can be a necessary condition for knowledge whilst retaining the thesis that inductive knowledge is fallible. A proposition is certain when the evidence in favour of the proposition exceeds the stake size on the proposition, provided that there is no evidence against the proposition.

This gives us the thesis that certainty and conclusive evidence are jointly necessary for knowledge.

# Chapter 3.

# STAKE SIZE VARIANCE AND THE PROBLEM OF RISK

## Introduction

The aim of this chapter is to resolve the long standing tension between the certainty condition for knowledge, and the fallibility of inductive inference. The problem can be stated in this way: much of our knowledge is held on inductive grounds, beliefs held on inductive grounds are fallible, so certainty cannot be necessary for knowledge. The same point can be made about belief. Mark Kaplan (1996, p. 93) argues that a certainty condition for belief is too demanding on the grounds that this would require betting astronomical sums for no gain, not only on the proposition believed, but any logical entailment of the proposition. “Given this much, it is hard to see how there can be very much you have any business believing.” He concludes that “the certainty view must be mistaken.” The same thought is presented by many different philosophers and can be traced back to Locke.

The argument can be put like this:

1. There is little, if anything, we are warranted in being certain of.

2. We know many things.

3. Therefore warranted certainty is not a necessary condition for knowledge.

If we calibrate probabilistic measures of evidence and conviction by setting certainty as probability one, then this argument implies a probabilistic threshold for knowledge, or warranted belief. Certainty, according to the Lockean threshold view, requires probability 1, but knowledge requires some probabilistic threshold less than 1. When I presented a paper at St Andrews, my commentator, Carrie Jenkins, challenged my claim that certainty was necessary for knowledge with the comment: “We are all fallibilists now.” The idea is that if knowledge is fallible, then it is not certain, and if it is not certain, then the evidential probability threshold must be less than 1.

Lewis in *Elusive Knowledge* claimed that “S knows that *p* if *p* holds in every possibility left uneliminated by S’s evidence.” (Lewis, D. 1996)

“It seems as if knowledge must be by definition infallible. If you claim that S knows that *p,* and yet you grant that S cannot eliminate a certain possibility in which not-*p,* it certainly seems as if you have granted that S does not after all know that *p.* To speak of fallible knowledge, of knowledge despite uneliminated possibilities of error, just *sounds* contradictory.” (Lewis, D. 1996).

He then noticed that the scope of “every” is context dependent. In an everyday context, sceptical hypotheses do not fall under the scope of “every uneliminated possibility”, but in the context of a sceptical argument, the sceptical hypotheses become salient and therefore must be eliminated. This has the consequence that it is possible to know that *p* given evidence E in one context, and not know that *p* given E in another context. But what is left unclear is what it is that makes an alternative relevant.

In this chapter we will be pursuing the natural idea that it is the *practical interests* of the agent that determines whether an alternative is relevant. An agent’s belief in *p* has practical consequences in terms of the range of actions that the agent will perform/refrain from. This range of actions is determined by the practical interests of the agent. An alternative possible world in which *~p* is only relevant if, in that possible world, it makes a practical difference to the desirability of an action whether or not *p*. The difference the truth of *p* makes to the desirability of an action we will call the *stake size.* The relevance of *p* to an action is determined by the difference the belief in *p*makes to the desirability of acting versus not acting. This will depend on the stake size on *p* of acting and of not acting. The stake size will vary over epistemic possibilities. In some possibilities it will no longer be relevant whether or not *p* since the stake size on *p* is zero. When the stakes are high, we can suppose that the stakes will remain high over a greater range of possible alternatives, whereas if the stakes are low, then the range of alternatives where it is relevant that *p* will be narrower.

This thought leads to Interest Relative Invariantism, (IRI). Many philosophers view the important difference between contextualism and IRI as being whether it is the context of the attributer or the context of the subject that determines the truth of knowledge attributions. However, Jason Stanley linked the truth of knowledge attributions to what was at stake for the subject. What is at stake for the subject is not context dependent, he claimed, so his Interest Relative Invariantism is not contextualism about knowledge. The basic IRI thought is that it is possible to know that *p* given E in one situation, where the stakes are low, but in another situation where the stakes are higher, to not know that *p* given the same evidence E. Once what is at stake for the subject is determined, then “S knows *p*”will have the same truth value in *all* contexts.

Stanley compares two possible types of IRI, *probabilistic strength of evidence* IRI and *relevant alternatives* IRI.

*“Probabilistic strength of evidence* IRI, practical facts about a subject’s environment at time *t* might make it the case that that subject must have stronger evidence than usual in order to know a proposition *p* at that time than she must possess in order to know that proposition at other times, where the strength of evidence is measured in probabilistic terms.

[ … ] *Relevant Alternatives* IRI, practical facts about a subject’s environment at time *t* might make it the case that, in order to know a certain proposition *p* at that time, that subject must rule out a different set of alternative propositions than she has to rule out at another time *t’*.” (Stanley 2006. pp. 85-86)

Now it should be clear that probabilistic strength of evidence IRI is straightforwardly incompatible with the thesis that an evidential probability of 1 is a necessary condition for knowledge. This is because probabilistic strength of evidence IRI has the consequence that it is sometimes true that S knows *p* but that S possesses evidence for *p* with an evidential probability of less than 1. Therefore an evidential probability 1 is not a necessary condition for knowledge.

Relevant alternatives IRI on the other hand is compatible with the thesis that an evidential probability of 1 is a necessary condition for knowledge. The scope of *every* is conditioned by the practical interests of the subject. So long as *p* is true in *every* uneliminated possibility, then the evidential probability that *p* is 1, and S knows that *p.* However, if ~*p* is only eliminated in some of the possibilities, then the evidential probability that *p* will be less than 1 and S does not know that *p*. If what counts as an uneliminated possibility depends on the practical interests of the subject, it is possible that the same proposition *p* could have both evidential probability less than 1 when the stakes are high and evidential probability 1 when the stakes are low given the same evidence. As more possibilities become relevant, then the evidential probability measured as a proportion of uneliminated possibilities in which *p* is true can go down. A more overtly probabilistic way of saying the same thing is that probability ranges over a mutually exclusive partition of events such that the probability of the entire partition sums to 1. In Dorothy Edgington’s words:

“The fundamental principle governing degrees of belief, I shall argue, is this: take a set of exclusive and exhaustive possibilities – a set of propositions such that it is impossible that more than one of them is true, and necessary that at least one of them is true. Call such a set a *partition.* A person’s degree of belief in the partition must sum to 1, the value assigned to certainty.” Edgington, D (1996) p. 40

The scope of relevant possibilities thus defines the partition. As more possibilities become relevant, the partition as a whole becomes larger, and consequently the degree of belief in *p* can go down from 1 to less than 1. This does not only affect probability 1. If the “probability that *p*” is the proportion of the partition in which *p* is true, then as more possibilities are added to the partition, this proportion can change.

The aim of this chapter is to show that it is perfectly reasonable to suggest that evidential probability varies with practical interests. We will introduce the Stake Size Variation Principle which states that the larger the positive stake size on *p* the lower the evidential probability. We will argue that this variation itself varies in accordance with the weight of knowledge that supports the probability assessment. We will show that this Stake Size Variation Principle fits the actual decisions people tend to make in risky situations, especially in situations where orthodox decision theory fails to explain the data. We will also argue from considerations of deviation and reference class selection that the principle is a normative principle.

In a nut shell the idea is that evidential probability is calibrated to the evidential requirement for knowledge, rather than to infallibility. If none of our beliefs are infallible, then there is little use in calibrating probability against infallibility. It would be a little like calibrating our measures of velocity as a proportion of instantaneous travel. Since instantaneous travel is impossible, it is better to calibrate velocity to the speed of light. In the same way, given that knowledge is possible, it is much better to calibrate our evidential probability measure against knowledge rather than infallibility. So the slogan “Knowledge requires evidential probability 1” is more about evidential probability than it is about knowledge.

The conclusion of the chapter will be that it is true that certainty is a necessary condition for knowledge. The objection that the evidential standards for knowledge can vary according to what is at stake is met by applying this same principle to evidential probability itself. The evidence necessary for knowledge is equivalent to the evidence necessary to silence all doubt. But the evidence necessary to silence all doubt is greater when there is more at stake. The evidence necessary to silence all doubt warrants a credence of 1, since such evidence will suffice to settle a bet, and thus make bets at any odds advantageous. It is evidential probability itself that varies according to stake size and this not only explains the pattern of knowledge attributions, but also explains people’s preferences under risk, that have been problematic for decision theory.

## 1 The problem of stake size variation for the certainty condition

### 1.1 Rational degree of belief 1 necessary for knowledge

The principle claim of this dissertation is that a necessary and sufficient condition for knowledge is a rational degree of belief of 1. Rational degree of belief is understood to be a probability measure. Given the standard betting interpretation of probability, a rational degree of belief 1 is the degree of belief had in a proposition when the bet pays off, in other words when the bet is settled. The evidence necessary for rational degree of belief 1 then is the evidence that would suffice to settle a bet. Once a bet on a proposition has been settled, then it would be irrational to place another bet on the negation of the proposition whatever the odds. So, according to the betting interpretation, the rational degree of belief in the negation of a proposition that is known to be true is zero, and the rational degree of belief in the known proposition is 1. We can call this claim the “certainty condition”.

### 1.2 But evidential standards for knowledge vary with stake size.

One big problem with the certainty condition for knowledge is the now commonly held view that the truth of knowledge ascriptions can depend on non evidential features of the context. For example, according to Jason Stanley, (2006) it is possible for a subject with the same evidence that *p* to know that *p* at low stakes, but to fail to know that *p* at high stakes. If we also suppose that evidential probability is uniquely determined by the evidence, then it follows from the stake size dependence of knowledge ascriptions that the certainty condition is false, because the evidential probability must be invariant across high stakes and low stakes, and therefore, according to the certainty condition, the truth of knowledge ascriptions should likewise be invariant across stake sizes.

The principle objective of this chapter is to defend the certainty condition for knowledge against the objection that the evidential standards for knowledge vary according to what is at stake. I will name the claim that evidential standards for knowledge vary with stake the “stake size sensitivity principle”. The stake size sensitivity principle says that the evidential standards necessary for knowledge can sometimes get higher as the stakes get higher. The stake size sensitivity principle I take to be about knowledge the natural kind, rather than “knowledge” the English word. Knowledge plays a central role in guiding intentional action. The stake size sensitivity principle is the claim that the evidential threshold for knowledge is normatively higher when the importance of the actions it guides is greater. We will argue that the stake size sensitivity principle follows from a normative principle governing evidential probability that we will call the “Stake Size Variation Principle”, and is thus not incompatible with the certainty principle.

### 1.3 Resolution of conflict: degree of belief 1 varies according to stake.

The two principles are on the face of it in conflict.

1. The certainty principle states that the rational degree of belief necessary for knowledge is invariant: it can be no lower nor higher than 1.

2. The stake size sensitivity principle states that the evidential sufficiency condition for knowledge varies according to what is at stake.

The conclusion of an argument formed from these two premises is that:

3. The evidence sufficient for a rational degree of belief of 1 varies according to what is at stake in the situation.

In this chapter I will be taking it as given that the two premises are true. I will be exploring a possible way of accepting the conclusion with the minimal damage to orthodox probability mathematics. In fact, accepting the conclusion actually provides solutions to well known paradoxes for decision theory, so far from damaging orthodox probability mathematics, accepting the Stake Size Variation Principle exonerates it.

### 1.4 General principle: evidential probability varies according to stake size.

The general principle of stake size variation is that the rational degree of belief to have in a proposition *p* given a certain body of evidence E is relative to the stake size U. We can call this the *evidential probability p* | EU. Evidential probability can be characterised as the degree of belief one *ought to* have given the evidence. This is distinct from the *objective* probability, and also from the relative frequency. It is often the case that the relative frequency or the objective probability is not determined given a particular set of evidence, in which case, the evidential probability will not be identical with either of these quantities.

Timothy Williamson (2000) devotes a chapter to the discussion of evidential probability. He gives good arguments that the evidential probability of a proposition *p* given evidence cannot be defined as the degree of belief (which he calls credence) that an ideally rational agent would have in *p* | E. We accept that the degree of belief that one ought to have in *p* | E is not the credence that an ideally rational agent would have in *p* | E. There is no such thing as an ideally rational agent in our system. The degree of belief you and I ought to have in *p* | E is not necessarily the degree of belief that an ideally rational agent would have, since an ideally rational agent may have a rather different set of interests from you or I.

Williamson argues that evidential probability is not the degree of belief that anyone or any set of people actually has. He imagines a case where everyone is irrationally certain of a proposition which is in fact not certain. Given our simple definition, we can accommodate this case. The case is simply one where everyone is certain that *p* when they really ought not to be given the evidence.

Williamson furthermore argues against any decision theoretic definition of evidential probability. His argument relies on the assumption that there must be a unique value for *p* | E that is invulnerable to a Dutch book for any decision theoretic account of *p* | E. It is not clear however whether Williamson himself is making this uniqueness claim. Maurice Schulz in the 2009 Philosophy of Probability conference at the LSE interpreted Williamson as being committed to a uniqueness principle. The main thrust of this chapter is to deny that there is a unique value for *p* | E, since *p* | E varies according to what is at stake. Some probability distributions P may be vulnerable to a Dutch Book *only if* degrees of belief are not sensitive to changes in stake size. So Williamson’s arguments against a decision theoretic analysis of evidential probability on the grounds that evidential probabilities may be vulnerable to Dutch Books are also arguments that evidential probabilities understood decision theoretically should be sensitive to stake size. An agent who had no evidence for anything could have a degree of belief 0 in all propositions at positive stakes (when buying bets) and a degree of belief 1 (when selling bets) in all propositions at negative stakes, and would be completely invulnerable to any Dutch book, since he would neither buy nor sell any bets. However, such an agent could considerably improve his situation in life by gaining more evidence, so his strategy is subject to epistemic criticism of a different kind from probabilistic incoherence.

In broader terms, a decision theoretic approach to evidential probability is merely the claim that *p* | EU is the best degree of belief to have in *p* given E at stake size U, in other words the degree of belief that will lead to the most gains and the least losses as far as you can tell from the evidence. This is a matter of judgement. A secular humanist approach is to replace the ideally rational agent with a market place of wise people. The evidential probability of *p* | EU is then the fair price for a bet on *p* at stakes U. The problem of global error is then easily dealt with. If everyone has, in your opinion, the wrong value for *p* | EU, then don’t complain. Do as Thales did: put your money where your mouth is and make a killing.

Williamson gives further arguments against subjective Bayesianism that supports the Stake Size Variation Principle. He points out that according to Bayes’ theorem, P(e | e) = 1. This entails that the evidence on which one updates has probability 1. Given his equation K = E, this has the consequence that known propositions have a probability 1. He then distinguishes this from absolutely certainty as follows:

“For subjective Bayesians, probability 1 is the highest possible degree of belief, which is presumably absolute certainty. If one’s credence in *p* is 1, one should be willing to accept a bet on which one gains a penny if *p* is true and is tortured horribly to death if *p* is false. Few propositions pass this test. Surely complex logical truths do not, even though the probability axioms assign them probability 1. But since evidential probabilities are not actual or counterfactual credences, why should evidential probability 1 entail absolute certainty?” (Williamson, 2000, p.213-214).

Now it should be observed that rhetorically, Williamson does not simply state very big odds, but also gives a very large stake size. The intuitive force of his argument would evaporate entirely if he stated the bet thus: “One should be willing to accept a bet on which one gains nothing if *p* is true but forfeits a single penny if *p* is false.” Many propositions pass *this* test. Yet the odds on this bet are 1 : 0, whereas on Williamson’s bet they are death by torture : penny. Let’s assume this is odds 1010 : 1. If an ideally rational agent is committed to having credences constrained by the axioms of probability that were stake size invariant, then it would be *irrational* to accept the 1 : 0 bet but decline the 1010 : 1 bet. This would be because, relative to the large bet the subject’s credence is less than 1/(1010 +1), whereas relative to the small bet it is higher than 1/(1010 + 1). Williamson’s arguments therefore are not against linking evidential probability with credences, but against any interpretation of Bayesianism that insists on stake size invariance.

Williamson qualifies certainty with “absolute” certainty. Evidential probability 1 is not absolute certainty he claims. We are in complete agreement. All that follows from this is that certainty is not absolute certainty. The non equivalence of certainty with absolute certainty shouldn’t surprise anyone. Certainty can be construed as evidential probability 1 relative to the stake size. Absolute certainty then has a clear meaning: evidential probability 1 relative to all stake sizes. We distance ourselves from subjective Bayesians as Williamson understands them to be by denying the claim that “probability 1 is the highest possible degree of belief.” We deny this claim for all the reasons that Williamson gives. With Williamson we claim instead that outright belief also comes in degrees.

“Outright belief still comes in degrees, for one may be willing to use *p* as a premise in practical reasoning only when the stakes are sufficiently low. Nevertheless, one’s degree of outright belief in *p* is not in general to be equated with one’s subjective probability for *p*.” (2000, p.99). We suggest that these degrees of outright belief are measured by the maximum stake size at which that belief is certain. It thus follows that degree of belief 1 in any specific context is not necessarily the highest possible degree of belief, since there could be a context in which the stake size is higher in which the degree of that belief falls below 1. For every stake size there is a higher stake size, for every level of certainty, there is a higher level of certainty, and there is no absolute certainty, except as a kind of mathematical fiction, like infinity.

## 2. The Stake Size Variation Principle

In this section we will introduce the Stake Size Variation Principle which we give here as:

When K*~p* ≠ 0:

PK (*p* | E)B-C = (C + K*p*)/ (B + C + KP).

And when K*~p* = 0, in which case;

If K*p* > B - C then PK (*p* | E)B-C = 1, with certainty level = K*p*/(B – C)

If K*p*< B – C then PK (*p* | E)B-C = K*p*/(B – C).

Where B is the wealth if *p*; C is the wealth if *~p*; K*p* is the positive value of knowledge in favour of *p* | E; K*~p* is the negative value of knowledge in favour of *~p* | E; and KP is the total of the absolute value of knowledge relevant to whether or not *p* | E.

This principle allows that evidential probability varies according to interests and value of knowledge. We will argue that the principle is a normative principle. In order to do this we need to define our terms and give some preliminary definitions. In what follows we will give precise meaning to the value of knowledge and wealth so that the formula can be applied to actual decisions.

### 2.1 Descriptive and normative belief.

Certainty is a measure of both degree of belief and of evidential support. We should call these two aspects of certainty: ***descriptive*** and ***normative***.

A subject is ***descriptively certain*** that *p* relative to her evidence and interests, when she has no doubts and therefore considers bets at any odds on *p* advantageous and bets at any odds on *~p* disadvantageous. A subject who is descriptively certain that *p* will consider bets on *p* to be settled, and will think that she knows that *p*. Descriptive certainty is thus relative to bets and the kind of actions that can be modelled by bets. We do not assume that descriptive certainty relative to low stake bets entails descriptive certainty relative to high stakes bets.

A subject who has an attitude towards *p* that is less than certain, has a ***partial belief*** that *p*. This can be measured by ***subjective probability*** in terms of betting odds. The precise measure is that a subject indifferent between A for certain and B if *p* and C if *~p* will have a degree of belief equal to (A – C)/(B – C). For the purposes of notation:dbkw(*p* | E)U is the degree of belief a subject descriptively has in *p* given E, at stake size U as measured by betting attitudes, given that they know K and have wealth W. Subjective probability is often called “credence”.

A subject is ***normatively certain*** that p given her evidence and interests, iff she ***ought*** to be descriptively certain that p.

One can have normative partial beliefs too. These are measured by ***evidential probability***. Evidential probability is objective in that the evidential probability is completely determined by the evidence and interests, so that any subject with the same evidence and interests ought to have the same degree of belief which is equal to the evidential probability. We will write this Pkw(*p* | E)U. This is the degree of belief a subject ought to have in *p* given E relative to knowledge K, wealth W, and stake size U.

It is possible to be descriptively certain without being normatively certain. We can call this over confidence.

It is likewise possible to be normatively certain without being descriptively certain. We can call this under confidence.



For partial beliefs it is also possible that the degree of belief does not match the evidential probability. One is over confident if one has a higher descriptive degree of belief that *p* than the evidential probability that *p*. One is under confident if one has a lower degree of belief than the evidential probability that p. One should proportion one’s beliefs to the evidence. If one proportions one’s beliefs to the evidence then one has a ***rational degree of belief*** that p.

If S has dbkw(*p* | E)U = Pkw(*p* | E)U = X, then S has a rational degree of belief that (*p* | E)U = X. This means that S is proportioning her belief to the evidence. Rationality is thus conceived of as the best possible fit between action and evidence. In a particular case it means that you actually do have the degree of belief that you ought to have given your evidence and interests.

S has a rational degree of belief 1 that *p* iff dbkw(*p* | E)U = Pkw(*p* | E)U = 1. In this case the substance of our thesis is that S is rationally certain that *p* and S knows that *p*.

### 2.2 Evidential externalism

Evidential certainty on this view entails the truth of *p*. One ought never to be certain of a proposition if it is false. This amounts to ***evidential externalism***. Evidence in this sense has to be public; it has to be presentable to third parties who share the necessary concepts and language. Two people with access to the same documents have access to the same evidence in this sense. Two people with the same view of a crime scene, have the same evidence. Whereas two people having identical experiences in a phenomenological sense do not necessarily have the same evidence. A brain in a vat does not have identical evidence to an ordinary person, even if their sensory inputs are identical. Someone who dreamed that they witnessed a murder is not on the same evidential footing as someone who actually witnessed the murder; however lifelike the dream. Illusions, hallucinations and all forms of deception do not count as evidence from an externalist point of view.

Given evidential externalism, evidential certainty entails truth. But this does not mean that whatever is evidentially certain is necessarily true. Inductive evidence is ampliative, in that an inductive argument contains more information in the conclusion than exists in the premises. This means that it is always *possible* that a proposition which is evidentially certain is false. But this possibility is metaphysical and is consistent with certainty. If the possibility were *epistemic* on the other hand, then the proposition would not be evidentially certain.

### 2.3 Propositional functions

We will think of the conditional *p* | E not as a proposition, but rather as an inductive inference between two ***propositional functions***. For a belief in *p* | E to have any practical use, E and *p* must refer to repeatable classes of propositions capable of both relative frequencies and inductive generalisations. In this case one’s degree of belief in *p* | E can inform one’s decision whether or not to bring E about; and past instances of *p* & E and ~*p* & E can provide inductive evidence for one’s degree of belief in *p* | E. A belief in a propositional function *p* | E is not a belief in a fact that it is true or false, but a belief in the validity of an inference from E to *p*. If such a belief is stake size variant, it simply means that you have more confidence in the inference when the stakes are lower. This means that you have no particular belief in the propositional function, but rather, evidence for and against the propositional function, that issues in particular degrees of belief in particular propositions at different stake sizes. The reference class problem is tackled head on here. The reference class in use is determined by the propositional function. The correct propositional function to use is determined by the agent’s interests and evidence. It is the one that best applies to the future actions of the agent so that gains and losses can be balanced.

Propositional functions are incomplete and need to be completed with individuating space/time and world arguments in order to become propositions that obey the law of excluded middle. We can indicate the completed propositional function with an individuating superscript. So *p* | Ei becomes a proposition equivalent to the disjunction *p*i or ~Ei. In other words, if E is true at time/space, world i, then *p* is true at i. The propositional functions must be such that the individuating superscript is ***evidentially neutral*** with regard to the evidential probability PKW(*p* | E)U. Individuating features indicated by superscripts x and y are ***evidentially neutral*** iff for all x and y, PKW(*p* | Ex)U = PKW(*p* | Ey)U.

To illustrate: The statement “A coin will land heads given that it is flipped” is not a proposition that obeys the law of excluded middle. It is neither true nor false. It is a conditional statement involving two propositional functions: coin flipping, and landing heads. We can however attach a probability to this conditional statement. We can say that the probability that a coin will land heads given that it is flipped is equal to ½. In order to make the conditional into a proposition we must specify the particular instance of coin flipping. We can do this by specifying time/space and world. So the statement: “If a coin is actually flipped on this table top at noon on 11/02/2010 then it will land heads” expresses a proposition that is either true or false. It is true if the coin is flipped and lands heads, false if the coin is flipped and does not land heads, and true if the coin is not flipped.[[10]](#footnote-10) The individuating features “on this table top at noon on 11/02/2010” are evidentially neutral unless we have special information. So the evidential probability that a coin will land heads given that it is tossed is equal to the evidential probability that a coin will land heads given that it is actually tossed on this table top at noon on 11/02/2010.

### 2.4 Reference class

We can then see that *p* | E can range over a restricted domain. Instead of specifying a single proposition, *p* | Ei, we can specify a ***reference class*** of propositions *p* | E1-N. The superscript 1-N defines the set of propositions that form the reference class. In this case, there can be a ***ratio*** of cases in which *p* & E is true to cases in which E is true in the reference class and this ratio will satisfy the axioms of probability. Our belief in *p* | E ought to be proportional not only to our evidence, but also to the range of cases that will practically concern us. In the course of the practical life of an individual there will only be a finite number of times when *p* | E is relevant. These cases will be conditioned by the utility differences between *p*&Ei, *~p*&Ei and ~Ei. As a convention we can call:

The utility of *p* & Ei: “Bi”, This is the state of wealth having won a bet on *p* | E at time/ space world i.

The utility of *~p* & Ei : “Ci”, This is the state of wealth having lost a bet on *p* | E at time/ space world i.

The utility of ~Ei: “Wi”, This is the state of wealth having declined to bet on *p* | E at time/space world i.

*p* | Ex is only relevant to an action resulting in Ex when Wx falls within in the interval Bx to Cx. The stake size can then be given as Bx – Cx.

We can then define the reference class *p* | EB - C as being the set of all instances of E that have a stake size above zero and below B - C. PKW(*p* | EB-C) is ***evidentially certain*** if *p* is true in all members of this set. In this case the ***ratio in the reference class*** is equal to 1. We will use the symbol “r” to signify the ratio in the reference class.

### 2.5 Certainty external to the evidence

Evidential certainty is therefore ***external*** to the evidence and the interests, since it is metaphysically possible that the same evidence and interests could lead to certainty in one world, and less than certainty in another world. This is because whether E&~*p* is ever true in a situation relevant to the subject at a stake size less than or equal to B - C is contingent.

To make this less abstract we could imagine the multi-verse of possibilities as a warehouse full of urns. Let W | D be the inference between propositional functions:

D (*a bead is drawn from the urn*); and

W (*the bead drawn is white*).

Each urn is an analogy for a whole universe. A subject can therefore only draw beads from a single urn. Let us suppose that a Man can choose to draw a bead from the urn every day, and if it is a white bead, he is rewarded, and if it is a black bead he is punished. If he does not draw a bead from the urn, then its colour makes no difference to him. In the course of his whole life, he draws N beads from the urn. The relevant reference class of drawings is thus D1-N. Suppose that we gather all the beads that a man draws from the urn in the course of his life and place them in a sack. The sack then is an analogy for all the times that the inference W | D is relevant to the Man. The sack then comprises the relevant reference class for this particular man’s degree of belief db(W | D) as conditioned by his interests. Other men and women will draw other beads from the urn, and so will have different reference classes. (Of course our Man can benefit from the experiences of other people through testimony, so N is not necessarily restricted to the beads he has personally drawn, but also includes all the bead drawings that he knows about. However to simplify the example, we will assume that in this warehouse world, testimony is impossible. The Man is thus a proxy for any community of information sharers.)

Supposing the urn itself to only contain white beads, then the inference will always be certain for any agent. However, were the urn to contain a few black beads among the many white beads, then some communities may occasionally draw a black bead and thus have an evidentially probability of less than 1, while others could be lucky enough to pick only white beads and still be evidentially certain.

Let’s suppose that after half his life the Man has drawn K beads from the urn and they have all been white. He is now certain that in general, beads drawn from the urn are white. But the fact that K beads are white does not yet determine that the evidential probability PK(W | D)N-K = 1, for it is *possible* that there are some black beads among the remaining N – K beads in the sack. But this *possibility* does not determine that PK(W | D)N-K < 1 either, for it is also *possible* that there are only white beads in the sack. The evidential probability depends therefore on the contents of the whole sack of N beads, K of which are *internal* to the evidence, and N – K of which are *external* to the evidence. Whether or not there are any black beads in the urn but not in the sack is irrelevant to the Man in terms of expectation but might be of speculative scientific interest to those whose intellectual curiosity extends beyond the narrow confines of their own interests. Whether or not there are any black beads in any of the multi-verse of urns is of interest only to the speculative metaphysician and is surely not relevant to practical certainty for this Man in this world.

Internally speaking, evidence for a judgement *p* | E ranging over a reference class will involve a ***sample*** of the reference class, which we will call *p* | E1-k. The sample is ***representative*** if the ratio “r” of *p* & E true to E true in *p* | E1-K falls within + (1- r)/√|K and – r/√K of the ratio of *p* & E true to E true in *p* | E1-N. (The error margin + (1- r)/√K and – r/√K is approximately equal to the appropriately skewed expected deviation for the expected value for r in E1-Kdivided by the sample size were P(*p* | E1-N) = r). As long as the sampling technique is evidentially neutral, then it is reasonable to expect the sample to be representative if k > N - K. If the ratio in *p* | E1-N = 1, (in other words, if *p*is true in the entire reference class), then the ratio in every possible sample will be 1 and it is necessary that *p* | E1-K will be representative.

For example, if you drew a handful of beads from an urn which contained only white beads, then the ratio of white beads to beads in the handful would necessarily equal the ratio of white beads to beads in the urn. However if the urn contained some non white beads, so that the ratio of white beads to beads was less than 1, then the ratio in the handful would not necessarily match the ratio in the urn. How close the ratios are likely to be is a function of the number of beads in the handful and the number of beads in the urn. If the handful contains more than half the beads in the urn, then it is reasonable to expect that the ratio in the handful will be close enough to the ratio in the urn for predictive purposes. (Don’t get too hung about the error margin + (1- r)/√|K and – r/√K. This is just my attempt at giving a function to make precise what is meant by “close enough”, which is sensitive to the relativity of the term.)

However, if the ratio in *p* | E1-K is 1, (in other words, if *p* is true in the entire sample) then it is not necessary that the sample is representative (i.e. it is not necessarily true that *p* in the entire reference class). This is where the nub of the matter lies. If the ratio in the reference class is 1, then it is necessary that the ratio in any sample taken from that reference class is 1. Any future bets on *p* | E will be taken from the reference class EK-N, since this is how the reference class is delimited, so this necessity applies to all future bets. So the appearance of a possibility that ~*p* | E for some Ei is only internal to the evidence. It is not an actual possibility. For example, from the evidential perspective of an early chemist, it may have appeared to be possible that some samples of water could have a structure other than H2O. But this possibility was only internal to the evidence. It was never *actually* possible that a sample of water could have a structure other than H2O.

If K > N – K, and the ratio in E1-K = 1 then it is quite reasonable to suppose that *~p* | Ei is not actually possible in any E1-N, and it is therefore quite reasonable to be certain that *p* given E. In other words it can be quite reasonable to be certain of a belief given evidence that does not entail the truth of that belief. (For example, it is quite reasonable to be certain that no two non-twins have the same DNA structure).

However, it is also possible to take a sample where r = 1 from a reference class where r < 1. In this case it would be possible to be reasonably certain of a false proposition. So reasonable certainty is not sufficient for evidential certainty. For reasonable certainty, it is sufficient that:

i) The evidence sample must have a ratio of 1;

ii) K must be high enough relative to N to make it reasonable to suppose that the ratio in the reference class is 1.

For evidential certainty a third condition must hold:

iii) The evidence sample must be sampled from a reference class where the ratio is 1.

This third condition is the surrogate for the truth condition for knowledge. It follows from this third condition that all known propositions are true propositions. We can thus grant the factivity claim that knowledge that *p* entails *p*, without granting the infallibility claim that if the evidential probability of *p* given E relative to K is 1 then it is necessarily true that *p* given E and K.

If PK(*p* | E)1-N was certain under ***evidential internalism***, then *p* would have to be true in all possible worlds where K and E were true. Evidential internalism is committed to limiting evidential certainty to tautologies. Thus evidential internalism is infallibilist about evidential certainty and entails either scepticism about knowledge of contingent fact or a Lockean threshold theory of justification. But, as Hume revealed, the problem does not just apply to certainty, it also applies to any threshold probability applied to unobserved cases, so the Lockean threshold theory cannot solve the problem. This provides a pragmatic argument against evidential internalism. The argument is that internalism can never allow inductive certainty, since no amount of evidence will necessitate ampliative conclusions. In short: evidential internalism fails as a theory of justification because no inductive arguments are justified under evidential internalism.

### 2.6 Stake size, wealth and utility

Although we do not wish to give a substantive theory of value in this thesis, the concepts of wealth and stake size need careful treatment. ***Wealth*** defined as the sum total of a subject’s resources must be distinguished from ***subjective utility*** defined as a measure of the subject’s preferences. Descriptive degrees of belief should be measured in terms of preference based expected subjective utility. Evidential probability should be measured in terms of resource based expected wealth.

Wealth is important epistemologically because of the relationship between wealth, stake size and repeatability. The relevant reference class for an individual will depend on the amount of losses the individual can sustain before ceasing to function. This is a fully objective and mind independent quantity.

The important feature of wealth in terms of bets and probability is repeatability. A person with wealth “W” can afford to make a loss “L” W/L times before he must stop. This has significance in terms of the amount of lifetime error one can expect in calculations of expected utility. It is an objective fact that one will make a certain number of bets, {E1, …EN}, on *p* | E in one’s life, with E1 being the first and EN being the last. This set we call the ***reference class*** which defines the relevance of judgements of P(*p* | E) to the subject of these bets. The size of the reference class, N, will depend on the agent’s decisions. If the stakes are very high relative to the agent’s wealth, then it is likely that N will be low. “Life time” need not be the literal life of the subject, but the lifetime of any function or activity which requires a certain level of success to continue. Wealth then is a measure of whatever the self sustaining success consists in.

We can think of a particular decision maker as adding to his knowledge “K” through each successive bet. The judgement then will be of the likely frequency of *p* given EL-N on the basis of the frequency of *p* in E1-K, supposing the present to lie between K and L. If there are big losses early on, then it is likely that N will be small, whereas if there are big gains early on, then it is likely that N will be much larger. The smaller N, the larger the expected deviation between the relative frequency of *p* in EL-K and the relative frequency in El-N­.

This gives us two mathematical arguments for stake size variation. The arguments depend on the principal that evidential probability relative to a reference class should aim to match the ratio within that reference class, adjusted for stake size. We can first assume some stochastic process with an objective probability of 1/2, like coin tosses. The interpretation of objective probability here is simply any interpretation that allows us to say in advance what the probability of any frequency outcome is given the number of trials. Supposing there to be N coin tosses in the reference class, and H out of those N times the coin landed heads, then the evidential probability of heads should aim at H/N. We then consider two cases:

**1. A subject with initial wealth W bets on heads at stakes U and fixed odds 1 : 1 and continues betting until she runs out of money.**

In this case, the subject will run out of money whenever the number of tails T exceeds the number of heads H by 2W/U. With this stopping rule, then for all finite N, T > H so that H/N <½ . If U is very large relative to W, then N is likely to be small and H/N has a high probability of being a lot smaller than ½. For example, if U = 2W, then H/N has a 0.5 probability of being equal to zero and a further 1/8 probability of being equal to 1/3. However, if U is very small compared with W, then N is likely to be very large and consequently H/N is likely to be close to ½. For very small U, we might suppose the game to continue indefinitely for most practical purposes since in reality, the game is more likely to stop for some extraneous reason, like boredom. In conclusion, in games with this structure, then H/N will be less than the objective probability, and this difference will increase in magnitude as the stakes get higher. H/N ought to guide evidential probability, not the objective probability; so evidential probability ought to vary with stake size when this kind of model is applicable. We will call the phenomena of the frequency tending to be below the objective probability at positive stakes: ***Gamblers Ruin***. This is because a gambler who repeatedly bets his entire holdings at even odds on an event with an objective probability of ½, will *certainly* be ruined. Whereas the expectation of this strategy, using objective probabilities that ignore stake size, is equal to the gambler’s initial holdings. Such a gambler will not *necessarily* be ruined, because it is *possible* that he will win again and again infinitely. An infinite run of heads on a fair coin is metaphysically possible, but has an evidential probability of 0. So we have here an illustration that probability 1 does not entail metaphysical necessity, and probability 0 does not entail metaphysical impossibility.

**2. A subject with initial wealth W1 bets repeatedly at even odds at stakes equal to a fixed proportion 2L of his current wealth**.

We are to imagine that “L” is a fraction between 0 and 1. Each time the coin lands heads, the Gambler wins LWN and his new wealth WN+1 = WN + LWN . Each time he loses he will pay LWN and his new state of wealth WN+1 = WN – LWN . In this case, the gambler will never run out of money just so long as L < W. But there will be a tendency to negative rate of growth with the rate becoming more extreme as L approaches W. This phenomenon was given detailed treatment in Kelly (1956). In this game, although H/N will tend to ½ as N tends to infinity, the rate of growth G per bet will tend to √(1 – L)(1 + L) as N tends to infinity. Whenever G < 1, the growth will be negative, and the gamblers fortune will tend to decline the more times he gambles. It is the rate of growth which in this case is relevant to evidential probability since it is the rate of growth which calculations of expectation are *about* and to which the law of large numbers applies. We can derive the degree of belief from the rate of growth using Ramsey’s formula.

A subject who is indifferent between

i) √(1 – L)(1 + L), and

ii) 1 + L if *p* and 1 – L if ~*p*

Has a degree of belief in *p* equal to:

{√(1 – L)(1 + L) – (1 – L)} / 2L

This makes the degree of belief in *p*< ½ for all values of L where L is between 1 and 0.

Many people find this phenomenon surprising and are initially incredulous, so it might help to illustrate with an example. Suppose L = 1/2W and the starting wealth was £100. Now suppose that H/N = ½ for all even N. On a naïve view, we might expect that W would remain constant since the gambler would win as many bets as he lost and each bet would be at 1 : 1 odds. But this is not so. The growth rate is √0.75. So after N bets, the median value for W = √0.75N. If N = 2, then the gambler will end up with £75. He will either win the first bet, gaining £50 and grow to £150, then lose the second bet and fall back to £75; or he will lose £50 on the first bet, and fall to £50 and win £25 on the second bet and grow to £75. If N = 4 then the final wealth W = £56.25. If N = 6, W = £42.1875. If N = 40 the final wealth will have dwindled to £0.317. As N tends to infinity, W tends to 0. If the subject is indifferent between these bets and the median value, then his degree of belief in H = 0.366.

Due to the law of large numbers, the probability that H/N will equal ½ when N is even tends to 1 as N tends to infinity. Therefore the probability that W at N will equal the median will approach 1 as N approaches infinity. If we assume the infinite divisibility of wealth and assume that any game is non terminating, then all games will tend to zero wealth with probability 1. We will call this the ***Kelly phenomena***.

Now let us suppose that in stead of betting a fixed portion of our wealth we consulted Kelly’s formula to find out the ideal value for L (See p.115). The fixed proportion is calculated by the formula edge/odds. In the current example, we can assume that the optimum value for e is going to be 0.5, making the edge equal 0 and therefore L equal zero. So the Kelly criterion suggests never betting at fair odds. But what if we knew in advance that the coin always landed heads once for three tails so that e = ¾? As it is even odds we can use L = 2e – 1. So with this special information L = 1 ½ - 1 = ½. Now we imagine a sequence where every fourth toss lands heads and the others tails. The Kelly gambler, let us say, starts with £100. First toss he wins £50, second toss he wins £75, third toss he wins 112.50 and on the fourth toss he loses £168.75, making his wealth after four tosses equal to £168.75. Once again the order isn’t important. We can see that with this betting system the gamblers wealth will grow faster than any smaller value for L. But what if L was larger than ½? Let’s suppose it was .8. First toss W = £180; Second toss W = £324; Third toss W = £583.2; Fourth toss W = £116.24. So all though he still makes a profit, it is not so great as it would have been had he kept his stakes lower.

It should be noted that the *mean* return over every possibility in this game given fixed N is still W. But as N gets larger, the probability of ending up with greater than W becomes increasingly tiny, and this super unlikely gain will become increasingly astronomical. In the above example where N = 2, two consecutive wins results in £225, a gain of £125, whereas two consecutive losses results in £25, which only results in a loss of £75. So the average across possibilities still comes out as £100. But even with N as low as 2, there is a 0.75 probability that your final wealth will be less than the mathematical expectation of £100. As N increases, this effect increases and the probability that your final wealth will be less than the expectation increases apace. Eventually, as N tends to infinity, the probability that your final wealth will be less than the expectation will tend to 1, and the expectation will be balanced by the infinitesimal probability of an infinite gain. This, I put forward, is a kind of reductio argument against the principle of expected utility using objective chances. The reductio argument has it best expression in the absurd infinite expectation of the St Petersburg game, where every possible outcome of the game is infinitely less than the mathematical expectation calculated by taking the sum of the products of the objective probabilities and the values of every outcome.[[11]](#footnote-11)

### 2.7 Wealth is directional

Wealth and value are directional, in that there is a fundamental difference between loss and gain. There is a lower bound to wealth. An agent with wealth W cannot, by definition, make a loss of more than W. Of course in this age of credit, it is possible to mount up debts that you will not be able to repay. But the loss then is passed on to your creditors. There is no corresponding conceptual upper boundary to wealth.

Evidential probability is objective in the sense that it is about what anyone rational *ought* to believe. So the stakes should be given in a measure of *objective* value. This is how much a subject *ought* to value a prospect. Wealth is measurable in units derived from any finite regenerative resource. Money is the chief unit since it is so generally applicable .Labour hours, energy, water, food, and lives are all regenerative resources. A business with capital of W can only afford to make a loss of L, W/L times. But there is no theoretical limit to how many times such a business can gain G. A species with W breeding pairs, can only afford to lose L breeding pairs through predation W/L times before they become extinct. An airplane with W reserves of fuel can only lose L fuel W/L times before it falls out of the sky. These are hard objective facts governed by physical laws and have nothing to do with preference or subjective value. Of course, a rational creature who values its own function ought to prefer gains in wealth to losses in wealth. But there is no logical reason why a creature cannot prefer its own destruction to its own reproductive success. In spite of this, the logic of evolution tells us that most surviving creatures capable of choice will prefer reproductive success to an early barren death.

Empirical upper boundaries can exist. There may well be an upper limit to any natural metric of growth. Human population is an example. We may have reached a critical level where a doubling of the population from six billion to twelve billion will not be possible due to the natural limitations of the earth. The probability that a couple will have four children that survive to maturity is consequently lower when applied to the entire population of six billion people, than it is when applied to the previous generation. These kinds of upper limit appear all over the place. Markets can be saturated, storage facilities can be full to capacity and natural resources can be depleted. When any metric of wealth reaches such a saturation point, the probability of gains will fall and the probability of losses will rise accordingly until an equilibrium is reached.

### 2.8 The Value of Knowledge

The ***value of knowledge*** is a term we have borrowed from Ramsey, and is a measure of, in Ramsey’s own words: “*How much is it worthwhile to find out K?*” (Ramsey, in Galavotti 1991, p. 285). The value of knowledge has a family resemblance to Turing’s “Score”, Good’s (1950) “Weight of Evidence” and Popper’s (1974) “Severity of test”, in that it measures the value of evidence in terms of change in probability. But the value of knowledge is different from these measures because these measures ignore stake size, and therefore implicitly assume stake size invariance, which we reject. The value of knowledge, on the other hand, determines how much the probability varies according to stake size. The value of knowledge thus plays a central role in the Stake Size Variation Principle. For this reason we are coining the term “value of knowledge” to avoid confusion with weight of evidence and other measures.

For notation:

k*p* = The value of knowledge in favour of p.

k*~p* = The value of knowledge against p.

KP = The absolute value of knowledge relevant to whether or not p.

The value of knowledge is measured in terms of the increase in value that the evidence (or knowledge) brings. The value of knowledge is commensurate with the value of expectation, wealth and stake size, and can be measured on the same scale.

The value of knowledge that *p* relative to a bet resulting in B if *p* and C if *~p* is equal to B – C, or the whole of the stake. The value of knowledge that increases the probability of *p* by x relative to a bet resulting in B if *p* and C if *~p* is therefore equal to x(B – C), which is the increase in expectation that the knowledge would bring. The Size Variation Principle reverses the direction of this inference, and has the probability increase due to new knowledge as being a function of the value of knowledge and the stake size. But since there is no privileged stake size, then we read the probability directly off the stake size and the value of knowledge.

We can explain this with reference to a literal bet on (*p* | Ex)B - C. Once you have laid a bet, the money you put into the stake no longer belongs to you, but the entire stake belongs to the winner. In ignorance of whether you are the winner or not, then the stake is in a state of neither being owned by you, nor not owned by you. If you lose the bet, then you forfeit the entire stake and your wealth amounts to C, and if you win the bet and the entire stake was added to your wealth, then your entire wealth amounts to B. Therefore, bet settling evidence that *p* is worth the entire stake which is equal to B – C. Bet settling evidence that *~p* is worth C – B, which is a forfeit of the entire stake. Since our thesis is that bet settling evidence is sufficient for knowledge, then the value of knowledge that *p*& Ex relative to this bet is equal to B – C. And the value of knowledge *~p* &Ex *­*is worth C – B.

Now suppose that we had the knowledge that Pk*p*(*p* | E)B – C = 1. This would entail evidential certainty that (*p* | E) at stake size B – C and therefore k*p* would provide bet settling evidence for any bet on *p* given Ex at stakes up to B – C. So the value of the knowledge k*p* relative to the bet on (*p* | Ex)B - C = B – C. This means that the value of knowledge k*p* is sufficient to settle bets on any bets (*p* | Ey) B – C just so long as the individuating superscripts x and y are evidentially neutral.

We propose that the value of evidence sufficient to settle a bet on (*p* | E)B – C is thus equal to B – C and therefore the value of evidence necessary for certainty and knowledge at stake size B – C is a minimum of B – C, or the whole of the stake size.

Given the same bet, we can also give the value of knowledge that PK*p*(*~p* | E)B - C relative to a bet on (*p* | Ex)B - C. In this case, the knowledge is equal to (C – B) since you *lose* the whole of the stake. Because in this case C – B is negative, we count this as negative value of knowledge pertaining to (*p* | E).

Positive and negative value of knowledge can be understood simply in terms of the type of error the knowledge avoids. Positive knowledge that *p* | E will avoid underestimates of the expectation of a bet on *p* | Ei, whereas negative knowledge that *p* | E will avoid overestimates. Positive knowledge provides reasons for an action, whereas negative knowledge provides reasons against an action. For short hand we can call positive value of knowledge “k*p*” and negative value of knowledge “k*~p.*” “Positive” and “negative” here have quite a special meaning that is distinct from plus and minus, and is more akin to for and against. With a bet on (*p* | A) B – C, where the agent has a choice over whether to act so that A, or not act so that ~A; positive value of knowledge *p* takes for granted that the expectation has value C, and the value of knowledge that *p* increases the expectation from C upwards. In other words, no knowledge whatsoever in favour of *p* would still give the wager a minimum expectation of C. So if all other options had an expectation of less than C, no positive knowledge whatsoever would be required to act so that A. Positive knowledge is thus *added* to C, more value of knowledge makes the prospect *more valuable than C*.

But negative knowledge is just as valuable. Negative knowledge takes for granted that the wager will be no more valuable than B. So no negative knowledge whatsoever is required to refrain from acting if there is any other option with an expectation more than B. Negative knowledge is thus *added* to B. Negative knowledge makes the prospect *less valuable than B*, and thus adds to the desirability of *not acting* so that A. Both positive and negative knowledge thus add to your wealth, but in different ways. Because knowledge is factive, the external situations are different when you know that *p* and when you know that *~p*.

The total value of knowledge pertaining to whether or not *p* is equal to the sum of the *absolute* value k*p* and k*~p*. A good metaphor to explain the value of knowledge is the scales of justice. Probability can be seen as the ratio between the evidence for and against, whereas the value of knowledge is the entire mass of evidence for and against. The more value of knowledge is already on the scales, the less difference a new piece of evidence for or against will make to the probability. The stakes create a bias in the scales, so that reasons against weigh more than reasons for when the stakes are positive. To keep with the image, the Stake Size Variation Principle adds the value C on the “for” side and the value B on the “against” side, thus creating a stake sensitive bias towards *~p* at positive stakes.

We take it to be a general principle of inductive evidence that instances of (*p* &Ei) count as positive evidence for (*p* | E) and instances of (*~p* & Ei) count as negative evidence for (*p* | E). We can enumerate this principle by keeping score of both positive and negative value of knowledge k*p* and k*~p*, in PK (*p* | E), and adding the stake size of any bet that is settled accordingly. So if the old positive value of knowledge (*p* | E) = k*p*, and a bet on (*p* | E)U is settled in favour of *p* through bet settling evidence, then the new value of knowledge (*p* | E) = k*p* + U. Whereas if the old negative value of (*p* | E) = k*~p*, and a bet on (*p* | E)U is settled in favour of *~p* with bet settling evidence, then the new value of knowledge (*p* | E) = k*~p* + U. The value of knowledge relevant to *p* | K will then equal the running total k*p* + k*~p* = KP. The stake size variant evidential probability that *p* | E will then be a function of the ratio k*p*/KP and the stakes according to the formula (C + k*p*) / (B + C + KP). In a simple case where each bet was at the same stake size, then the running total k*p*/KP will equal the relative frequency.

For example, suppose you start with £100 and you kept betting at stakes £2 that a slice of toast dropped from a height of three feet will land butter side up. Let us call this propositional function *b* | *d*. Suppose you start with no knowledge. Each time you drop a piece of toast, you will add £2 to either k*b* or k*~b* depending on what way the toast lands. So k*b*/ Kb is the ratio of times when the toast fell butter side up to the total number of toast droppings. Now let us see what happens to the evidential probability. In order to keep the example simple, let us suppose your wealth remains constant due to outside influences and the stakes on each bet are £101 - £99 at even odds. We will then take a time slice and compare evidential probabilities at stakes £101 - £99 with stakes £150 - £50.

At first you have no evidence, so there is no evidential probability. The second time, supposing the toast landed butter side up the first time, you will be certain that *b | d* at stakes £101 - £99. But due to the externality of evidence, you will only be evidentially certain if the second slice lands butter side up. At stakes £150 - £50, Pk(*b | d*)£150 - £50 = k*b*/U = 2/100. So at this value of knowledge, the low stakes certainty is very fragile and becomes almost complete doubt at high stakes.

Let us say the frequency of *b | d* tends to 1/3 in the long run. Let us say that after six droppings k*b* / Kb = £4 / £12. Your evidential probability *b | d* = (C + k*b*) / (B + C + K) = (99 + 4) / (£200 + £12) = 103/212. We can see that, as the value of evidence is so low relative to the wealth, and because the stake size is so low relative to the stakes, then there is a strong pull towards the logical probability, and the evidential probability is close to ½. But if the stake size were significantly increased at this level of knowledge, then this would drive the evidential probability down. So suppose B – C = £150 - £50; then Pk(*b | d*)£150 - £50 = (C + k*b*) / (B + C + K) = £54/£212, which is closer to 1/4. The ignorance makes the degree of belief much more sensitive to the stakes, avoiding high stakes bets and encouraging experimentation at low stakes.

Now we can fast forward to a time when we have dropped thousands of pieces of toast so that k*b* = £10 000 and k*~b* = £20 000 making Kb = £30 000. The same bets at stakes £101 - £99 now has an evidential probability Pk(*b | d*)£101 - £99 = (£99 + £10000)/(£200 + £30 000) = 10 099/30 200, which is now approximately 1/3, the observed frequency. If we raise the stakes again to £150 - £50, then Pk(*b | d*)£150 - £50 = £10 050/£30 200, which, although lower, is still approximately 1/3. The general pattern is as your knowledge increases relative to your wealth, then the evidential probability tends towards the relative frequency, and the probability does not vary so wildly with changes in stake size.

In order to derive the stake size variant evidential probability from the value of knowledge and the stake size we must examine the connection that the value of knowledge has with expectation. First we will consider the case when there is only positive knowledge that *p* | E. As we have said, certainty that *p* at stake size B – C requires that k*p* = B - C. The expectation on a bet B if *p,* C if *~p,* given only positive knowledge k*p* is in general equal to:

C + PK*p*(*p* | E)B – C(B – C).

As we have seen, if k*p* = B – C, then PK*p*(*p* | E)B – C = 1. In this case the expectation is equal to:

C + 1(B – C) = B.

Therefore, the expectation is equal to:

C + k*p*.

In constructing the measure for the value of knowledge, we propose using certainty as an anchor point and generalising from the case when k*p* = B – C, to all values for k*p*.

So, whenever there is only positive value of knowledge and k*p* < B - C, the expectation is also presumed to be equal to:

C + k*p*.

And therefore PK*p*(*p* | E)U = k*p*/(B – C).

But if k*p* is larger than (B – C) then we have an embarrassment of riches and k*p*/(B – C) is greater than 1. This means that relative to a bet at stake size B - C, *p* is *more than certain*. There is nothing self contradictory about the concept of being more than certain, and in fact it is a fairly common expression. So there no reason to reject the principle on the grounds that it makes it possible to have a surplus of positive evidence for a proposition. It often is the case in science and practical reasoning that we have more evidence than is necessary for certainty. It is clearly possible to be in an evidential position such that we are certain of the results of a particular experiment before we have performed it. In which case, when we do perform the experiment, the experimental results will increase our store of knowledge to beyond that which is sufficient for certainty.

But the evidential probability cannot be greater than 1 for the simple reason that a bet that results in B if *p* and C if *~p* cannot have an expectation higher than C + 1(B – C) since this would make the expectation higher than B, which is absurd. The value for k*p*/U, when it exceeds certainty, can be called the ***level of certainty*.** Although increases in the level of certainty make no difference to the evidential probability, the level of certainty can still make differences in terms of decisions. Firstly it can give a weak preference ordering between prospects with the same expectation. For example, given a choice between a pair of prospects;

1. B if *p,* C if ~*p*
2. B if *q*, C if *~q*.

Where there is only positive knowledge, and k*p* > k*q* > B – C, then i) is more preferable than ii) even though both *p* and *q* are certain relative to the stake size and the expectation on both prospects is B.

Secondly, it can motivate actions that increase the stakes. For example, given a choice between

1. B if *p*, and C if ~*p*
2. A if *p*, and D if *~p*

Where A > B > C > D, and there was only positive knowledge that *p*, then the exact value of k*p* could make a difference between whether i), or ii) was preferable. For example, if k*p* ≥ A – D then ii) would be preferable, whereas if k*p* ≤ B – C, then i) would be preferable.

Also, it serves a function in belief updating in cases of *epistemic shock,* where we get a piece of negative evidence for a conditional of which we were formally certain. A disadvantage of classical Bayesianism is that certainty is monotonic, and evidence is taken to be certain. This has the unpalatable consequence that one can never revise one’s evidence propositions. But in reality, many observations later turn out to be errors, and probability 1 is often assigned to propositions which later turn out to be less than certain or even false. But stake size variable certainty is blatantly not monotonic, because certainty at low stakes can become less than certainty at high stakes. Cases where k*p*/U > 1, and yet we come across an instance of ~*p* at stakes < U, we will call an *epistemic shock*, and the precise magnitude of k*p*/U will determine the post shock evidential probability assigned to *p*. We will go into this in more detail in section 7. For now, allow that when k*p*/U > 1, then PK*p*(*p* | E)U = 1, with ***certainty level*** k*p*/U.

Now we consider the case when there is only negative knowledge that *p*. The expectation on a bet B if *p* and C if *~p* is equal to:

B + P(~*p* | E)B - C(C – B)

As we have seen, when K*~p* = (C – B) then P(*p* | K) = 0 and the expectation becomes:

B + 1(C – B) = C.

And because k*~p* = (C – B), then the expectation equals:

B + k*~p*.

As with the case of positive knowledge we generalise from the case where the value of knowledge is the minimum necessary for knowledge to all cases where there is only negative value of knowledge that *p,* so that:

PK*~p*(*p* | E)B – C = k*~p*/(B – C).

Now, to form the Stake Size Variation Principle in cases where there is both positive and negative evidence for *p* | E, we propose forming an odds ratio of the expectation given only positive knowledge k*p* to the expectation given only negative knowledge k*~p*.

C + k*p* : B + k*~p*

Converting this into a probability we get:

PK(*p* | E)B – C = (C + k*p*) / (C + k*p* + B + k*~p*)

Adding k*p* to k*~p* to form the absolute value K we get the Stake Size Variation Principle:

PK(*p* | E)B – C = (C + k*p*) / (B + C + K).

And, for *~p*

PK(~*p* | E)B – C = (B + k~*p*) / (B + C + K).

### 2.9 The Stake Size Variation Principle

The equations at the end of the last section are of central importance to this thesis. They allow the evidential probability to be calculated from the stakes and the knowledge of the subject. So for emphasis we will repeat the definition.

**Stake Size Variation Principle**:

The evidential probability of probability of proposition *p* relative to stakes B if *p* and C if *~p,* and to knowledge KP comprising of only positive knowledge k*p* is given by:

In cases where KP< (B – C), then PK (*p* | E)B-C = KP/(B – C).

In cases where KP> (B – C), then PK (*p* | E)B-C = 1 with certainty level KP/(B – C).

The evidential probability of proposition *p* relative to stakes B if *p*, C if *~p*, and to knowledge KP comprising of both positive knowledge k*p* and negative knowledge k*~p* is given by:

PK (*p* | E)B-C = (C + k*p*)/ (B + C + KP).

From this the following useful results follow:

1. Whenever there is only positive knowledge and KP > U, then *p* | EU is evidentially certain. This makes inductive certainty possible, as well as allowing that S can know *p* at low stakes and not know *p* at high stakes given the same evidence measured in value of knowledge.

2. For all bets at all stake sizes B - C and all values of knowledge KP and evidence E: PK (*p* | E)B-C + PK (*~p* | E)C – B = 1.

3. The evidential probability PK (*p* | E)B-C will tend to get lower as the magnitude B – C gets higher. We can call this ***stake size variation***.

4. The greater the knowledge relevant to *p* | E, the slower the ***rate of stake size variation****.*

5. There is a bias in favour of the logical probability ½. This bias is proportional to the wealth of the individual relative to the relevant knowledge.

This last feature resonates with the Objective Bayesian intuition that the logical probability should form the objective prior. In the absence of any information, then the probability of a proposition at zero stakes is ½. As the stakes B - C increase positively relative to C, then the probability decreases as a function of C / (B – C). When C = 0, in other words, when losing the bet means ceasing to function, then the probability goes right down to zero. As more knowledge is gained, then the tendency to ½ is retained but lessens in effect depending on how resilient the subject is to losses. This would have the normative effect that, if you are rich and the stakes are low, you should give better odds to unlikely events, because the value gained in information outweighs the cost in wealth. The tendency to zero as C tends to zero is also lessened with any amount of positive knowledge. This has the normative effect that if your range of options is very poor, then you should value any option which has ever had any success in the past above options about which you have no information.

This simple formula then does explain a good deal of information guided decisions in a principled way that gives a unified approach to the evaluation of evidence and of outcome. However, it is to a certain extent arbitrary in that it is chosen from a continuum of possible formulas. In this thesis I shall argue that evidential probabilities derived from the Stake Size Variation Principle are better calculators of expected utility that stake size invariant beliefs based on objective probabilities. I will not argue that the Stake Size Variation Principle here presented is better than any other conceivable stake size variation principle.

**Multiple Valued Stake Size Variation Principle:**

One obvious supplement to the formula is to extend it to partitions of more than two possible outcomes. In order to do this we must rank the value outcomes and pair them off so that the highest is paired with the lowest and the second highest is paired with the second lowest etc. As in the binary case, the numerator is formed by the value paired with the *p* value plus K*p*, whereas the denominator is formed by the sum of all the values plus the total knowledge.

So a prospect {*p*1; V1,..,*p*i; Vi ,.., *p*N; VN}, with N propositions *p*1-N and N values, V1-N has evidential probability: PK(*p*i)V1-N = (VN+1-i + K*p*i) / (∑1-NVi + K).

For example if there was an option:

A if *p*1, B if *p*2, C if *p*3 and D if *p*4 such that A > B > C > D, and *p*1-4 formed a mutually exclusive jointly exhaustive partition, then A would be paired with D and B would be paired with C.

And PK(*p*1)U = (D + K*p*1)/(A + B + C + D + K)

PK(*p*2)U = (C+ K*p*2)/(A + B + C + D + K)

PK(*p*3)U = (B + K*p*3)/(A + B + C + D + K)

PK(*p*4)U = (A+ K*p*4)/(A + B + C + D + K)

It follows immediately that PK(*p*1 OR *p*2 OR *p*3 OR *p*4)U = 1.

With odd numbers the middle value is paired with itself.

There are a few problems with this. In a case where there are many identical low values, and one high value, then the evidential probability would depend arbitrarily on which of the low values was paired with the high value. This would be the case in, for example, a lottery. A pragmatic solution here would be to treat any propositions with identical values as the same proposition. In this case a lottery would be seen as the binary proposition: win or lose. This would have the effect of making lotteries with high prizes and low ticket costs as preferable to their actuarial value. So it would explain why people buy lottery tickets. This is not entirely satisfactory, since in many cases we will want to discern between probabilities with equal value outcomes, and we might want to retain the idea that it is in fact irrational to buy lottery tickets. But these problems are not directly relevant to the theory of knowledge and can be set aside for further research.

A further problem comes with deciding how to deal with a continuous case. This could be approached by taking the deviation from the mean value and reversing the sign for the numerator. This again requires careful thought and can be set aside for present purposes.

### 2.10 Bid - Ask Spread.

Given a way of calculating the stake size variant degree of belief, we can give a value for the ***bid-ask spread***. The ***bid-ask spread*** is the degree to which a rational degree of belief swings as the stakes switch from positive to negative. According to the Stake Size Variation Principle, you should typically have a higher degree of belief in *p* when betting against *p* than when betting for *p*. We can call this the ***bid-ask spread*** since in asset pricing terms it is the difference between the amount you would ***bid*** for a bet where you gained U if *p* and how much you would ***ask*** for a bet where you had to pay U if *p*. If we divide through by the stake we get a ***probability interval*,** which is the difference in degree of belief at positive stakes and negative stakes.

Bid-ask spreads are a much studied phenomena and there is a good deal of evidence that bid-ask spreads vary with indicators of stake size (per unit volume), indicators of repeatability (market volume) and indicators of epistemic strength (variance). (Frank & Garcia, 2009; Gregoriou, Ioannidis &Skerratt, 2005; Mian, 1995).

The bid ask spread is thus PKW(*p* | E)-U ­– PKW (*p* | E)U.

This is:

(B + K*p*)/(B + C + K) - (C + K*p*)/ (B + C + K)

So:

S (bid-ask spread) = (B – C) / (B+ C + K)

We can then isolate K and work out the value of knowledge from the spread and the utilities.

K = (B – C)/S – (B + C).

### 2.11 Probabilistic theory of knowledge

The probabilistic theory of knowledge can now be stated in terms of two necessary and jointly sufficient conditions. S knows that *p* given E at stake size U with knowledge K iff

1. S has a degree of belief 1 that *p* given E at stake size U.

2. The evidential probability of *p* given E at stake size U relative to knowledge K is equal to 1.

Knowledge, then, entails both subjective and evidential certainty.

The concept of evidential probability plays an important part in the theory. Evidential probability in the probabilistic theory of knowledge has the following substantive features:

1. Evidential probability is external to the evidence.

2. Evidential certainty entails truth.

3. Evidential probability varies as a function of the value of knowledge Kp + K~p and the stake size B - C according to the formula: PK (*p* | E)B-C = (C + K*p*)/ (B + C + K) when K*~p*> 0; or (C + KP) when K*~p* = 0.

The theory rests on the idea that degrees of belief and evidential probability are relative to decisions. The evidential probability (*p* | E)B – C is ***relevant*** to all decisions where the value of ~E falls between the value of B and the value of C. A degree of belief 1 is where the value of a bet (*p* | E)B –C = B.

Evidential certainty is when certainty is backed by knowledge of greater value that B – C. This is when KP is all positive and when KP> B – C. Certainty is thus relative to the stake size B – C.

This means we have an updating rule which is sensitive to the value of gathering evidence relative to a decision. Beliefs backed by more value of knowledge are more resilient in the face of new evidence. The relative value of gathering new evidence relevant to a decision will thus be well defined in each particular case.

## 3 Objection: Stake Size Invariance is a constraint on probabilistic coherence.

### 3.1 The age old measure.

The main objection to the Stake Size Variation Principle is that degrees of belief that vary according to the stakes are probabilistically incoherent and leave the Subject vulnerable to a Dutch book. In order to examine and counter this objection we need to be clear about our concept of degrees of belief and why probabilistic coherence is relevant.

The age old measure of degree of belief is in terms of the minimum odds at which you would accept a bet. The immediate problem with the age old measure is that it leaves the stakes unspecified. We can remedy this by setting a fixed prize on a successful bet and taking the value of this bet divided by the stakes as the degree of belief. The value of the bet is the price of the bet at which the subject would be indifferent whether he accepted the bet, or left it. We will not concern ourselves with the practical difficulties involved in eliciting the precise price at which the subject would be indifferent. The measure is a *conceptual* measure, not an *operational* measure. We presume that there is a price or range of prices at which the agent would or should be indifferent whether to take or leave the bet. The agent’s degree of belief will then equal the quotient: price/prize.

There are several strong reasons to use this measure as a conceptual measure for degree of belief.

Firstly, the betting odds measure is the “only show in town”. There is no alternative conceptual measure of degree of belief that allows for a quantitative measure. Seeming alternatives like introspectible strength belief feeling are not amenable to inter-subjective quantification of any testable reliability.

Secondly, the measure gets directly at the *causal efficacy* of the belief. Therefore it lays a claim to be a measure of the essential property of belief, which is its causal role in guiding deliberative action. This means that the measure will apply to any deliberating agent, regardless of the idiosyncrasies of its cognitive chemistry.

Thirdly, the measure has specific *normative consequences*. Once every available action has been assigned an expected value based on the degrees of belief in and the values of all the various outcomes, then the agent *ought* to perform the action with the highest expected value. Normative theories of evidence can then be compared and tested in terms of their performance in maximising whatever parameter of value is under consideration.

Finally, the measure screens out insincerity and epiphenomenal cultural effects. This is very important in the field of scientific endeavour. Statements of belief given in terms of odds probabilities can be put to the test. You can lie with words, and use social intimidation to crush opposing voices. But you can’t consistently lie with your actions. To put your money where your mouth is has to be the best proof of your conviction, because this way, it is you who pays the price of false belief.

### 3.2 Ramsey’s formulation

Ramsey formalised this age old measure of belief so that he could demonstrate that the basic laws of probability follow from the simple axioms of preference. He defined degree of belief in the following passage:

“If the option of A for certain is indifferent with that of B if *p*is true and C if *p*is false, we can define the subject's degree of belief in *p*as the ratio of the difference between A and C to that between B and C; *which we must suppose the same for all A, B's and C's that satisfy the conditions*.” [My italics] (Ramsey, 1926)

For present purposes we will endorse this measure of degree of belief with two reservations.

Firstly, in Ramsey’s paper, the values “A” “B” and “C” refer to values attached to possible courses of the world in a scale that Ramsey devised. It is a measure of desirability of possible future courses of the world derived from the hypothetical preferences of the agent. However, we do not accept this measure of value. Instead we take Socrates’ line and remain non committal about the best way to measure value. We assume that in certain cases value can be measured easily by some regenerative resource like money, population or energy.

Secondly, we do not accept the italicised final clause. This is the main issue. The final clause we name the ***Supposition of Stake Size Invariance***. The magnitude of difference “B – C” is the stake size of the bet. “A” is the expected value of the bet. Ramsey supposes that whatever the magnitude of difference in outcomes of the gamble, the value for price/prize will always remain the same. He must suppose this for his system of value to work. Since Ramsey’s representational theory is so excellent, this is a prima facie reason to accept Stake Size Invariance, which is a formidable objection to our thesis. We hope to show that the supposition of Stake Size Invariance is not essential to the betting measure of degree of belief, and is not necessary to provide consistency arguments for the laws of probability, in which case it can be abandoned. Stake Size Invariance, we argue, is a substantive empirical claim that is held without argument and, without Ramsey’s theory of value measurement, is observably false.

### 3.3 Value odds not money odds

The betting odds measure of degree of belief has some immediate difficulties when it is given its most literal interpretation. The subject may accept bets out of a love of gambling or refuse bets out of a hatred of gambling. The subject also may despise money and be happy with his current lot which would make him refuse bets at good odds. In general, the subject will accept or reject a bet because of the relative *desirability* of winning and losing to the subject, not the relative *market price*. It is always possible that the subject will value the bet in terms of their own personal values for the outcomes, which may differ from the market price of the outcomes. It is the odds in terms of the personal system of value that will measure the subject’s degree of belief, not the odds in terms of market price.[[12]](#footnote-12)

The use of cash equivalent, or willingness to pay, to measure personal value is insufficiently general since many important decisions have outcomes not immediately measurable in financial terms. Furthermore there is reason to suppose that deliberating agents such as animals and small children could have partial beliefs, without being able to understand the value of money at all. But if we move to any other metric of value the problems get worse. Money has the advantage of being transferable and non perishable, although even money loses value over time. Other quantities of value, like time, energy, food etc. are not so easily storable and transferable.

The general solution to problems of this kind is to think of the degree of belief as calculating expectation in terms of value odds, rather than money odds, where value is a measure of whatever it is that the subject ultimately desires.

### 3.4 Stake Size Invariance

Merely stating that the measure of degree of belief attaches to value odds rather than money odds does not immediately solve the problem until it is specified exactly what value is, and how to measure it quantitatively. Ramsey conceived of value as a measure of the relative desirability to the agent of possible courses of the world, with the most desirable course of the world having the highest value. Thus it is true by virtue of the way that value is defined that any course of the world of higher value is preferable to any course of the world with lower value. This gives us an ordinal scale.

But in order to obtain a price/prize ratio that measures degree of belief, we need an interval scale. An interval scale is where the intervals are preserved through transformations.

To obtain an interval scale of goods and bads, Ramsey assumed that a degree of belief in an ethically neutral proposition would remain invariant over changes of stakes. An ethically neutral proposition *p* is one such that any two courses of the world identical in all respects, other than the truth of *p*, are also identical in terms of value. Ramsey could then obtain an interval scale by comparing gambles of different stake sizes over ethically neutral propositions.

To obtain an interval scale one must give a general meaning to the statement A – B = C – D. To do this Ramsey assumed two axioms: Stake Inversion Invariance and Stake Size Invariance.[[13]](#footnote-13)

**1. Stake Inversion Invariance.**

Ramsey assumed that an agent’s degree of belief would remain invariant when the stakes were inverted. This is to assume that one necessarily has the same degree of belief in *p* relative to a prospect when one is betting *for p* as when one is betting *against p*. This assumption allowed Ramsey to define a degree of belief ½ in an ethically neutral proposition. Suppose the agent to prefer B to A. Let *p* refer to an ethically neutral proposition. If the agent is indifferent between:

Gamble 1) A if *p* and B if *~p*

Gamble 2) B if *p* and A if *~p*,

then Stake Inversion Invariance allows us to deduce that the agent’s degree of belief in *p* is equal to ½.

Without Stake Inversion Invariance, all that follows is that the degree of belief in *p* in gamble 1 is equal to the degree of belief in *~p* in gamble 2 and *vice versa*.

Counterexamples to Stake Inversion Invariance are not hard to find. For example, if the proposition *p* was true iff the subject chooses Gamble 1), then the subject would be indifferent between Gamble 1) and Gamble 2), while having a degree of belief 1 that *p* relative to Gamble 1) and a degree of belief 0 that *p* relative to Gamble 2).

It might be objected that *p* is not ethically neutral in this case. There are two counters to this objection:

1) If *p* is not ethically neutral then it is because of the difference that *p* makes to the prize. In which case, no proposition that is being gambled on is ethically neutral.

2) Even if we concede that *p* is not ethically neutral in this case, the example still shows that Stake Inversion Invariance does not necessarily apply to propositions that are not ethically neutral.

**2. Stake Size Invariance.**

Ramsey then assumed that an agent’s degree of belief would remain invariant over increases in the stake size. This allowed Ramsey to create a scale of personal value that preserved intervals. Given the stake size invariant degree of belief ½ in *p,* Ramsey obtained an interval scale by finding D and C such that the agent is indifferent between:

3. A if *p*, B if *~p*

4. C if *p*, D if *~p*.

In this case the interval A – D is equal to the interval B – C.

Thus Ramsey established a conceptual measure resulting in an interval scale of value based on the assumptions of Stake Inversion Invariance and Stake Size Invariance.

However, there is good reason to suppose that degrees of belief vary with stake size, and stake inversion, and that the two assumptions of invariance necessary to obtain Ramsey’s measure of value are empirically false. Ramsey’s system of value measurement, while admirable, *presupposes* Stake Size Invariance and Stake Inversion Invariance, so cannot provide an argument for them. The problem with assuming Stake Size Invariance is that it means there is no way to distinguish *genuine* changes in marginal utility from changes in degrees of belief due to stake size. While it is quite obviously true that we sometimes will decline bets at goods odds because the price/prize ratio in terms of personal desirability is lower than the price/prize ratio in terms of money, it is also true that a change of our personal stake size in an issue can affect our degrees of belief.

Kant provides an example:

“It often happens that someone propounds his views with such positive and uncompromising assurance that he seems to have set aside all possibility of error. A bet disconcerts him. Sometimes it turns out that he has a conviction which can be estimated at one ducat but not at ten.” (Kant, 1961, A825 B853 p. 648)

In this example, the person is changing their *degree of belief*. In the passage from which this quotation is extracted, Kant uses “conviction” to refer to inductive certainty. The person here described is *certain* at low stakes, but *not certain* at high stakes. This makes him willing to bet at low stakes while reluctant to bet at high stakes at the same odds. The person’s *behaviour* could, admittedly, be explained in terms of the diminishing marginal utility of money. But, according to Kant, the person at low stakes “seems to have set aside all possibility of error.” This is a purely epistemic statement. It is not that an extra ten ducats means very little to him relative to the loss of ten ducats. It is that the stimulus of this increase in stake size alters the range of possibilities he is prepared to consider as *epistemically possible*.

I am quoting Kant here, because Kant preceded Ramsey and did not conjure this example up in order to undermine Ramsey’s theory of value measurement. Kant was merely presenting an example of what he took to be an observable truth about human practical beliefs in contingent facts: that practical certainty is sensitive to differences in interests in the proposition.

However, it still remains theoretically consistent to insist on Stake Size and Stake Inversion Invariance, and put such commonplace behaviour down to an underlying non linear value function. But Ramsey’s system as a whole has faced more precise and technically fatal objections. Later in this chapter we will look at the Allais preferences, the Ellsberg preferences, and Kahneman and Tversky’s reflection effect, all of which can be explained by stake variant degrees of belief, but cannot be explained by value functions and non linear marginal utility. But first we will defend stake size variation against Dutch book arguments for Stake Size Invariance.

### 3.5 Dutch book arguments

Ramsey proved that, given his representation theory, it follows from the basic axioms of preference that degrees of belief must conform to the laws of probability or the subject will make inconsistent preferences. He commented that if a person’s beliefs violated these laws, then her choice of options would depend on the precise form in which the options were presented, which would be absurd. He then said that a cunning bettor could make a book against the agent which he would lose in any case. This has been subsequently termed the Dutch book argument for probabilism. Probabilism is the theory that degrees of belief ought to conform to the laws of probability. Ramsey’s Dutch Book argument is an argument that degrees of belief measured in terms of betting odds must conform to the laws of probability on the pain of inconsistency. But Ramsey assumed Stake Size Invariance and Stake Inversion Invariance. Are these assumptions necessary for the Dutch book argument? The answer is a qualified no. It is possible to prove that there are some probabilistic constraints on degrees of belief from Dutch book arguments that do not rely on Stake Size Invariance.

We demonstrated in chapter 2 that the three laws of probability below follow from the axioms of preference given any additive measure of value supposing only that any higher value is preferable to any lower value, and that two gambles with the same pay offs for the same outcomes have the same value.

1. db(*p* OR *q*)U = db(*p*)U + db(*q*)U – db(*p* AND *q*)U

2. db(*p*)U = 1 - db(*~p*)­­-U

3. db (*p* AND *q*)U = db(*p*)db(*q given p*)U db(*q* given *p*)U.

We also proved the following constraints on cross stake size degree of belief:

MINIMUM. If S has P(*p*)U = x, then S has P (*p*)V ≥ Ux/V whenever V > U

MAXIMUM. If S has P(*p*)U = x, then S has P(*p*)V ≤ x whenever V > U.

So the assumption of Stake Size Invariance and Stake Inversion Invariance are not necessary to show that degrees of belief are constrained by some probabilistic laws.

A stronger objection to the Stake Size Variation Principle is that stake variant degrees of belief are probabilistically incoherent, and thus are vulnerable to Dutch books. For this objection to carry, it must be proven that there is a Dutch book argument for Stake Size Invariance and for Stake Inversion Invariance.

Arguments of this kind are not hard to find. But they depend on a method of measuring degrees of belief which differs subtly from the one we are using. This method involves naming the price for a bet at which You[[14]](#footnote-14) would be indifferent whether You sold it at that price, or bought it at that price. The method requires two players and it is brilliantly explained by Skyrms (1984).

“What do You consider a fair price to pay for a wager which pays $1 if *p* is true; nothing otherwise? To keep things honest, we can rely on the wisdom of Solomon: You set the price, but I decide whether you buy the bet from me or I buy the bet from You. The price You judge to be fair for this bet we will take as Your personal probability for *p*.” (Skyrms, 1984, p. 21)

Henceforth, we shall call degrees of belief measured in this way “Personal Probabilities”.

If we use this method of eliciting personal probabilities then Stake Inversion Invariance follows immediately. The stakes in Skyrms’ passage are $1. Using $1 simplifies the exposition because at $1 the personal probability price/prize = price, so there is no need to divide by the stakes. It follows from Stake Inversion Invariance that if your personal probability at +$1 stakes is different from your personal probability at -$1 stakes, then you will be vulnerable to a Dutch Book. This is of course the case! This is because a single price for buying and selling a bet at $1 is identical to a single price for selling and buying a bet at - $1. Selling a promise of $1 if *p* for $0.5 is identical to buying a promise of –$1 if *p* for - $0.5.

We will run through an example of how a stake size variable bid-ask spread of 0.6 around a mid-point of $0.5 can be Dutch booked. Suppose you declare 0.2 at $1 and –0.8 at -$1. This means that to “buy” a bet at -$1, You “pay” – $0.8. Paying a negative sum of money is equivalent to receiving a positive sum. This entitles You to win -$1 if *p* is true. Winning a negative sum of money is winning a debt, so it means You must pay $1 if *p* is true. So this pair of prices is equivalent to a “bid” price of $0.2and an “ask” price of $0.8. A bid price of $0.2 is the price You would be prepared to *buy* a $1 bet on *p*; an ask price of $0.8 is the price you would be prepared to *sell* a $1 bet on *p*. Skyrms simply buys the positive bet at $0.2 and “buys” the negative bet at – $0.8, making a net gain of $0.6 for Skyrms, and a net loss of $0.6 for you. Since he now has a positive bet and a negative bet on *p*, then if *p,* Skyrms pays You $1 for the negative bet, and you pay Skyrms $1 for the positive bet. If *~p* then no one owes anyone anything. So whatever happens You lose $0.6, which is your db(*p*)-U - db(*p*)U. This is equal to your bid – ask spread.

It should be clear that this Dutch book argument is a scam. All You have done is give a cautious bid-ask spread of $0.6. Skyrms has bought a bet off You at Your bid price, and sold it back to You at Your ask price (which is equivalent to buying it from You at your negative stakes price). So of course You have lost money. If it became the general law that people should sell at their own bid price and buy at their own ask price, then all commercial currency dealers and book makers would go broke over night.

A similar Dutch book could be made against stake size variant degrees of belief. Suppose You had a low stakes $1 personal probability of 0.5 that *p* and a high stakes $100 personal probability of 0.3 that *p*. Skyrms could buy a single high stakes bet at the lower price of $30, and sell it back to You piece meal as one hundred low stakes bets at a per bet price of $0.5. Skyrms would thus make a zero risk gain of $20. As long as Skyrms is allowed to buy or sell as many bets as he wants at the same price he can make arbitrage off any stake size variation.

This Dutch book argument is just as much a scam as the previous argument. It relies on the extra rule that you must buy and sell multiple bets at the same price. Such a practice, were it to be made law, would make supermarkets go broke over night. You could simply fill up your basket with “two for the price of one” offers, and demand a refund at the face value, doubling your money at the supermarket’s expense. And poor old banks would go into melt down. You could simply borrow billions of pounds at their current account interest rate of 2% and then lend it back at their savings account rate of 6% and drain the entire economy. If only!

These Dutch book arguments work because Personal Probabilities so defined have Stake Size Invariance built in. You are required to give a price at which you are both prepared to buy and sell. A defender of Personal Probabilities might say that this is an advantage, because this forces You into a situation where You give a unique point probability that is Stake Size Invariant. This means that each event has a unique point probability from any evidential perspective. They might say that convention is on their side and it fits in better with the way people think about probability to measure degrees of belief in this way. Furthermore, if we do measure beliefs in this way, then we have the advantage of being able to use Ramsey’s system of value measurement to explain the phenomena of stake size variation.

But the bigger picture falls apart. For Personal Probabilities are Stake Size Invariant *on money odds*. They are Stake Size Invariant because they are defined in such a way that they have to be. So, according to Ramsey’s system of value measurement, then value will *necessarily* be linearly related to money, or the agent will be vulnerable to the Dutch books given above.

This is because, for any degree of belief above ½, and any non linear value function that gives a diminishing marginal utility to money, then the asks are more attractive than the bids, and so the agent wishing to maximise her expected value will adjust her price accordingly. The fact that the expected value of money bets at the same odds with the same money expectation varies with stake size is the mechanism by which Ramsey elicits his interval scale of value. But it also means that, by the rules of Skyrms’[[15]](#footnote-15) game above, the agent with a non linear value function can be Dutch booked.

For example, suppose there was an objective probability of 0.9 that *p.* (If there is a problem with the interpretation of objective probability, let us say that there is a lottery with 9 tickets marked “*p”* and 1 ticket marked “*~p*”, and further assume that the ticket is drawn in such a way that each ticket has an equal probability of being drawn.) Julie has just £10 in all the world and has a logarithmic value function to money. Skyrms proposes that they use the wisdom of Solomon to measure each others’ personal probability that *p*. Skyrms proposes stakes of £10 and declares his fair price for this bet at £9 on *p*. This makes Skyrms’ personal probability that *p* equal to £9/£10 which is equal to 0.9, the objective probability. In assessing Skyrms’ personal probability that *p* at stakes £10Julie finds that she is given the choice of what side of this bet to take:

1. Bid: Lose £9 if not *p*, Gain £1 if *p.*

2. Ask: Gain £9 if not *p*, Lose £1 if *p*

Using her value function she calculates the expectation on each option by multiplying the objective probability of each possibility by the logarithm of her end state wealth and taking the sum.

Expected value. Bid = log £1 x 0.1 + log £11 x 0.9 = 0.937 = log £8.655

Expected value. Ask = log £19 x 0.1 + log £9 x 0.9 = 0.9867 =log £9.699.

So at supposedly fair odds, Julie finds the Ask option much more attractive than the Bid option. In other words, at these odds, she would much prefer to sell a bet on *p* than buy a bet on *p*. Furthermore, since both have an expectation of less than log £10, she would prefer the option of walking away from Skyrms altogether to either buying or selling. This means that, in a sense, she loses out in expectation just by accepting Skyrms’ rules of play. This ought to signal that there is something unfair about this particular Dutch book from the outset.

Now let us suppose it is Julie’s turn to set the price. She can increase her expectation by lowering the price. A lower price would mean that Skyrms would to choose to bid. Let us suppose she lowers it to £8.5, making her personal probability 0.85. Her expectation then becomes:

Exp. Bid = log £1.5 x 0.1 + log £11.5 x 0.9 = 0.9722 = log £9.38

Exp. Ask = log £18.5 x 0.1 + log £8.5 x 0.9 = 0.9632 = log £9.188

Now the Bid option is more attractive, the values are closer together and the average expectation is higher. We can see that Julie, given time and a calculator, would set her personal probability at somewhere between 0.85 and 0.9. Given different stakes, she would set her personal probability at a different point. The smaller the stakes, the closer to 0.9 she would set her personal probability. This is on the basis that she has a Stake Invariant degree of belief equal to the objective probability; has a logarithmic value function; and is maximising her expected value.

Skyrms can get her into a Dutch book by getting her to set one price for £10 bets and another price for £1 bets. She will set the stakes £1 price at closer to 0.9 than the stakes £10 price, and Skyrms will be able to profit from the arbitrage by buying 1 bet at the £10 price and selling 10 bets at the £1 price.

What this shows is that the only Dutch Books that can provide an argument for Stake Size Invariance are too powerful to be plausible, since they also prove that any non linear value function to money results in vulnerability to Dutch books. Furthermore, Dutch books that prove Stake Size Invariance also prove that the majority of commercial asset pricing is probabilistically incoherent.

A further interesting revelation is that if we assume that an agent has a degree of belief x that *p* and has a non linear value function to money; then the agent’s personal probability measured by the fair odds they declare will be slightly less than x and therefore not equal their degree of belief.

This leads to a dilemma for the supporters of Stake Size Invariance: If we assume Ramsey’s system of personal value measurement, then we cannot use Dutch Books to establish Stake Size Invariance. But if we establish Stake Size Invariance using Dutch Book arguments, then we cannot use Ramsey’s system of personal value measurement.[[16]](#footnote-16)

In conclusion, it cannot be proven using Dutch book arguments that degrees of belief ought to be Stake Size Invariant.

### 3.6 The conjunction rule, conditionalization and stake size escalation.

In chapter two we demonstrated the identity of bets we called the Conjunction rule:

CONJUNCTION: db (C & A)U = db(A)db(C | A)U db(C | A)U.

This followed from Ramsey’s definition of conditional degree of belief, with the added concept of a multiplier bet, which is where the entire stakes of the first bet are placed on the second bet in case the first bet wins. In this case, the multiplier bet on the Antecedent multiplied by a conditional bet on the Consequent given the Antecedent wins just in case the conjunction A & C is true and loses just in case the conjunction A & C is false.

We have noticed that this identity involves what we call ***stake size escalation*.** In other words, the stakes on the conjunction are greater than the stakes on the antecedent by a factor of 1/db(A | C)U. This reflects the fact that a conjunction has more empirical content and therefore constitutes more valuable knowledge. The evidence necessary to settle a bet on a conjunction is therefore of greater value than the sum of evidence necessary to settle bets on the conjuncts.

Under the assumption of Stake Size Invariance, there is a Dutch Book argument that degrees of belief ought to conform to the conjunction rule at all stake sizes. De Finetti provided just such a Dutch Book argument based on the two player game above. This allows us to derive the standard Kolmogorov ratio definition of conditional probability:

P(C | A) = P(A & C) / P(A).

We now have a concept of central importance to Bayesianism: the concept of conditionalization on the evidence.

***Conditionalization***: If you have a prior degree of belief P(H | E) = X, and you become certain that E through observation; then you should update your degree of belief so that P(H) = X.

However, Ramsey, who many cite as the author of contemporary Bayesianism, explicitly denies the Conditionalization rule.

“We are also able to define a very useful new idea -- the 'degree of belief in *p* given *q*'. This does not mean the degree of belief in ' If *p* then *q* '*,* or that in '*p* entails *q* ', *or that which the subject would have in p if he knew q, or that which he ought to have*. It roughly expresses the odds at which he would now bet on *p*, the bet only to be valid if *q* is true. Such conditional bets were often made in the eighteenth century.” [My italics] (Ramsey, 1926 p. 21)

“This is not the same as the degree to which he would believe *p*, if he believed *q* for certain; for knowledge of *q* might for psychological reasons profoundly alter his whole system of beliefs.” (Ramsey, 1926 p. 21)

The difference between the conjunction rule, and the conditionalization rule is that the conjunction rule can be argued for by a Synchronic Dutch book as we have shown above. And if we include stake size escalation, then this identity can proven to hold at escalated stakes without the assumption of Stake Size Invariance. Whereas the conditionalization rule can only be argued for with recourse to a Diachronic Dutch book.

Diachronic Dutch books are more controversial than synchronic Dutch books. Skyrms (1987) provided a Dutch book argument, which is attributed to David Lewis, which showed that if your degrees of belief were diachronically incoherent, then a cunning bettor will be able to buy and sell bets from you at your posted prices that will guarantee a net loss in all eventualities. He also provided the converse argument that shows that if your degrees of belief *are* diachronically coherent, then you *cannot* be Dutch booked.

Both the synchronic and diachronic Dutch book arguments rely on the two player game version of Personal Probability. This means that you have to post a unique price at which you must either buy or sell bets whatever the stakes. However, while the synchronic Dutch books do successfully dramatize the conjunction rule at escalated stakes; diachronic Dutch books fail to dramatize the conditionalization rule.

The reason they fail is informative about the nature of stake variant degrees of belief so is worth examining. The Dutch book involves a Bookie, who posts odds, and a Cunning Bettor, who places a series of bets at the Bookies odds, which taken together form a Dutch book, such that the Bookie loses money in every possible outcome. The game structure is that the Bookie, who is the subject of the personal probabilities, posts a synchronically coherent probability distribution for t1 including Pt1(~E), andPt1(*p* | E) and an updating rule that gives the Bookies new probability distribution at t2 when it is known whether or not E. The Dutch book conditionalization theorem is that any updating rule that results in Pt2(*p*) ≠ Pt1(*p* | E) will be vulnerable to a Dutch book.

**Proof.**

Either (1) Pt1(*p* | E) < Pt2(*p*)

Or (2) Pt1(*p* | E) > Pt2(*p*)

If (1) then let *d* = Pt2(*p*) - Pt1(*p* | E). The Cunning Bettor bets as follows:

At t1 he proposes

i) A bet on t1(*p* | E) at stakes U. This means he pays Pt1(*p* | E)U if (E &*~p*)*,* wins U - Pt1(*p* | E)U if (E &*p*) and the bet is cancelled if *~*E.

ii) a bet on Pt1 (~E) at stakes dU, meaning he pays Pt1(~E)dU if E and wins dU – Pt1(~E)dU if ~E.

At t2 if ~E then the cunning bettor wins a net of dU – Pt1(~E)dU,

iii) if E he then bets on *p* at – U at the new odds Pt2(*p*). This last bet is at negative stakes so that the Cunning Bettor gains the price Pt2(*p*)U if *~p* and has to pay U - Pt2(*p*)U if *p*.

|  |  |  |  |
| --- | --- | --- | --- |
| BET | Pay Off ~E | Pay Off E&*p* | Pay Off E&*~p* |
| i) t1(*p* | E)U | NA | U - t1(*p* | E)U | - t1(*p* | E)U |
| ii) Pt1 (~E)dU | dU - Pt1 (~E)dU | - Pt1 (~E)dU | - Pt1 (~E)dU |
| iii) Pt2(*p*)–U | NA | Pt2(*p*)U – U | + Pt2(*p*)U |
| Net Total | dU - Pt1 (~E)dU | dU - Pt1 (~E)dU | dU - Pt1 (~E)dU |

As we see from the last line, the Cunning Bettor wins dU - Pt1 (~E)dU in all cases other than when the bookie sets Pt1(~E) = 1.

In case (2), *d* becomes negative, so the Dutch book works by simply inverting the stakes, in other words by substituting all instances of –U with +U and *d* with ­–*d*.

Now, the remarkable thing about this Dutch book argument is that the structure of the game doesn’t allow the Stake Size Variation Principle as an updating strategy. The cunning bettor can bet how he likes at t1 knowing that this will not affect the bookies prices at t2. But, because of the nature of conditional bets, the stakes on Pt2(*p*) vary according to, not only how the cunning bettor bets at t2, but also how he conditionally bets at t1 according to the stake escalation phenomena. Many criticisms in the literature of diachronic Dutch books point out that the bookie ought to be able to “see the cunning bettor coming” (Levi, 1987), (Maher, 1992). The stake size variation principle we laid out at the beginning of this chapter has a built in immunity to this diachronic Dutch book precisely because the bookies Pt2(*p*) will be sensitive to the stakes on *p* and, via the stake size escalation principle, to the stakes on E. These stakes will in part depend on the way the cunning bettor bets, so the agent following the stake size variation principle will automatically *see the Dutch book coming* and avoid it.

In case (1) when the Cunning Bettor hoped at t1 that (1) Pt1(*p* | E) < Pt2(*p*), he attempted to get the bookie into a Dutch book by placing i) a bet on Pt1(*p* | E) at stakes U ii) a bet on ~E at stakes dU, and iii) a bet on Pt2(*p*)at stakes –U.

Now, since the Bookie takes the other side of the bet from the Cunning Bettor, all the stake sizes are inverted from the Bookies point of view. So the Bookies bet on i) t1(*p* | E) is at stakes – U, and the Bookies bet on iii) P t2 (*p*) is at +U. According to the stake size variation principle, then this means that the Pt1(*p* | E) ≥ Pt2(*p*) since the degree of belief at negative stakes must be greater than or equal to degrees of belief at positive stakes. So if the bookies updating rule conformed to the Stake Size Variation Principle, then, as soon as the Cunning Bettor placed his bets at t1, the Bookie would respond by updating his degree of belief in P­t2(*p*) so that it was less than or equal to Pt1(*p* | E) in conformity with the conjunction rule and MAXIMUM (If S has P(*p*)U = x, then S has P(*p*)V ≤ x whenever V > U) and thereby prevent the Dutch Book.

The inverse case (2) is similar. If the Cunning Bettor places bets at t1 under the assumption that the bookies Pt1(*p* | E) > Pt2(*p*), then the bookie will respond by updating so that Pt1(*p* | E) ≤ Pt2(*p*).

What is especially pleasing about this result is that the Bookie who follows the Stake Size Variation principle will naturally set odds that not only avoid diachronic Dutch books themselves, but will tend to profit from bettors whose diachronic betting odds do not obey the stake size variation principle. We can see that the Cunning Bettor in the original diachronic Dutch book is actually behaving as if she had Stake Size Variant degrees of belief.

We ought to observe at this point that many dealers in commercial risky assets, for example race track book makers, will post odds that are Stake Size Variant in this way. The aim of the commercial book maker is to balance the books and profit from the arbitrage present in the variation of opinion in the general public. As long as the odds are set with a bid ask spread in the bookies favour, and the volume of bets against and the volume of bets for balance, then the book maker can make a risk free profit without any commitment to any particular outcome. However, if the entire public bet one way, then the book maker is vulnerable to a considerable risk. So the common practice is to vary the odds in inverse proportion to the amount of bets sold. For example, the *favourite,* i.e. the horse with the highest negative stakes from the bookies point of view, is given the lowest odds, in other words, is accorded the highest probability from the bookies point of view. This makes the option to the general public of betting against the favourite more attractive, tempting cancelling bets that will balance the books.

We can also see how stake size escalation works in a commercial setting. If many people had high stakes multiplier bets on the results of two consecutive races, say Black Beauty to win the first and Red Rum to win the second, but there weren’t any balancing bets, then the odds on Red Rum winning the second race would swing heavily depending on whether Black Beauty won the first race. If Black Beauty won the first race, then the Stakes on Red Rum would be very high, so the bookie would lower the odds to attract cancelling bets. If Black Beauty lost the first race, on the other hand, then there would be no stakes on Red Rum and the Bookie would base his odds on the market. So it is the standard practice of commercial bookmakers to vary the betting odds in accordance with the Stake Size Variation principle.

### 3.7 Gambler’s Ruin

But is it possible to provide some kind of Dutch book argument for conforming to the Stake Size Variation Principle as opposed to having Stake Size Invariant beliefs? I think it is. The argument I will give I will call ***Gambler’s Ruin.*** The argument is based on the law of large numbers and the relationship between probability and frequency.

The argument is that given advantageous odds, the Invariant Bettor will maximise expectation by betting at the largest possible stakes. This will tend to lead to “ruin”, meaning that the Invariant Bettor’s holdings will almost certainly be wiped out after a very short run. This has the added effect that the mean win/lose ratio will tend to be lower as the stakes get higher. The Stake Size Variant Bettor, on the other hand, will be able to sustain a much longer run and therefore tend to achieve long run win/lose ratios closer to the objective probability. So although the Invariant bettor has a higher expectation on any particular bet; in the long run the Invariant bettor is almost certain to be ruined with a poor mean win/lose ratio; whereas in the long run, the Variant bettor can safely expect a steady rate of growth. This Dutch book argument is admittedly different from the usual Dutch Book argument which involves a single “book” that involves a “sure loss”. This Dutch book argument involves a long run betting strategy that will lead to almost certain ruin. The long run betting strategy has the agent only betting at positive expectation. No particular bet will involve a sure loss, but ruin is certain in the long run. A simple example of this is “triple or quits.” An invariant bettor with coherent subjective probabilities who conforms to the expectation principle and has bel(0.5) that a coin will land heads can be ruined by offering him a bet where he triples his money each time the coin lands heads, and every time he bets he is offered the same bet again at triple the stakes. The invariant bettor will continue to play this game until he is ruined, since every bet will have a positive expectation. The certainty of his ruin is not absolute certainty but *almost* certainty. As the number of these bets tends toward infinity, the probability of ruin tends to 1.

The Dutch book requires a few extra assumptions. This time, the Bookie posts odds, not of bets on one off propositions, but of bets on repeatable propositional functions. The Cunning Bettors must then choose how to bet on the assumption that the antecedent of the propositional function will be repeated each time the Bettors buy a bet. This models any repeatable action over which a deliberative agent can make a decision whether to act or not. The two competing cunning bettors’ aim is to maximise long term growth. The Dutch book argument will assume that the Bookie has infinite wealth and will honour any bets at odds that are fixed from the outset. It will also assume that the cunning bettors can choose the stake size so long as it does not involve a price greater than their entire holdings. Given a scenario where the Bookie offers odds that give the bettors a slight advantage, the argument purports to show that the Invariant Bettor is almost certain to lose his entire holdings to the Bookie in spite of the advantage; whereas the Variant Bettor is certain to at least avoid ruin, and is highly likely to make a considerable profit at the Bookie’s expense. Since it is rational to prefer a high probability of gain to certain ruin, then this argument should show that it is rational to adopt the Stake Size Variation Principle.

The Bookie posts coherent prices this time, with no arbitrage so that B (*p* | E) = ½ at all stake sizes forever. The Bettors are given £100 start up and have the knowledge that K*p*/K = 6/10 and K = £1000 with respect to *p*| E. We will assume that the Invariant bettor has an Invariant degree of belief equal to K*p*/K that *p* | E, in other words P(*p* | E) = 0.6. E is a repeatable event under the Bookie’s control that takes place only if a bet is made. There are then only two possible outcomes for each bet Ei: either *p*i or *~p*i. The number of bets N is therefore equal to the number of occurrences of E. If all bets are on *p,* then call the number of wins “W”. We shall also assume that the *objective probability* of *p* | E = 0.6. This has the causal consequence that as N → ∞ then W/N → 0.6. This is short hand for the more comprehensive statement that for any N, standard probability calculus can deduce a probability distribution for W/N such that W/N is increasingly likely to be within diminishing margins of 0.6 as N gets larger, and that the mean value for W/N in this distribution is 0.6.

In these conditions, the Invariant Bettor, to maximise her expectation, should bet her entire capital on *p* and expect a profit of 20% per bet. We can then use standard probability calculus to calculate the probability of going broke at each value of E = N. For each value of E = N where the Invariant bettor goes broke, there is a unique value for W/N. We can then work out the average actual rate of growth by taking the sum of the products of probability of the number of bets E = N by the frequency W/N at E = N.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| N | P(E = N)  0.4x0.6N-1 | Running total;  P(E ≤ N) | Frequency W/N given  E = N | Product  W/N. P(E=N) | Total:  Expected W/N  given (E ≤ N) |
| 1 | .4 | 0.4 | 0 | 0 | 0 |
| 2 | .24 | .64 | ½ | 0.12 | .12 |
| 3 | .114 | .754 | 2/3 | 0.076 | .196 |
| 4 | .0864 | .8404 | ¾ | 0.0648 | .2608 |
| 5 | .05184 | .8994 | 4/5 | .041472 | .3023 |
| 6 | 0.031 | .9304 | 5/6 | .02583 | .32813 |
| 7 | 0.01866 | .94906 | 6/7 | .0159943 | .3441 |
| 8 | 0.0112 | .96 | 7/8 | .0098 | .3539 |
| 9 | 0.0067 | .966 | 8/9 | .0060 | .3600 |
| 10 | .004 | .9709 | 9/10 | .0036 | .3636 |
| →∞ | → 0 | → 1 | → 1 | → 0 | → < 0.4 |

What this table shows is that it is very likely that the Invariant bettor will be broke in less than ten rounds, and the mean frequency W/N over all possible games weighted by probability is less than 0.4. This shows that the expected profit of 20% per bet is *almost certain to be false*. And the true per bet return is likely to be in the region of minus10% per bet. It might seem paradoxical that the mean frequency W/N < 0.4 given that we have defined the objective probability of *p* | E as 0.6 in terms of the mean value for W/N in the probability distribution for W/N for all N. This disparity is explained because the mean frequency of < 0.4 is conditional on the stopping rule that each run only terminates when ever the bettor can no longer place bets and so the trial ends; whereas the mean frequency of 0.6 is conditional on there being no stopping rule.

The Variant Bettor, on the other hand will aim to maximise C + (B – C)(C +K*p*)/(C + B + K) on each bet. The maximum expectation on the first bet is achieved by setting C = £50, B = £150, in other words the ***optimal stake size*** is £100 with a bet price of £50. The ***optimal stake size*** is the stake size that gives the highest expectation according to the Stake Size Variation Principle. The expectation on this bet is £50 + (£100)(650)/(1200) = £104.1666, (+ 4.166%). The optimal stake size on further bets are dependent on the results of previous bets in two ways: 1) The results of previous bets will increase the value of knowledge (K) by the stake size, and thereby affect the ratio Kp/K. 2) The gains and losses from previous bets will add to the Bettor’s capital effecting the ratio (B – C)/(B + C) and thus affecting the evidential probability (*p* | E)B – C.

Both 1) and 2) will have the affect that wins will increase the optimal stake size on the next bet, and losses will decrease the optimal stake size on the next bet.

Given these factors, the Stake Size Variant Bettor will out perform the Invariant Bettor, because the Invariant Bettor will be ruined with a probability of 1 after a fairly short amount of time, whereas the Variant bettor will tend to continue to bet indefinitely, learning through accumulated experience as she goes. Another advantage for the Stake Size Variant Bettor is she can track *changes* in probability, with more recent experiences tending to have more weight than older experiences. So although the Variant Bettor has a more modest rate of growth in the short term, her degrees of belief will more accurately reflect the relevant frequencies, and her long term prospects are impervious to ruin.

In short, the Bookie will almost certainly win money from the Invariant bettor, and will almost certainly lose money to the Variant bettor.

### 3.8 Stake Size Variation Principle gives a better estimation of future frequency.

As well as out performing the Invariant bettor in terms of betting strategies, the Variant bettor out performs the Invariant bettor at straight forward predictions of frequency. Let E be some repeatable event so that the frequency of *p* in a sequence of N events of type E is written *p*/N. At zero stakes, a degree of belief ½ in (*p* | E) implies a prediction of a frequency *p*/N = ½. This means that, in ignorance of the size of N, the best estimate of the frequency of *p* / N is ½ . We can make this more precise by saying that for any N, a degree of belief *p* implies a probability distribution of all possible values for *p*/N. The sum of the product of each value *p*/N by its probability will equal ½. We can call this sum the *expected frequency* of *p* | E. For this reason, the expected frequency for a series of bets on *p* | E at even odds is always going to be ½ for an Invariant bettor with a degree of belief ½ in *p* | E.

The Variant bettor, on the other hand, will have a best estimate of the frequency of *p*/N that diminishes as the stakes go up. For this reason, a Variant bettor with a degree of belief *p*| E0 = ½, will predict that the frequency of *p* | ES diminishes in inverse proportion to the magnitude of S. In what follows we shall attempt to show that the Variant bettor’s prediction is more likely to be correct. This is under the assumption that repeatable trials will cease whenever losses exceed the total wealth. I argue that this assumption is essential to the meaning of value. If there is no limit to how much you can lose, then nothing is at stake. For stake size to have any meaning, there must be a limit to how many times one can sustain a loss of a specific stake size relative to a specific level of wealth.

Let us suppose that a coin is flipped every time a bet at even odds with a fixed stake size of 2 units is placed on tails by a bettor with a variable wealth of W units. We can thus increase W in order to decrease the significance of the stake size. The significance of the stake size relative to the wealth of the agent is 2/W. The larger W, the smaller the significance of a 2 unit stake size bet. This models a choice of action where the stakes are fixed externally to the agent and the agent merely has a choice to act or not to act. The decision whether to act involves a long term forecast of the likely success rate of the action. The betting continues until the Nth bet when bettor runs out of money. A bet with stakes 2 units at even odds means that the bettor’s wealth increases by 1 unit for each tails and decreases by 1 unit for each heads. Therefore the bettor runs out of money whenever the number of heads “H” exceeds the number of Tails “T” by W, so that H – T = W. Therefore the frequency of tails at the Nth bet, hence forth “FN(T)”, when the Nth bet is the terminal bet, is equal to (N-W)/2N. Under the assumption that the objective probability of tails = ½, there is a probability distribution for the value of N for each value of W. Call the probability that N = x, “P(Nx)”. When W > x then P(Nx) = 0. When x = W then P(Nx) = ½ W . When x > W then P(Nx) = fN(½ N) where fN depends on the number of possible combinations of H and T in N such that H – T = W where the last term is H and where there is no predecessor Ni where Hi – Ti = W. fN will be less than 2N and will be greater than or equal to fN-1 due to the law of large numbers. We can now give a general formula for the expected frequency of tails given a wealth of W and a stake size of 2, “Ex F(T)2/W”:

Ex F(T)2/W = ∑x = 1 - ∞ fxi( ½xi ) (xi – W)/(2xi)

We can see straight away that Ex F(T)2/W will be less than ½. This is because for all x, x ≠ ∞, (xi – W)/(2xi)< ½. It is also the case[[17]](#footnote-17) that as W gets larger relative to the stake size, then ExF(T)2/W tends towards ½, with ExF(T)2/W → ½ as W → ∞.

The Stake Size Variation Principle has db(*p* | E )B-C = (C + K*p*)/ (B + C + K). If K = 0, this becomes (C)/(B + C). In the above scenario C = W – 1 and B = W + 1. The degree of belief at stake size 2/W = (W – 1)/(2W). We can see that this too will always be below ½, will be lower as the stakes get higher relative to W, and will tend to ½ as W tends to ∞.

If we add values for K*p* and K, then the ratio of K/W determines the degree to which the stake size reduces the degree of belief. But however great K relative to W, the tendency will still be the same. Because we measure K in terms of the stake size on previous events, then K cannot be infinite. If K is very high relative to W, then the degree of belief will tend to be not so influenced by stake size. This means that the Variant bettor will have a degree of belief in (*p* | E)W> Ex F(*p* | E)W; although still smaller than ½ . We can justify this by saying that the judgement of the frequency of *p* | E takes into consideration the past frequency, whereas Ex F(*p* | E)W only gives the expected future frequency.

In conclusion the Variant degree of belief (C + K*p*)/(B + C + K) gives a better estimate of the frequency of *p* | EU than any Invariant degree of belief, including cases where the Invariant degree of belief matches the objective probability. The only case where both estimates are equally good is where the stake size is zero. These are cases where no action of the agent makes any difference to the length of the run, and the frequency makes no difference to the agent and so the agent’s degree of belief has no function. Any cases where there is a motive for having an accurate prediction of the frequency, then the Stake Size Variation principle gives a more accurate prediction than the objective probability.

### 3.9 Kelly Gambling

It might be thought that the arguments for Stake Size Variation here rely on the negative utility of ruin, and that an Invariant bettor could avoid ruin by never betting his entire capital but only betting a fraction of his capital L. Let us suppose that the bettor only ever risked ½ his capital on any bet. In this case, the bettor would never be ruined so long as the bettor had control of the stakes and the currency was infinitely divisible. Because the length of the run would be limitless, then the Invariant bettor would have no reason to over estimate the frequency.

Now, it is true that this strategy would save the Invariant bettor from ruin, but it would not save the Invariant bettor from being out performed by the Variant bettor. In fact, it can be argued that the Invariant bettor would still be in effect ruined, because his capital would converge on zero as the run tended to infinity.

As we discussed in chapter 2, Kelly JR (1956) wrote a paper which calculated the exact proportion of one’s capital than one should gamble when one has an “edge”[[18]](#footnote-18). The “edge” is the difference between one’s own degree of belief, and the odds offered by the bookie. The Invariant bettor will always bet his entire capital if they have any edge whatsoever, but he will naturally prefer larger edges. The Kelly gambler, in contrast, will bet a proportion L of her capital W, where the proportion of the capital gets larger in direct proportion with the edge according to the formula L = edge/odds.

Kelly wrote his paper in 1956, with the help of Shannon. Since then, Kelly Gambling has been demonstrated to out perform any other betting system in terms of growth. The Kelly bettor is behaving as if he had stake variant degrees of belief, since the Kelly bettor will accord bets with a higher stake size a lower expectation than bets at a lower stake size at the same odds. If degrees of belief are measured in terms of the expectation of bets, then this means that the Kelly bettor has Stake Size Variant degrees of belief. As we discussed in Chapter 2, the Kelly Gambler is not exactly the same as an SSVP gambler, but were we to assign degrees of belief to the Kelly gambler using Ramsey’s formula, then we would find that the Kelly Gamblers degrees of belief varied with the stakes in the same direction, except that L would be linear with regard to W for a Kelly gambler, whereas for any fixed k*p*/K, then L would tend to diminish as C increased relative to K. (It is a little more complicated than that because as C gets higher relative to K, then the probability will tend towards C/(B + NC) for N mutually exclusive propositions, rather than tending towards 0, so the tendency for L to diminish as B and C increase relative to K only applies when o > C/(B + NC) and when k*p*/K ≥ o).

## 4. Objection: Stake size variation adequately accounted for by non linear utility attached to risky outcomes.

Perhaps the obvious objection to the Stake Size Variation Principle is that all the preferences explained using the Stake Size Variation Principle are adequately explained with recourse to *utility theory*. Utility theory is simply the general term for any theory that attempts to give a measure to the subjective value of various goods. Money and the willingness to pay will not do because it is possible that some people put a lot more value on money itself than other people. It also is fairly obvious that the value most people attach to money and other goods is not linearly related to the quantity. Two portions of chips is not twice as valuable as one portion of chips to a single diner. A windfall of a million pounds will dramatically change my life, but a second windfall of a million pounds will not have so great an impact. All this is undoubtedly true and is not in any way incompatible with the Stake Size Variation Principle. The Stake Size Variation Principle applies to the utility of outcomes, not to the cash value.

But utility can also be applied to risk itself. Some people may like gambling so much that they are prepared to play at disadvantageous odds. Other people may turn down advantageous odds through a hatred or a moral disapproval of gambling. At this point the waters become very muddy, since there is no operational measure that can distinguish between a change of degree of belief in winning due to stake size variation, and a change in subjective value of the prospect due to the risk, since both affect the odds at which an agent will be indifferent without affecting the utility score of the outcomes.

### 4.1 Bernoulli’s formula for calculating expectation.

Utility theory as an explanation for anomalous preferences due to risk has a long history which can be traced to Daniel Bernoulli’s 1738 paper *Exposition of a new theory of risk measurement* (Republished in English translation in *Econometrica*, Vol. 22, No. 1. (Jan., 1954), pp. 23-36.) Bernoulli distinguished between the price of a risky prospect, which is the same to everyone, and its expected utility, which can vary from individual to individual. Observing that risky prospects have a higher expected utility to the rich than the poor, Bernoulli suggested that people cannot be multiplying the chances by the gains in order to derive the expected utility. He then provided an alternative calculation of expected utility taking a logarithmic function of the gains, chances and the subject’s holdings. This has the consequence that the marginal expected utility diminished as the gains get larger relative to the subject’s holdings.

We can now demonstrate that Bernoulli’s calculation of expected utility is *compatible* with stake size variation.

Bernoulli begins his paper by stating the orthodox expected utility principle.

“*Expected values are computed by multiplying each possible gain by the number of ways in which it can occur, and then dividing the sum of these products by the total number of possible cases where, in this theory, the consideration of cases which are all of the same probability is insisted upon.*”

Although the mode of expression is slightly archaic, this is identical to the orthodox expected utility principle that calculates the actuarial value from the objective chances and the utility outcomes using the standard decision theoretic formula: EU xi = (ux1, px1) + …. (uxi, pxi,) + ..... (uxn, pxn).

He then argues on the basis of some of examples of decisions under risk, that all men cannot use the same calculation when evaluating risky prospects. He offers this alternative rule:

*“Any gain must be added to the fortune previously possessed, then this sum must be raised to the power given by the number of possible ways in which the gain may be obtained; these terms should then be multiplied together. Then of this product a root must be extracted the degree of which is given by the number of all possible cases, and finally the value of the initial possessions must be subtracted therefrom; what then remains indicates the value of the risky proposition in question.”*

Ibid. p.29

So we have a way of calculating the expected utility which is sensitive to risk in the right way. We also have a formula, thanks to Ramsey, that shows us how to derive the rational degree of belief from the expected utility and the utility of the outcomes. So we can stick with the orthodox expected utility principle, and use Bernoulli’s risk measurement function to derive the expected utility, and then Ramsey’s formula to derive the risk sensitive degree of belief.

We feed the gains, the chances and the subject’s wealth into Bernoulli’s formula. This gives us the expectation. We then take Ramsey’s definition of degree of belief:

Ramsey’s degree of belief defined in terms of expectation.

An agent indifferent between:

1. A for certain.

2. B if *p* and C if ~*p*

Has degree of belief equal to (A – C)/(B – C).

It follows from the definition of expected utility that an agent will be indifferent between the expected utility of a gamble (A) and the gamble itself (B if *p* and C if ~*p*). Since we have values for A, B and C, we can calculate the stake size sensitive degree of belief. So Bernoulli’s formula provides a method for calculating stake size sensitive degrees of belief from the chances, the stake size and the subject’s wealth.

For example: Suppose Jack had a fortune of 4 ducats and was considering a prospect that would pay 1 ducat if *p* and nothing if *~p*. Let’s assume there is one way that *p* and one way that *~p,* making a total of two possible cases. To make this vivid let us suppose that *p* is true if the coin lands heads and *~p* is true if the coin lands tails. Since there is one way that *p* out of two possible ways, we can stipulate that the “chance” of *p* is ½. We can interpret this as the stake size independent probability, i.e. the limiting frequency in a hypothetically infinite series of trials. According to Bernoulli, the expectation on this bet for Jack is:

Expected utility = {2√(4 + 1)1 (4)1 } – 4 = 0.4721 ducats

This would make his degree of belief in *p* at stakes (1 – 0), given Ramsey’s formula:

(*p |* E)1 = (0.4721 – 0)/(1 – 0) = 0.4721.

And his degree of belief in *~p* at stakes (0 – 1)

(*~p* | E) –1 = (0.4721 - 1)/(0 – 1) = 0.5279

We can see here that the degree of belief at positive stakes is lower than the zero stakes chance, and the degree of belief at negative stakes is higher than the zero stakes chance. This is what we should expect given the Stake Size Variation Principle. Furthermore, the degrees of belief thus derived are clearly coherent, since the probability of (*p* v ~*p*) = 1 relative to every possible prospect.

Now suppose Jack considered a prospect with twice the stakes size, with everything else held fixed. This time the prospect pays 2 ducats if *p* and nothing if ~*p*. According to Bernoulli, the expected utility of this prospect is worth

Expected utility = 2√(4 + 2)1 (4)1 } – 4 = 0.8990 ducats

Which makes the degree of belief that *p*, given Ramsey’s formula:

(*p* | E)2 = (0.8990 – 0)/(2 – 0) = 0.4494

And the rational degree of belief that ~*p*:

(*~p* | E)-2 = (0.8990 - 2) / (0 – 2) = 0.5506

As we can see, in this particular example the variation in degree of belief nearly doubles as the stake size doubles, from 0.0279 to 0.0506. As far as I am aware, deriving the Ramsey degree of belief from the Bernoulli expected utility is original to this thesis.

In order to compare these Bernoulli degree’s of belief with the Stake Size Variation Principle we need to interpret the knowledge that the chance of *p* = ½ in some way. I suggest that we let K*p* be equal to the wealth of the individual, and K*p*/K = ½ . We can then calculate the evidential probability using the Stake Size Variation formula:

P(*p*)­­B-C = (C + K*p*) / (B + C + K)

So, for the 1 ducat bet, with fortune 4 ducats and K*p* = 4

P(*p*)5 -4 = (4 + 4)/(5 + 4 + 8) = 0.4705

And for the 2 ducat bet, with fortune 4 ducats and K*p* = 4

P(*p*)6-4 = (4 + 4)/(6 + 4 + 8) = 0.4444

Although not an exact match, we can see that the stake size variation degree of belief is pretty close in this example to the degree of belief derived from Bernoulli’s expected utility using Ramsey’s formula.

### 4.2 The Weber-Fechner Law

However Bernoulli’s alternative calculation was widely ignored, and instead people cherry picked the idea that anomalous preferences under risk could be explained by the diminishing marginal utility of money.

The theory of diminishing marginal utility is that there is a non linear relationship between money gains and utility gains. Typically utility is seen as a logarithmic function of money. The theory gained popularity in the 19th century when it was applied to sensation in general. The Weber-Fechner law states that any change in sensation is related to the change in stimulus as a proportion of the size of stimulus according to the following equation:

Change in Sensation = Constant x Change in Stimulus/Stimulus.

This formula successfully predicts, for example, that it is easier to sense the difference between a 1g weight and a 2g weight pressed on your palm, than a 1000g weight and a 1001g weight pressed on your palm. Similarly it would predict that a change from 30 decibels to 31 decibels would involve a greater change in loudness than a change from 100 decibels and 101 decibels.

If we view utility as a kind of sensing of value, then adding an extra £100 to a fortune of £100 will involve a greater increase in utility than adding the self same £100 to a fortune of £1 000 000. If this is accepted, then it follows that people will prefer smaller stakes bets to larger stakes bets at the same odds just so long as all the outcomes are positive states of wealth. Let’s suppose that a person with 1 unit of wealth were to consider a bet that paid 0.5 units of wealth if he won and cost half a unit if he lost. The stimulus of ending up with 0.5 units from 1 unit would involve –0.5/0.5C = – 1C (where “C” is the constant) of change in utility; whereas the stimulus of ending up with 1.5 units of wealth would involve + 0.5/1.5 C = 1/3C of change in utility. If we suppose that it is utility we are after, then it would be rational to take this bet only if we considered the probability of winning to be greater than ¾, since although the odds in money are 1 : 1, the odds in change in pleasure are 1 : 3. So if the Weber-Fechner law gives a good approximation of the relationship between money and utility, and if people actually calculate expected utility, rather than expected money gain, then a certain type of anomalous preference is explained, whilst preserving the theory that people’s degrees of belief match the zero stake chances when the chances are known, regardless of stake size.

I would just like to highlight that at this point it is not clear what this “utility” refers to. It is not clearly a sensation. Neither is it obvious that it is an objective property of the outcomes. It only shows up in calculations of expected utility, since the functions from money to utility preserve ordering under certainty. So it seems to be a function of people’s reasoning about risk; in particular how we calculate the expectation. This then begs the question: ought the ideally rational agent calculate expectation using utility or money? If the answer is in money, then the utility theory is an error theory, it merely describes and predicts a certain type of systematic error. But if the answer is in utility, then utility theory is a normative theory of rationality.

### 4.3 Friedman and Savage and the wiggly line utility curve.

Ramsey invented a measure theory for utility, or “good” as he called it, based on preferences. Given a measure theory for utility, a formula relating money gains to utility gains has empirical meaning. By 1950 it was established that peoples’ actual preferences don’t conform to any simple diminishing marginal utility curve since there are too many exceptions, notably the market price of state lotteries and insurance premiums. In their ground breaking paper, Friedman and Savage (1948) examined the actual preferences of large numbers of people and hypothesised a “wiggly line” utility curve that modelled overvaluing of lottery tickets and insurance premiums against the *actuarial value.* The actuarial value of a lottery ticket or an insurance premium is the sum of the stated probabilities of the outcomes multiplied by the cash value of the outcomes. If we then assume that the value a subject attaches to a prospect is its stated probability multiplied by its *utility*, then we can calculate the utility a subject attaches to an outcome. Friedman and Savage admit in their paper that the wiggly utility curve that follows lacks plausibility. For one thing it is clear that no one *actually* calculates expected utility in this way, by consulting a wiggly utility curve. But it is not obvious that they *ought to* either. The most that can be claimed is that they behave *as if they do*. This descriptive claim is compatible with the Stake Size Variation Principle.

Although there are many well known criticisms of Friedman and Savage’s approach, I will not rehearse them here. This is because, in essence, Friedman and Savage’s approach can be considered friendly fire. Empirically, we are in agreement, it is only in the fine tuning that a difference shows up. But someone entirely hostile to the idea of stake size variable rational degrees of belief may object that anything that the Stake Size Variation Principle explains can be adequately explained by “wiggly utility curves” (ibid. 297), and this may form a kind of weak objection to the Stake Size Variation Principle. In answer to this type of objection we will argue that although a wiggly utility curve might do as well as the Stake Size Variation Principle to explain risk aversion and lotteries, there are a number of famous cases where people’s preferences can be explained by the Stake Size Variation Principle, but can not be explained by any utility analysis whatever. These are: the Allais paradox, the Ellsberg paradox and the Reflection Effect. In what follows we will show that the preferences revealed in these three cases follow from the Stake Size Variation Principle, but do not follow from any Stake Size Invariant expected utility theory. Given the greater generality of the Stake Size Variation Principle, it should be preferred as an explanation even when the utility theory provides an adequate explanation. We will then go on to show that the SSVP can give an adequate account of both the Lottery and the Preface paradox, showing it to be preferable to threshold theories of rational belief that necessarily fall foul of one or the other paradox if they insist on stake size invariant evidential probability.

## Chapter 4

## SOLUTIONS TO PARDOXES FOR DECISION THEORY

### 1. Introduction

The SSVP combines the ratios k*p* / KP and C / (B + C) in order to give a two dimensional account of evidential probability that is sensitive to the stake size (B – C), and at the same time sensitive to the value of evidence (KP). This two dimensional account has the evidential probability varying across different prospects whilst retaining probabilistic coherence within any prospect thus providing immunity from Dutch books where any Dutch book is taken to be a single prospect. The extra dimension of stake size variability means that a deliberating agent can discriminate between probabilities that are backed by a great deal of evidence and probabilities that are backed by very little evidence. The difference will not be in the point probability relative to any particular bet, but will be in the rate of change of probability relative to stake size. This can be measured by the bid ask spread at any particular stake size. In general, stake sensitive degrees of belief recommend *buying low and selling high*. The difference between your buy price and your sell price is your bid ask spread. The SSVP has the bid ask spread get narrower as the value of evidence gets greater. This feature of the SSVP provides the missing link in one dimensional treatments of epistemic probability. The missing link is how to distinguish cases where the agent is very confident in their probability judgement because it is grounded in a great deal of evidence from cases where the agent is unsure of their probability judgement because they have very little evidence. This lacuna is sometimes referred to as the problem of second order probability, or of levels of certainty. The problem with existing treatments is they are often unclear how a different level of certainty can impact on decision and expected utility. The SSVP on the other hand simply records the value of evidence and gives a bid ask spread interval for any stake size. The greater the value of evidence supporting an evidential probability, the smaller the bid ask spread at any stake size. It follows from the SSVP that if you have a greater value of evidence relevant to *p* than to *q,* so that Kp > Kq, while the stake free probability is the same, so k*p* / KP = k*q* / Kq, then you will prefer bets on both *p* and on *~p* to bets on *q* and on *~q* at the same stakes. In other words you prefer to buy a bet on *p* or sell a bet on *p* at specified odds and stakes, to either buying or selling a bet on *q* at the same odds and stakes. This feature explains the Ellsberg paradox.

The other aspect of the SSVP is that it explains why people seem to prefer lower stakes bets at the same odds. This feature is a rival explanation to the orthodox expected utility explanation. The SSVP is less orthodox so has the burden of argument. The argument in this chapter is that the SSVP explains preferences that cannot be explained by orthodox expected utility theory, namely the Allais paradox and the reflection effect. The SSVP again out preforms its rival because of the two dimensional account. The SSVP assigns different evidential probabilities at different stake sizes given the same evidence. The Preface paradox, the Allais paradox and the Reflection effect all assume that rational agents *must* assign the same probability across stake sizes, and discover that actual agents are inclined to have irrational preferences. But if we assume that people’s evidential probabilities conform to the Stake Size Variation Principle, then these preferences are rational, because it is rational to proportion your belief to your evidence in accordance with the SSVP, and if you do, then you will make the preferences that test subjects tend to make in these decision games.

**1. Two principles**

In order to discuss: the Lottery paradox; the Preface paradox; the Allais paradox; the Ellsberg paradox; and the Reflection effect sensibly we will confine our discussion to explanations of the phenomena that are consistent with the following two principles:

**Principle of Preference**:

A prospect with a higher expected utility is preferable to a prospect with a lower expected utility.

This principle follows from the definition of expected utility.

**Expected Utility Principle**:

The expected utility of a prospect is the sum of the products of the utilities and the degrees of belief in the outcomes.

This principle follows from the way in which degrees of belief are defined.

If we adhere to these two principles, then the SSVP accounts for people’s preferences in the paradoxes in a way that standard expected utility theory cannot.

**2 Stake Size Variation Principle**

The Stake Size Variation Principle can be summed up as follows:

1. k*p* is the value of knowledge in favour of *p*.
2. k*~p* is the value of knowledge in favour of not *p*.
3. KP is the total value of knowledge relevant to whether or not *p*.
4. k*p* + k*~p* = KP
5. The ratio k*p*/KP gives the chance of *p* in cases where no stake size is relevant.
6. In a prospect with outcomes B if *p* and C if *~p*, then the stake size = B – C, which as a convention we will sometimes call “U”. The evidential probability of *p* relative to this gamble and knowledge KP will be written PK(*p*)U.
7. In case of two outcomes B if *p* and C if *~p*, and value of knowledge k*p*/KP such that 0 < k*p*/Kp< 1, then the evidential probability PK(*p*)B-C = (C + k*p*)/(B + C + KP)
8. If Kp is all positive, so that k*p*/KP = 1, then the probability PK(*p*)U = KP/U. (If KP/U > 1, then this means that P(*p*)U is certain with certainty level KP/U).
9. This has the general effect that degrees of belief attached to positive stake sizes get lower as the absolute stake size gets higher, and degrees of belief attached to negative stake sizes get higher as the absolute stake size gets higher.
10. The rate of stake size variation is inversely proportional to the value of knowledge.

### 2. The Ellsberg Paradox

The Ellsberg paradox is a decision game where people reliably prefer to bet on a proposition about which they have more information to one about which they have less information. On a fairly standard representational theory, it appears that people with these preferences are probabilistically incoherent. The SSVP solves the Ellsberg paradox by allowing that propositions about which there is more information have a higher value for K, and therefore a narrower bid-ask spread. In certain situations, like the Ellsberg paradox, this will lead to a preference to both sides of a bet on *p* over both sides of a bet on *q* without incoherence. This then shows that the standard response to the Ellsberg choice game is not incoherent.

Imagine there is an urn containing ninety balls. You are told that exactly thirty of the balls are red, and the remaining sixty balls are either yellow or black. You are provided with absolutely no information about the number of black balls or yellow balls other than that they sum to sixty. In other words, there could be no yellow balls, or there could be sixty, or any number in between.

You are now told that a ball will be chosen randomly from the urn. You are offered a choice between choosing a red ticket, or a yellow ticket.

Choose the red ticket, and if the ball chosen is red, then you will win £100; but if the ball chosen is yellow or black, you win nothing.

Choose the yellow ticket, and if a yellow ball is chosen, then you will win £100; but if black or red then you win nothing.

Which ticket would you choose? Most people presented with this choice choose the red ticket. The reason for this preference is that they have more definite information about the red ball. They know that there are exactly 30 red balls, whereas they only know that there are between 0 and 60 yellow balls.

So far there is no problem for standard decision theory. Call the agent’s degree of belief in red: “bel(red)” and in yellow: “bel(yellow)” where “bel” is a stake size invariant probability function that conforms to the axioms. The expected utility of the red ticket is bel(red)\*£100, whereas the expected utility for the yellow ticket is bel(yellow)\*£100. The final values are the same, so we must suppose that bel(red) > bel(yellow). This is explained by the fact that there is less information pertaining to yellow, so the probability is lower.

But now you are offered a choice between a red/black ticket and a yellow/black ticket. A red/black ticket pays £100 if the ball chosen is either red orblack, but pays nothing if it is yellow. A yellow/black ticket pays £100 if the ball is eitheryellow orblack but pays nothing if it is red.

People, perhaps including you, tend to prefer the yellow/black ticket over the red/black ticket when offered this choice. And the reason given is the same. There is more information pertaining to the yellow/black ticket than to the red/black ticket. We know that there are exactly 60 yellow/black balls, whereas we only know that there are between 30 and 90 red/back balls.

But this is where standard decision theory hits a snag. Using the same reasoning as before, the expected utility of the red/black ticket = bel (red/black) \* £100, whereas the expected utility of the yellow/black ticket = bel (yellow/black) \* £100. So preferences reveal that bel(red/black) < bel(yellow/black). This is explained by the lower level of information pertaining to red/black as compared with the information pertaining to yellow/black.

The snag is that taken together, these bel assignments are incoherent. This is because the yellow/black ticket is equivalent to the negation of the red ticket. So the disjunction of yellow/black and red should sum to 1. So bel(red) + bel(yellow/black) = 1. Another way of putting this is that it ought to be evidentially certain that the ball will be either red or not red.

But the yellow ticket is equivalent to the negation of the red/black ticket. So it follows that bel(yellow) + bel(red/black) = 1. But this cannot be because we have revealed that bel(yellow) < bel(red) and bel(red/black) < bel(yellow/black). So if it is certain that the ball is either red or not red, then it must be less than certain that it is either yellow or not yellow which is absurd. It seems that the revealed degrees of belief are on the face of it probabilistically incoherent. Yet they seem reasonable. Further more, it is clear that no expected utility account can explain these preferences in terms of wonky utility curves since the outcomes are constant across the decisions.

The reason that the decision theoretic analysis hits this snag is that it assumes Stake Inversion Invariance. In other words it assumes that the bid price probability PK(*p*)B – C = the sell price probability PK(*p*)C – B. In the decision game this general assumption leads to the specific assumptions that PK(red)B – C + PK(~red)B – C = PK(yellow)B – C + PK(~yellow)B – C = 1. The SSVP on the other hand allows that at positive stakes PK(yellow)B – C + PK(~yellow)B – C < PK(red)B – C + PK(~red)B – C < 1. This means that as long as we don’t assume Stake Inversion Invariance, then the choices most people give in response to the Ellsberg game are the right ones to have for the reason that they give: the red ticket and the not red ticket are preferable because they are the more informed choice.

Let us see how this works in the SSVP formula. For all four tickets the stake size is the same. B – C = (W + £100) – (W + 0) where W is the starting wealth of the subject. (For convenience we will assume W = 0, although this is an extreme case). We haven’t been given a value of evidence, but we can say that in general: kred > kyellow. We can also assume that kred/Kred = kyellow/Kyellow = 1/3. This is all we need to show that the Ellsberg preferences are in accord with the SSVP.

Since there are three mutually exclusive jointly exhaustive propositions in the partition, making two propositions and the negation of their disjunction, then the denominator becomes B + 2C + KP. Thus the formula for PK(*p*)B – C = (C + k*p*)/(B + 2C + KP).

Given that:

i) kredis greater than kyellow, and

ii) B is greater than C, then

iii) (C + kred)/(B + 2C + Kred) > (C + kyellow)/(B + 2C + Kyellow); therefore

iv) Pk(red)B – C > PK(yellow)B – C

Likewise, since yellow/black and red/black are disjunctions of two mutually exclusive propositions the general formula is PK(*p*)B – C = (2C + k*p*)/(B + 2C + KP)

i) k~redis greater than k~yellow, and

ii) B is greater than C, then

iii) (2C + k~red)/(B + 2C + K~red) > (2C + k~yellow)/(B + 2C + K~yellow); therefore

iv) Pk(~red)B – C > PK(~yellow)B – C

Coherence is maintained because when the stakes on the red ticket are positive, then the stakes on the yellow/black ticket are negative and vice versa.

We can see then that the Ellsberg preferences accord with degrees of belief that conform to the SSVP. But to make it clearer, let’s plug in some real values. Suppose you had £100 as a bank roll. The value of the evidence that thirty balls are red and the rest are either black or yellow is worth £3000 as evidence relevant to red, and £1500 as evidence relevant to yellow, making Kred = £3000; kred = £1000; Kyellow = £1500 and kyellow = £500.

Therefore:

Pk(red)B – C = (£100 + £1000)/(£200 + £200 + £3000) = 11/34 = approx. 0.324

PK(yellow)B – C = (£100 + £500)/(£200 + £200 + £1500) = 6/19 = approx. 0.316

PK(~red)B – C = (£200 + £2000)/(£3400) = 22/34 = approx. 0.647

PK(~yellow)B – C = (£200 + £1000)/(£1900) = 12/19 = approx. 0.632

With these degrees of belief, then the red ticket is preferable to the yellow ticket and the yellow/black ticket is preferable to the yellow/red ticket. These are the preferences that most people have with regard to the Ellsberg game.

### 3. The Allais Paradox

The Allais paradox presents a decision game involving very large cash prizes. When presented with this decision game, people consistently reveal preferences that cannot be explained by any representational theory which posits stake size invariant degrees of belief. Allais (1953, 1979) published a paper reporting a survey he had conducted in 1952 whereby test subjects who had good training in probability theory and could be considered rational, routinely violated the expected utility axioms in response to this game. Savage’s response to the Allais game was, ironically, that though he at first revealed the same preferences as the majority of respondents, and thereby violated his own axioms, he then purported to show that the preferences were in fact irrational because they defied the “sure thing principle.” However, if we accept the Stake Size Variation Principle, the Allais preferences can be explained without concluding that the preferences of seemingly reasonable numerate people are routinely irrational.

Examine the table below and consider, in a one off case, whether you would prefer 1A or 1B. Then consider whether you would prefer 2A or 2B. The prizes are distributed depending on which of a hundred numbered tickets are drawn from a hat. The tickets 1-89 are yellow, the ticket 90 is black and the remaining tickets 91-100 are red. Call yellow “y”, black “b” and red “r”. In ordinary treatments of the Allais Paradox, it is assumed that the correct degree of belief in these outcomes is stake size invariant: P(y) =0.89; P(b) = 0.01 and P(r) = 0.1. We will assume in what follows that these are the stake free objective probabilities of these events, in other words that ky/Ky= 0.89, k/Kb = 0.01 and kr/Kr = 0.1. We will also assume evidential completeness so that Ky = Kb = Kr and for the purposes of example set Kticket = $1 000 000.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Ticket  Drawn | 1A | 1B |  | Ticket  Drawn | 2A | 2B |
| 1-89 | $1 000 000 | $1 000 000 | Yellow | 1-89 | 0 | 0 |
| 90 | $1 000 000 | 0 | Black | 90 | $1 000 000 | 0 |
| 91-100 | $1 000 000 | $5 000 000 | Red | 91-100 | $1 000 000 | $5 000 000 |
|  | Prefer |  |  |  |  | Prefer |

In test studies, when presented with the pair of choices 1A or 1B, and 2A or 2B, the majority of people prefer 1A to 1B and 2B to 2A. But whatever private subjective utility functions people attach to the outcomes, if their beliefs are coherent, they should either prefer A to B in both cases or B to A in both cases. One way of seeing this clearly is to observe that for tickets 1 -89, there is no difference between A and B in either case. So whatever relative utility is attached to $1 000 000 and 0, it will make no difference to the difference in expected utility between 1A and 1B, or 2A and 2B. So any difference in utility must be in the remaining 2 rows. But in the remaining 2 rows, 1A : 1B is identical to 2A : 2B. So the *only* way for the expected utility of 1A > 1B and 1B < 2B is if the Stake Size Variation Principle is being applied and the probabilities of each event varies across the prospects.

**Solution to Allais Paradox**

To solve the Allais paradox using the Stake Size Variation Principle we need to assign a value for Kticket, which is what we shall call the total value of knowledge relevant to the colour of the draw. Let us assign the value Kticket = $1 000 000, and distribute this over the objective probability. So K*y*= $890 000; K*b*= $10 000 and K*r* = $100 000. This assignation is not completely arbitrary. It gives a value for Kticket that is equal to the median return of $1M. This is the minimum value for Kticket that allows the disjunction yellow, black or red, to be certain in option 1A. In every published treatment of the Allais paradox it is assumed that the probability of getting $1M in choice 1A is 1. If the value of knowledge Kticket were less that $1M, then Kticket/U < 1, and the probability of getting $1M in option 1A would be less than 1.

We also have to decide how to assign a stake size to a three value outcome. We can do this by taking each prospect as three individual bets on the three outcomes. The expected utility of each prospect is then the sum of the expected utility on each of these three bets.

We can now calculate the expected utility of each of the 4 prospects. We take each event and calculate the probability using the formula P = (C + k*ticket*)/(B + C + Kticket). For each event, C = 0; the value for B is the value of the prize on this outcome indicated in the table; Kticket = $1 000 000 ($1M); and k*ticket* is the number of tickets of this colour multiplied $10 000, so k*y* = $890 000; k*b* = $10 000; and k*r* = $100 000.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Ticket  Drawn | 1A | 1B |  | Ticket  Drawn | 2A | 2B |
| 1-89 (y) | $1 000 000 | $1 000 000 | Yellow | 1-89 (y) | 0 | 0 |
| 90 (b) | $1 000 000 | 0 | Black | 90 (b) | $1 000 000 | 0 |
| 91-100 (r) | $1 000 000 | $5 000 000 | Red | 91-100 (r) | $1 000 000 | $5 000 000 |
| EU | Prefer  $1 000 000 | Don’t Prefer  $528 330 |  |  | Don’t Prefer  $55 000 | Prefer  $83 300 |

Evidential Probabilities and expected utilities, 1A, 1B, 2A and 2B.

1A. $1 000 000 (y); $1 000 000 (b); $1 000 000 (r)

We assume that the probability of the partition is 1 at the stakes involved. It is not a part of the Allais problem to doubt that you will receive $1 000 000 in option A, so the expected utility of option A is clearly $1 000 000.

1B. $1 000 000 (y); $5 000 000 (r); 0 (b)

PK(*y*)$1M = (C + k*y*) / (B + C + Ky) = (0 + £.89M)/ (£1M + £1M) = 0.445;

PK(b)0 = (C + k*b*) / (B + C + Kb) = £0.01M/£1M = .01;

PK(r)$5M = (C + k*r*) / (B + C + Kr) = (0 + £0.1M)/(£5M + £1M) = 0.0166.

Expected utility = 0.445 \* 1M + 0.1 \* 0 + .0166 \* 5M = $528 330

Therefore, according to the Stake Size Variation Principle, a rational agent should prefer 1A to 1B.

2A. 0 (y); $1 000 000 (b); $1 000 000 (r)

PK(y)0= (C + k*y*) / (B + C + Ky) = £0.89M / £1M = 0.89;

PK(b)$1M = (C + k*b*) / (B + C + Kb) = £0.01M/£2M =0.005;

PK(r)$1M = (C + k*r*) / (B + C + Kr) = £0.1M/£2M = 0.05.

Expected utility = 0 +$0.005M + $0.05M = $55 000

2B. 0 (y); $5 000 000 (r); 0 (b)

PK(y)0= 0.89;

PK(b)0 = 0.01;

PK(r)$5M = 0.0166

Expected utility = 0 + 0 + 0.0166 \* $5M = $83 300

Therefore, according to the Stake Size Variation Principle, a rational agent should prefer 2B to 2A.

In this example we have stipulated that the value of knowledge Kr,y,b = £1 000 000 and C = 0. However, we could have chosen any positive value for C and the preference ordering would be the same just so long as the value of knowledge was sufficiently large.

These are the preferences that most people, including Savage, have when presented with the Allais choice game. This provides a very powerful vindication of the Stake Size Variation Principle against the objection that standard decision theory can explain the same results using non linear utility curves. It also preserves the reasonableness of the Allais preferences. This, in my opinion, is reason to prefer the Stake Size Variation Principle to orthodox expected utility accounts. This is because orthodox expected utility accounts have the consequence that the Allais preferences are irrational, yet many rational people have these preferences.

**Kahneman and Tversky**

In their ground breaking paper: *Prospect Theory: An Analysis of Decision under Risk*. (*Econometrica*, Vol. 47, No. 2. (Mar., 1979), pp. 263-292.) Tversky and Kahneman presented the results of some questionnaire studies in which they demonstrated that the Allais preferences were to be found in the majority of test subjects. This amounts to empirical evidence that the Stake Size Variation Principle is, at least in the majority of cases, descriptively true. In this thesis we have presented arguments for the normative force of the Stake Size Variation Principle. These are arguments that one rationally ought to vary one’s degrees of belief in accordance with the Stake Size Variation Principle. Supposing that the Stake Size Variation Principle is a rational principle, then the Allais preferences are explained. Most people have the Allais preferences because the Allais preferences are rational, and most people are rational.

### 4. The reflection effect

Tversky and Kahneman found another interesting effect in peoples’ preferences that they called the *reflection effect*. The reflection effect is their name for the preference of *more* risky prospects over losses. Basically, if you reverse the signs on all the gains so that they all become losses, then people will prefer the gamble to the sure thing when the expected utility calculated using objective chances is the same.

The phenomena that I have been seeking to explain using the stake size variation principle is often called “risk aversion”. The risk in a prospect can be associated with the probability and the size of the loss in the worse case. Using objective probabilities uncorrected by the Stake Size Variation Principle, people will tend to be risk averse. They will prefer less risky prospects over more risky prospects with the same actuarial value. It is thought to be possible to explain this due to the marginal disutility of money. The theory is that there is a logarithmic function from money gains and losses to utility gains and losses, making the utility odds of big bets skewed in favour of losses. This assumes that all outcomes are positive, counted up from some absolute zero utility.

But Tversky and Kahneman found that people are *risk seeking* over losses. In other words, when presented with an option of a certain loss, or a gamble over two losses with the same expected utility calculated from stake invariant chances, then people prefer the gamble. This really ought to spell the end for any non linear utility explanation for descriptive psychology since it directly contradicts the predictions made in terms of concave utility curves.

Tversky and Kahneman suggest an S shaped value curve which is concave in the gain region and convex in the loss region. This irregular, non linear mapping from money to utility shows what a huge strain maintaining the assumption of stake size invariance is under. However, the Stake Size Variation Principle explains the reflection effect preferences without the need for an S shaped utility function.

Kahneman and Tversky conducted a study to reveal the reflection effect preferences. The table below shows the study results. The subjects were asked to choose between G1 and H1, and asked to choose between G2 and H2 (Mnemonic; G for gamble, H for home).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Outcome | Stated  probability | Pay Off G1 | Pay Off H1 | Pay Off G2 | Pay Off H2 |
| P | .5 | +£1000 | +£500 | -£1000 | -£500 |
| ~P | .5 | +£0 | +£500 | -£0 | -£500 |
| Preference |  | 16% | 84% | 69% | 31% |

In this study around a hundred people were given the above pair of hypothetical choices. They were told explicitly the “probability” of the stated outcomes. The choice was between G1 or H1; and between G2 or H2. We can see from the bottom line the percentages that preferred the prospect over its partner. 84% of the test subjects preferred H1, a gift of £500; to G1, a 0.5 chance of £1000; whereas 69% preferred G2, a 0.5 chance of losing £1000; to H2, a straight loss of £500.

This is surprising for the standard expected utility theories, since losses are supposed to be of a greater marginal utility than gains, so if it is a non linear utility function that explains why H1 is preferable to G1, then the same utility function would have H2 preferable to G2.

The non linear utility function would explain the preference of H1 over G1 in terms of the decreasing marginal utility of money. One method is to assume an initial level of wealth and take the logarithm of the sum of this figure and the gains and losses. We then can have a 1 to 1 mapping from money to utility. This is easy to do, but it means that all money values must be positive and non zero. The assumption is that no one is ever completely broke, since they can always sell their labour. So let’s assume an initial wealth of £2000 and create a utility score for each of the values in the table.

|  |  |  |  |
| --- | --- | --- | --- |
| Gain/Loss | Wealth | Utility (log Wealth) | Expectation |
| +£1000 | £3000 | 3.48 | (½, 3.48) = 1.74 |
| +£500 | £2500 | 3.40 | 3.40 |
| 0 | £2000 | 3.30 | (1/2, 3.30) = 1.65 |
| -£500 | £1500 | 3.18 | 3.18 |
| -£1000 | £1000 | 3 | (1/2, 3) = 1.5 |

We can see that the expected utility of the options are:

G1 = 3.39; H1 = 3.40; G2 = 3.15 and H2 = 3.18.

So according to the diminishing marginal utility explanation of risk aversion, people should prefer H1 to G1, which they do, and H2 to G2, which they do not.

**The rationality of the reflection effect.**

However, if we apply the Stake Size Variation Principle, then the two preferences are mutually coherent given very basic rules of preference. Given that stake size variation is what makes the expected utility of H1 greater than G1, then it follows from the same principle that the expected utility of G2 is greater than H2.

Tversky and Kahneman point out that the objects of utility are *gains* and *losses*, not states of affairs. The sign in front of the sum of money gives this direction. So -£1000 is not a state of wealth, but a *loss* of £1000.

In an exchange between two people any transaction will have the reflected outcomes in terms of gains and losses for both parties. So for example if Alf sells Beth a bicycle for £100 pounds then Alf has *lost* a bicycle and *gained* £100, whereas Beth has *lost* £100 and *gained* a bicycle. The gains and losses cancel each other out and the exchange is zero sum. If we name the utility of the bicycle “B”, then the exchange can be represented as:

Alf +£100 – B.

Beth -£100 + B.

If we assume that B has the same monetary value to Alf and Beth, then the exchange is only fair if B = £100. Whereas if we assume that B has a differing value, BAlf for Alf and BBeth for Beth, then the exchange is fair whenever £100 – BAlf = B­Beth- £100.

To find this exchange preferable, then Alf must consider £100 > ~ -BAlf. This means that Alf prefers to *gain* £100 than to *not lose* a bicycle. Now if we reverse the signs, we get - £100 < ~ +BAlf. This means that Alf prefers to not gain a bicycle over losing £100. This had better be right, otherwise Alf’s actions would be contradictory, since as soon as the transaction was completed he would immediately want to buy the bicycle back from Beth at the same price he just sold it for.

Now let us consider the prospect {£1000 if *p*, £0 if *~p*}. Let us call this prospect G, and think of it as a contract, or a promissory note, that can be exchanged like a bicycle. Suppose Alf evaluates G at £400. This means that he would prefer to not lose £500 than to gain G. By the same logic, he would prefer to not to lose £500 than not to lose G. Therefore he would prefer to lose G than to lose £500.

In general we have the rule of exchange preference:

RULE OF EXCHANGE PREFERENCE.

If S prefers gaining H over gaining G, then S prefers losing G over losing H.

In the table above, “losing” G1 or H1 is equivalent to “gaining” H2 or G2. Most people prefer H1 to G1, so according to the rule of exchange preference, they should prefer G2 to H2. This is in fact the case.

To clarify, the prospects in the table are as follows:

G1 = + {£1000 if p, 0 if *~p*}

H1 = + {£500 if *p*, £500 if *~p*}

G2 = – {£1000 if *p*, 0 if *~p*}

H2 = – {£500 if *p*, £500 if *~p*}

So according to the rule of exchange preference then if S prefers H1 to G1 then S prefers G2 to H2.

According to the Principle of Preference stated earlier, if S prefers H1 to G1, then the expected utility H1S > G1S.

According to the Principle of Expected Utility, if H1S > G1S, then S’s degree of belief that *p* relative to prospect G1 < ½. This is because the expected utility of G1 = 0 + db(*p*) (1000 – 0) = db(*p*)(1000). By S’s preferences, then db(*p*)(1000) < (500), and therefore db(*p*) < ½.

The Stake Size Variation Principle explains this on the grounds that S’s degree of belief is sensitive to the stake size, which relative to G1 is £1000 – 0 on *p* and 0 - £1000 on ~*p*; and relative to H1 is (£500 - £500) = 0 on *p* and ~*p*. Therefore the degree of belief in *p* should be lower relative to G1 than relative to H1; whereas the degree of belief in ~*p* should be higher relative to G1 than relative to H1. The degree of this variation depends on variables to do with each individual subject, like their wealth and their certainty levels in the testimony of the experimenters.

Let us suppose, for example, that a subject had KP = £1000 and presume an initial wealth of £1500. In this case PK(*p*)£1000 = £2000/£5000 = 0.4, and the expectation of G1 would be £400. This means that the subject should be indifferent between the gamble G1 and £400. Now let us suppose that the agent took the gamble G1 and added it to his holdings before it was redeemed. This would mean that £400 would be added to the wealth of the subject in the same way that banks and insurance companies consider the expected value of risky assets to be a part of their stock. Now suppose that the subject was offered the choice between G2 and H2 and chose G2. This would have the effect of neutralising G1, since, if *p* he would lose £1000 and gain £1000, and if ~*p* he would lose nothing and gain nothing, with a net result of zero. Therefore, the effect of adding G2 is to subtract G1, which is to subtract an expectation of £400, and leave the subject with the same holdings that he started with. If he chose H2, on the other hand, he would subtract £500 from his holdings, leaving him with £100 less that his original holdings. Therefore a rational agent with stake size variant degrees of belief should prefer G2 to H2 if they prefer H1 to G1.

This argument assumes that G1 – G2 = 0, but given that the Stake Size as a proportion of the wealth on G2 is greater than on G1, we should re-calculate the probability *p* using lower values for B and C. We thus obtain PW£1500K(*p*)£1000 = 1/3. This makes the expectation on G2 a loss of £333.33. This is clearly preferable to a loss of £500, making G2 preferable to H2 in line with revealed preferences.

However there is something paradoxical here. We appear to have two different ways of calculating the expectation with two different results. The Stake Size Variation Principle, as we have been using it, does not have a value for the starting wealth of the individual. It only gives the two final values, B if *p* and C if *~p*. If B is higher than C, we consider it a bet *for p*, and P(*p*)B-C(B – C) gives the *bid price*, in other words the maximum price you would be prepared to pay in order to receive the prospect (B – C) if *p*. If B is *lower* than C, then the stake size is negative and we consider it a bet *against p*. In this case P(*p*)B-C(B – C) gives the *ask price*, in other words, the minimum price you would ask to receive for the obligation to pay (B – C) if *p*. Now, in prospect H2, if we assume a starting wealth of £1500, then it looks like a negative stakes bet on *p*, since B is lower than C by £1000.

BET against *p*: B = £500; C = £1500; Kp = £500; KP = £1000

PK(*p*)-1000 = 2/3.

So the *ask* price for this bet is PK(*p*)-1000 (£500 - £1500) = £666.66.

So it seems as if G2 has two different values; -£333.33; which is better than - £500, and ‑ £666.66, which is worse. But this is in fact a mistake. We can see the mistake if we consider that if one were to get offered *more* than the ask price of £666.66, then one would be pleased. But if one was offered a debt of *more* than £666.66, then one would not be pleased. The value of -£500 in H2 behaves more like a bid price, than an ask price, since it becomes more attractive the greater the probability of *p*.

Both *bid* and *ask* prices that we have discussed so far are *positive*, in that it is a positive amount paid in exchange for a positive prize. In a bid price, you calculate how much of a price you are prepared to *pay,* in order to receive a *greater or equal prize* if *p*. In an ask price, you calculate the minimum price you are prepared to receive in order to pay a greater or equal prize if *p*. So the value of prize – price is always positive[[19]](#footnote-19). Because these two options are *positive*, the probability gets lower for winning and higher for losing as the stakes get higher. We have discussed the rationale behind this. The greater the stakes, the greater the expected deviation, and if the negative deviation exceeds the wealth of the individual, it will cause *ruin* and a termination of the run when the losses exceed the gains therefore a greater average frequency of losses over the whole run than projected by the objective probability. But when the trend is reversed, i.e. when you are in a lose/lose situation and the value of prize – price is negative, as in H2 : G2, (-£1000 – – £500 = – £500), then a greater deviation can lengthen the run. For example, with our stated wealth of £1500; then H2 will mean certain ruin in 3 turns, whereas with G2, the subject has only a 0.25 chance of being ruined in less than 3 turns and a 0.25 chance of being ruined on the 3 turn, making a total of 0.5. So looked upon as a repeatable propositional function, you ought to go for G2. This is the rationale behind the Stake Size Variation Principle in general, and so it also provides a rationale to keep the degree of belief in *p* the same over the positive realm and over the negative realm. This means that PK(*p*)B – C = 1/3, and that £333.33 is the maximum price you would be prepared to pay to avoid paying £1000 if *p*.

Kahneman and Tversky actually managed to elicit this dual response in test subjects in the study above. When they presented the choice G1, H1 and G2, H2 above, they gave the subjects a gift of £1000 before hand in G1 and H1; and a gift of £2000 before hand in the choice between G2 and H2. The gift makes the end results identical according to Tversky and Kahneman: (1979. p.273).

G1: {£2000, if *p* (0.5); £1000, if ~*p*(0.5)}

G2: {£2000, if ~*p* (0.5); £1000, if *p*(0.5)}

H1 & H2: {£1500}

The majority of test subjects, as we have already mentioned, preferred H1 to G1, but preferred G2 to H2.

It is clear that absolutely no expected utility theory that assumes stake size invariance can explain this result since both the probabilities and the end state utilities are the same. So the only possible explanation is if the probability of *p* is stake size variable and not equal to 0.5. In our example the P(*p*)£2000 *-* £1000­ = 3/8. In this case G1 is {£2000, 3/8; £1000, 5/8} = £1375 and G2 is {£2000, 5/8; £1000, 3/8} = £1625 making H1 preferable to G1, but G2 preferable to H2.

### 5 The Lottery Paradox and the Preface Paradox

The SSVP allows for a probabilistic theory of knowledge that runs the gauntlet between the fallibility and the certainty of knowledge. We can illustrate how the SSVP navigates these issues by considering the Lottery paradox and the Preface Paradox. In the Lottery Paradox we are shown that no matter how high the probability of losing a lottery it can never be known that we will lose until the draw has taken place, showing that dropping the certainty condition for knowledge leads to paradox. The Preface paradox shows us that even evidentially certain beliefs can be fallible, and that this fallibility is in conflict with the certainty condition on knowledge. The SSVP resolves this conflict by allowing certainty at low stakes, but mere probable belief at high stakes.

In a lottery there are a large number of tickets each with a very low probability of winning. It might be thought that given high enough positive stakes, then the probability of winning will be so low that it ought to be considered *certain* that a ticket selected at random will not win, and therefore that we *know* that any ticket selected at random will not win. But of course this cannot be, because we already know that one ticket will win, and that is incompatible with knowledge that each ticket not win.

A particular misunderstanding of the probabilistic theory of knowledge might be to assume that because low stakes propositions can be known given less evidence, then propositions with very high probabilities can be known at low stakes even when the probability is less than 1. In a lottery the probability of losing is at negative stakes, so the greater the prize, the lower the stakes on losing and the higher the probability. It might be thought that given a big enough lottery with a big enough prize, then SSVP will assign knowledge to each ticket losing, leaving it vulnerable to the lottery paradox.

This is a misunderstanding because it assumes that the evidential probability remains constant relative to fixed evidence whereas the SSVP has the evidential probability vary with the stakes. We will now show that, formally speaking, it is never known that a ticket will lose a lottery before the draw.

Given a partition of N + 1 mutually exclusive and jointly exhaustive propositions *p*0 - N, where there is no empirical evidence in favour of any proposition, then:

PK( *p*i)B – C = (C + k*p*i) /(B + NC + Kp)

So in every case where C > 0 or k*p*i > 0 then PK( *p*i)B – C > 0.

We proved that normalisation holds at inverted stakes, so it follows that:

PK(*pi*)C – B = 1 – (C + k*p*i) /(B + NC + Kp)

So the evidential probability of any ticket losing a lottery of any size is less than one, provided that either C > or k*p*i > 0.

Let us briefly discuss the two exceptions. The first exception is that when both k*p*i and C = 0 then PK(*p*i) = 0. We can put this into words by saying that, according to the SSVP, if there is zero evidence in favour of winning, and if everything of value is at stake, then you are better off politely refusing the ticket. I take it that under these conditions the paradox does not arise, since if you do not have any evidence, then you do not know that one ticket will win, so there is nothing contradictory in believing that every ticket is certain to lose. The external nature of evidence will ensure that the evidential probability that the actual winning ticket will win is not 0 even in this degenerate case. Besides which, as Bernoulli said, you are never entirely broke while you are alive, and as I say, you are never entirely ignorant either. Although this case is interesting and there could be interesting discussion regarding what exactly this amounts to, this is an avenue for further research and for the purposes of this thesis we will just treat this as a degenerate case.

The other exception is when N = ∞, in other words in a lottery with infinite tickets.

In this case it is plausible that

PK( *p*i)B – C = (C + k*p*i) /(B + ∞C + Kp) = 0,

In which case

PK(~*pi*)C – B = 1 – (C + k*p*i) /(B + ∞C + Kp) = 1.

This result depends on the idea that any number divided by infinity is equal to zero. However, we prefer to assign this an infinitesimal value, so that PK(~*p*)C – B is not evidentially certain but *evidentially almost certain*. Evidential almost certainty does not entail truth and is not closed under logic. In this case the lottery paradox does not arise. We will come back to the infinite case at the end of this section.

Having briefly discussed these two exceptions, we can discuss the non degenerative cases of a lottery. We have shown that the SSVP respects additivity as long as the stakes are identical. This is not so straight forward as it first appears.

Identity of stakes holds between PK(*p*)B – C and PK(*q*)D – E iff B = D and C = E; it is not sufficient for B – C = D – E.

For example, let us imagine a punter called John with £1000 to his name. There is a lottery with 2000 tickets each costing £1 each. The prize for this lottery is £2000. We can call a lottery with (N+ 1) tickets where the prize is (N + 1) times the price a “fair lottery”. (We say N + 1 where N is the minimum number of mutually exclusive propositions necessary to make a lottery of this size, the extra ticket is then defined as the negation of the disjunction of all N propositions). Now let us suppose that John considers buying 1 ticket. We can calculate the expected utility like this:

C = £1000 - £1 = £999

B = C + £2000 = £2999.

N = 2000 – 1 = 1999

Let us suppose that the value of evidence that one ticket will win out of 2000 is worth £1 per ticket, so that k*t*i = £1 and KT = £2000.

PK(*ti*)B – C = (C + k*t*I)/(B + NC + KT) = (£999 + £1)/(£2999 + £1997001 + £2000) = 1/2002

The expected utility of this gamble then is C + PK(*ti*)B – C (B – C) = £999 + £2000/2002 = £999.999

This would lead John to be fairly indifferent between buying a lottery ticket for £1 and not buying one, since the difference is a fraction of a penny. We can see that this accords pretty well with a fairly reasonable attitude to lottery tickets. There is no harm in buying a lottery ticket, but neither is it a good way of making a profit.

Now let us suppose that John considers buying many (M) lottery tickets. Although the stake size will still be £2000, the stakes are different because he is investing a lot more. So the values for both C and B will be lower making the gamble higher stakes. He can afford up to £1000. To avoid the degenerate case let us say he buys 999 tickets. Now we have:

C = £1000 - £999 = £1

B = C + £2000 = £2001

M = 999

N = 1999

K*t*1 – M = £999

PK(*t1-m*)B – C = (MC + t1 - M )/(B + NC + KT) = (£999 + £999)/(£2001 + £999 + £2000) = 999/2500

The evidential expectation would then be: C + 2000\*999/2500 = £800.20

So this gamble is seriously disadvantageous and is equivalent to giving away £200. We can see how additivity does not hold here due to the change in stakes, making the latter gamble have a much worse per ticket expectation. If we imagine John to be an agent whose beliefs conform to the SSVP principle, then we would expect John to happily buy one ticket to this lottery and consider £1 to be a fair price, especially if there is any extra pleasure in the gamble. But we would also expect John to never buy large numbers of tickets. Since this is how people behave towards lotteries, (many people buying small numbers of tickets, very few people buying large numbers of tickets), then the SSVP comes out well in its analysis of the descriptive expectation of lotteries.

But this analysis may seem to offend against the axiom of additivity. Here is the argument:

We have seen that PK(*t*i)B – C  = 1/2002 and PK(*t1-m*)B – C = 999/2500. But *t*1-M is equivalent to the disjunction: *t1* OR *t2* OR…OR *ti* …OR *t*M. Therefore PK(*t*1 –M)B - C = ∑i = 1 – M PK(*ti*)B - C = M(PK(*ti*)B – C = 999/2002.

But 999/2002 > 999/2500, so therefore John has incoherent probability assignments.

This argument is flawed because the two probability assignments are at different stakes so additivity does not need to hold for probabilistic coherence. We can show that additivity does hold when identity of stakes holds if we recalculate PK(*t*i)B – C using the values C = £1 and B = £2001. In this case:

PK(*t*i)B – C = £2/(2001 + 1999 + 2000) = 1/2500.

So as long as we preserve the identity of stakes condition, then additivity holds and

PK(*t*1 –M)B - C = M(PK(*t*i)B - C = 999/2500.

We can see what this means in terms of ticket pricing by working out the expectation on a single ticket at these stakes.

C + PK(*t*i)B – C(B – C) = £1 + £2000/2500 = £1.8.

This should make John consider 80p to be the maximum price he should pay for a ticket at these stakes. But these stakes would only arise on a single ticket at this price if he started out with £1.80.

So we can see that there is no problem with additivity just so long as the identity of stakes is preserved.

Another objection to the SSVP analysis of lottery pricing is this: it appears as if the probability of any particular ticket winning in a lottery of (N + 1) tickets is less that 1 / (N + 1). This means that if we preserve identity of stakes, then the probability of one of the N + 1 tickets winning ought to be less than 1. But we know from the set up that one of the N + 1 tickets will win.

To work this into the example, let us give John an extra £5000 bringing his starting wealth up to £6000. Keeping everything else the same, let us suppose that he buys *all* the tickets. This makes:

C = £4000

B = £6000.

PK(*ti*)B – C  = (C + k*ti* ) / (B + NC + Ki) = 4001/ (6000 + 7 996 000 + 2000)

= 4001/8 004 000

PK(*t*1 to (N+1))B – C

= (NC + 1C + k*t*i) / (B + NC + Ki) = (N + 1) PK(*t*i)B – C

= (8 002 000)/(8 004 000) < 1.

It is also interesting to note that the value of knowledge missing from the numerator is £2000 which is (B – C), the stake size. This is equal to the value of knowledge gained by observing which ticket is drawn from the lottery. This is a constant, and the constant is that the value of knowledge necessary for certainty is (B – C), which is the stake size.

The reason for this anomaly is that the stake size itself is logically impossible once John buys all the tickets. As we said in chapter 1, the value B is retained for the negation of the disjunction of all the propositions bet upon. In the bet John is entering by buying all the tickets, the value C becomes £6000, not £4000. This is because there is no longer any alternative *~ti* in which John ends up with £4000.

We will go over the logic again here in this context. In buying all the tickets, John is betting for and against any proposition in the partition. To illustrate this let us number the tickets t1 to t2000. Now take the proposition t1, and its complement ~t1 which is equivalent to the disjunction of t2, t3, … t2000. So now we have the logical truth that either t1 will win, or one of tickets t2 to t2000 will win which is equivalent to the logical truth that either t1 will win or t1 will not win. Now, in the example John has:

1. A bet on t1 at stakes B – C, such that t1 results in B, and ~t1 results in C

And

2. A bet on ~t1 at stakes B – C, such that t1 results in C, and ~t1 results in B.

These two bets are inconsistent since t1 results in B and ~B and C and ~C.

Consequently the only way it is possible to have a degree of belief in the disjunction of all the tickets winning at a non zero stake size B – C is if you allow the possibility that all the tickets will lose. In which case there is no logical problem with having a degree of belief less than 1 in the disjunction of t1 OR t2 OR … OR t2000, because this proposition is no longer presupposed by the set up.

One way of conceptualising this is to think of the proposition that one ticket will win the lottery. Typically when considering lotteries in philosophical contexts, this proposition is taken as given and constitutes knowledge. However, in any actual lottery there is nothing sacrosanct or logically necessary about this proposition. Let us suppose you are ignorant of the proposition that one ticket will win. In this case, according to the SSVP, at stakes B – C, the evidence required to be certain of this proposition is equal to B – C, which is equal to the shortfall in the value of knowledge of the disjunction of the partition at stakes B - C. This is again equal to the increase in value of evidence given by the observation that one ticket won this lottery to the propositional function that in lotteries of this kind one ticket will win.

But as long as there remains a possibility that John can lose, then the partition can be divided up into some *p* and *~p* such that *p* results in B and *~p* results in C, and PK(*p*)B - C + PK(*~p*)C – B = 1, which is enough to ensure coherence and invulnerability from Dutch Books.

On this last point I refer to Jon Williamson (1999) where he gives a Dutch Book argument for countable additivity. He contemplates a game of “guess the number” where he thinks of a number. For any *n* what is the probability that *n* is the number that Williamson is thinking of? This can be seen as a lottery with infinite tickets, although there is no guarantee that each number is equally likely to be chosen. Given the normal Dutch Book set up Williamson proves that as long as your distribution sums to 1 and countable additivity holds, then you will not be Dutch Booked. However, if countable additivity does not hold, and your distribution sums to less than 1, then you will be able to be Dutch Booked. I see no problem here for the SSVP because the Dutch Book argument presented relies on Stake Inversion Invariance. In other words the bookie is allowed to choose whether the stakes are positive or negative, and the subject is required to give a single price at both positive and negative stakes.

J. Williamson considers the case where the sum of the probabilities on each proposition *qi* is less than one. The bookie then sets the stakes to negative on each proposition and proves thereby that whatever number wins, the winnings will be less than the loss. In the SSVP example above, this would be equivalent to John pricing each ticket at just under £1 at positive stakes, and then the Dutch bookie buying every ticket off him at this price. In this scenario John would lose whatever ticket won the lottery. However, this is illicit in two ways. Firstly, the stakes would no longer be non zero if the bookie bought *all* the tickets, so the identity of stakes would not hold. Secondly, by inverting the stakes, the Bookie is compelling John to sell lottery tickets at his bid price instead of his ask price, so again, the identity of stakes does not hold.

In a genuinely infinite lottery then *q*i would have to be infinitesimal and ∑i = 1 - ∞ *q*i = 1 – x where x = (B + k*i*)/(B + ∞C + Ki), which we take to be infinitesimal. So the price of each ticket would be *q*i(B – C) which would again be infinitesimal given finite stakes. The bookies strategy then would be to buy every ticket from John at *q*i(B – C) for a total of (1 – x)(B – C) and then expect a prize of (B – C) which ever ticket wins. The bookie would make a certain profit of x(B – C) in this case, which is an infinitesimal amount. But as above, that would be because she was i) buying tickets from John at his bid price rather than his ask price, ii) buying *all* the tickets, thus collapsing the stake size.

### 5. 1 THE PREFACE PARADOX

The lottery paradox has never been much of a threat to the probabilistic theory of knowledge because the probabilistic theory of knowledge has evidential probability 1 as a necessary condition for knowledge. However, the lottery paradox is often taken in conjunction with the Preface Paradox (David Makinson 1965). The Preface Paradox is a direct threat to the principle axiom of the probabilistic theory of knowledge, that knowledge entails an evidential probability 1.

A historian writes a book in which he includes 100 historical propositions, which we can call h1 – h100. The historian is very rigorous and intellectually responsible. He is quite certain that each proposition in the book is true given the evidence he has gathered, otherwise he wouldn’t dream of publishing. Let us call the evidence relevant to each proposition in the book e1 – e100 respectively. We can then say that for each hi, PK(hi | ei)B – C  = 1, where B – C are the stakes on each proposition being true. You may be wondering what the stakes are on a historical proposition being true. Good question! Don’t worry; we will come on to that later.

The historian being epistemically modest then writes in the preface to the book that given the large number of propositions in the book and given human frailty, he is very likely to have made at least one mistake, and he assures the reader that this mistake is all his own. We can then say that from the historian’s evidential perspective PK(h1 &…& hi &…& h100 | e1&…& ei &…&e100)D - E < 1, where D – E are the stakes on the conjunction of all the facts in the book. In other words, the probability that every proposition in the book is true given the totality of is evidence is less than certain.

This is a problem for the certainty condition for knowledge because it would appear that the historian knows each conjunct hi without knowing the conjunction of all hi (which we will henceforth abbreviate to “&h”). This would be fine if knowledge entailed an evidential probability less than 1, since the probability of the conjunction could be as low as you like if the probability of each conjunct was less than 1. We will call a theory than allows knowledge at an evidential probability of less than 1 a “threshold theory”. Threshold theories can also apply to rational belief. This is in fact how the Preface paradox is usually framed. The historian has an inconsistent set of beliefs, all of which appear rational. He rationally believes the facts when considered individually, but he rationally believes that the conjunction of the beliefs is false. To translate the Preface into a paradox about knowledge, we will assume that the facts in the book are all true, so that the only bar to knowledge is whether or not the Historian is rational to believe.

Hawthorne and Bovens, (1999) as well as Scott Sturgeon (2008), deal with this paradox easily with a threshold theory. The probability that each statement in the book is false is small enough so that the subjective probability the author has in the truth of each proposition exceeds some arbitrarily high threshold, high enough to warrant a knowledge attribution. But taken as a single proposition, the conjunction of all the statements in the book has an accumulated probability of falsehood that exceeds the threshold for reasonable belief.

Sturgeon also tackles the lottery paradox in the same fashion. He observes that plausibility of a reductio argument reduces in proportion to the number of claims in play. This is consistent with a sub unitary probabilistic threshold view of coarse grained belief, a view that Sturgeon attributes to Locke. The general story Sturgeon tells as a response to the Lottery and Preface paradoxes taken together is that coarse grained belief is determined by a fine grained belief threshold, and is therefore not closed under the conjunction rule.

The conjunction rule is, in Sturgeon’s words:

“If one rationally believes *p*, and rationally believes *q*, one should also believe their conjunction: (*p* & *q*).”

This rule can be waived if we have a sub unitary probabilistic threshold theory of course grained belief. But the conjunction rule for *knowledge* claims seems to be undeniable. Surely, if you know that *p* and you know that *q*, then you know that *p* & *q*?

The threshold theory solution is not available for the SSVP precisely for the reason that the Lottery paradox is not a problem for the SSVP: the certainty condition for knowledge. The threshold solution relies on PK(hi | ei) < 1; whereas we stipulate that PK(hi | ei) = 1. If the evidential probability of each conjunct is 1, then the conjunction rule holds, probabilistically or otherwise.

Our solution to the paradox again relies on the insistence on the identity of stakes as a condition on the rational requirement of probabilistic coherence. The solution is simply that the stakes are much higher on &h than on the sum of the stakes on hi, so that the value of evidence &e is not conclusive that &h. This means that the value of evidence can be higher than the stake size in the case of each individual conjunct where the sum of the value of evidence for the conjunction can be lower than the stake size on the conjunction. Thus according to the SSVP one sided evidence formula:

PK (hi)B – C = khi / (B – C ) or 1 if hi > B - C

Whereas:

PK(&h)D – E = k&h / (D – E) or 1 if k&h > (D – E)

Therefore it is possible that k&h is less than certain while hi is certain, provided that k&h < (B – C) whereas khi > (D – E).

This would be possible if khi> k&h, but we exclude this possibility, since any evidence for the conjuncts is clearly evidence for the conjunction. In fact, let us stipulate that khi is constant for all h1 – 100, and that the evidence relevant to the conjunction k&h = 100khi. This is about as hard as we can make the Preface, because ordinarily the evidence in support of a hundred historical propositions would include a lot of overlap, and would therefore be less than a hundred times the value of evidence supporting each conjunct. However, the value of evidence for the conjunction could be up to a hundred times the value of evidence on each conjunct, so if it works on this assumption, then it will work for any preface.

Under these conditions the Preface is still possible under the SSVP if (B – C) > 100(D – E). This result is in itself enough to show that the SSVP can handle the Preface paradox whereas no stake size invariant theory of probability that has evidential probability 1 as a necessary condition for knowledge can.

But the question remains: why must the stake size on the conjunction be so much higher than the stake size on the conjuncts? Further more, what does that even *mean*? The answer to these questions depends on the acceptance that knowledge is of genuine objective value, and that knowledge of some propositions is more valuable than the knowledge of others. The probabilistic theory of knowledge is based upon the axiom that the value of knowing a proposition is the difference in value that the truth of the proposition makes as compared with its falsehood given that you act on the belief that the proposition is true. We can also say that the difference the truth of a proposition makes as opposed to its falsehood is its *empirical content*. The difference the truth of a proposition makes compared with its negation is how we have defined stake size. It follows that the stake size on a proposition is its empirical content.

This is a substantive metaphysical claim with a heavy burden of argument. However it can be argued for in a formal way if we pay attention to the conditional probability constraint. The constraint is as follows:

PK(h | e)B - C < PK(h&e)B - C /PK(e)B - C

We said that the reason that the conditional probability is less than the ratio of conjunction over conjunct is that for the Kolmogorov ratio definition to hold under a betting interpretation, the stake size on the conjunction has to be higher than the stake size on the conjunct. Consequently, if the conditional probability equals the ratio, then the stake size on the conjunction must be higher than the stake size on the conjunct, driving the probability of the conjunction down, and thus making the ratio smaller. We can run this the other way and show that if the probability of the evidence is 1 and the probability of the hypothesis and the evidence is 1, then the probability of the conjunction given the evidence is less than 1. This only makes sense if we understand the stake size escalation of conditional probability.

To help understand this we can go back to Charles Peirce who explained to us in Love Chance and Logic (1949) that inductive inference was ampliative, in that the knowledge contained in the conclusion of an inductive inference is greater than the knowledge contained in the premises. It can be of no doubt that a conjunction of N propositions has a greater empirical content than any of the constituent propositions. What may not be quite so obvious is that a conjunction contains more information than the sum of the information contained in the conjuncts.

One way of arguing for this is to accept Lewis’s conception of evidence as ruling out possibilities. We could then measure empirical strength of evidence in terms of the sets of possibilities it rules out. A single proposition *e* rules out one set of possibilities: those in which ~e; whereas a conjunction of two propositions *e* & *f* rules out three sets of possibilities: those in which *e & ~f*; those in which *f* & ~*e*; and those in which ~*f* & *~e*. The more propositions we add, the more possibilities are ruled out by the largest conjunction. In fact the rule is simple: for a conjunction of N propositions, 2N – 1 possibilities are ruled out by the conjunction. In this way, the number of propositions creates an exponential expansion in the number of possibilities. It is in this way that inductive inference is ampliative. Each of N results of a repeatable trial rules out one possibility, but the entire sequence of N results rules out 2N – 1 possibilities. This makes the information in the conclusion of an inductive inference greater than the sum of its parts.

Now we look at it from the other direction and ask what value of evidence is required for knowledge of a single proposition, and what value of evidence is required for knowledge of a conjunction of N propositions. If the value of evidence is measured in the number of possibilities it needs to rule out, then the value of evidence needed for knowledge of the single proposition will be 1U, whereas the value of evidence needed for knowledge of the conjunction will be (2N – 1)U.

We can see the reality of the link between value and empirical content if we look at the concept of an amplifier bet. Suppose you had £1000 pounds to invest in 10 football matches. Suppose that the odds on the home team of each match was 1 : 1. The value of evidence sufficient for knowledge that the home team wins a particular race is no more than £2000 since that is the maximum stakes you can afford at these odds. You can put £100 on each match and win a profit of £1000, which together with your investment makes a stake size of £2000.

But if the bookies allow amplifier conditional bets, then you can bet: £1000 on the first home team; £2000 on the second home team conditional on the first; £4000 on the third conditional on the second and the first; and so on, giving you a massive profit of (210 – 1)£1000, which is £1023 000, on the 10th conditional bet, at stakes of £1024 000. In other words, the value of knowing that each home team will win is a maximum £2000, whereas the value of knowing that all ten home teams will win is worth over a million. This is because the conditional knowledge of the conjunction enables a stake size of £1024 000.

Of course this relies on the market odds being 1 : 1 on each match conditional on the previous match at unlimited stakes. This resonates with the logical probability formed by indifference between each home team winning or losing and the assumption that the matches are probabilistically independent. This would give 1/1024 as the logical probability of the conjunction. Now we can see that the observation that the home team had won the nth match would increase the logical probability of the home team winning the nth match by ½ to certainty, whereas it would only increase the logical probability of the conjunction of all the home teams winning by 1/1024 to 2/1024. Because the stake size on the conjunction is 1024 times as great as the stake size on the conjuncts, then the increase in expectation is the same: £1000. This accords with the principle that the *value of evidence is equal to the increase in expectation*. The evidence is the same in both cases, it is the observation that the home team won the nth match. So the evidence should increase the expectation by the same amount, which it does, provided the stake size on the conjunction is 1024 times the stake size on the nth match.

It is best to add now that there is no need to restrict ourselves to logical probabilities here. The odds reflect the conditional probabilities between the races. Because the races are presumed to be independent, the conditional probabilities between the races are also ½. The stake escalation is a function of these conditional probabilities. The rule is that if the conditional probability is x, then the stake escalation is 1/x. Of course, the conditional bets are at escalated stakes so the conditional probabilities will themselves be affected by stake variation. As long as the conditional probability between the hypotheses is less than 1, then the stake size on the conjunction will be exponentially higher than the stake size of the conjuncts, allowing for evidential certainty of the conjuncts while having considerable doubt at the level of the conjunction.

Now to return to the Preface and the Historian with his book, the Preface is coherent provided the conditional probabilities between the statements in the book are low enough to make the stake escalation greater than 100, in other words, so long as the stakes on the conjunction are over 100 times the stakes on each statement. (It is not necessary that the stakes on each statement are the same, it just makes it easier to express if we assume that they are the same).

Not all books of a hundred facts will meet this criterion, so it might be worth talking about possible books that do not meet this criterion so that we can see that, in these cases, the Preface would not be rational.

For example, suppose the conditional probability between the conjuncts was 1. This would mean that the conjunction had the same stakes as the conjuncts, so that the preface would be necessarily incoherent. Such a book would be one where each fact followed logically from the previous fact. So for example if h1 stated that the earth was over 4000 years old and h2 stated that the earth was over 3999 years old and h3 stated that the earth was over 3998 years old and so on, then the Preface would be absurd. It would be absurd to be certain of each of these facts, but not to be certain of all of them, simply because to be certain of h1 just is to be certain of all the facts.

To take another example, if the conditional probabilities between the propositions were high enough, then the stake escalation would not be that great, and the Preface would lose plausibility. In this case it might be that the stakes on the conjunction is lower than a hundred times the stakes on the individual conjuncts. But in this case it is not so clear that the evidence for the conjunction will be a hundred times the evidence for each of the conjuncts, since the evidence would necessarily over lap, the evidence for the antecedent increasing the posterior probability of the consequent. If we imagine such a book, it is also intuitively the case that the greater the conditional dependence, the less plausible the Preface.

In general stake size escalation has the consequence that isolated facts can be of little value, but when presented together with a whole battery of related facts, then their value increases. This is a very general human phenomenon.

Take for example the fact that Cicero was married to Terentia. In the absence of any other information relevant to Cicero and Terentia, this fact is, to put it bluntly, *boring*. Evidence needed to prove this fact is of little value since there is little motive to doubt it, and you can simply take my word for it. It is of no interest and therefore needs little evidence for its acceptance. However, in conjunction with the fact that Cicero was a Consul in ancient Rome, and Terentia was very wealthy and was from an aristocratic family and a hundred other facts about ancient Rome, their marriage takes on a lot more significance and would need a lot more evidence to provide certainty. In general, non experts find it hard to understand the excitement and enthusiasm that experts feel for the minutia of their area of expertise. An expert will generally require a great deal more evidence to accept a claim in his area than a non expert.

In the Probabilistic Theory of Knowledge, conclusive evidence is *bet settling*. This is in the literal sense that evidence sufficient for certainty is evidence sufficient to settle a bet and vice versa. The solution to the Preface paradox is that the evidence the historian has is bet settling for each isolated fact, but is not bet settling for the conjunction of all the facts due to stake size escalation.

For example, suppose two non experts, a man and wife let’s say, had a disagreement about whether or not Cicero was married to Terentia, and so they decide to bet on it. Being non experts, the bet is at low stakes. They decide to settle the bet by looking it up in the Historian’s authoritative book on Ancient Rome. They discover that in fact Cicero was married to Terentia. The Historian’s book thus provides bet settling evidence for the fact that Cicero was married to Terentia, and in the same manner, any other proposition written in the book would settle low stakes bets for non experts. We can likewise presume that were the married couple to bother to ask the Historian to provide evidence for each fact in the book, the Historian would be able to provide them with evidence that would settle their bet.

But now the married couple decide to have a bet on whether the Historian’s book is truly authoritative. They decide have a bet on whether *every fact in the book is true*. How can they settle this bet? Well, one way they could do this is to check each fact in the book against the book itself, just in the same way that the married couple settled their bet on whether Cicero was married to Terentia by checking it in the book. But it is clear that this procedure could not settle the bet, since it presupposes the very thing bet upon. In the same way, given time and trouble, they could find the Historian and go through all his evidence with him. But again, this is unlikely to settle the matter, since even the Historian still retains a level of doubt at the stakes of the whole book, which is the root cause of the Preface in the first place. This shows that the sum of evidence that is bet settling for each conjunct is not sufficient to settle bets on the conjunction. The SSVP can explain this in terms of the higher stake size on the conjunction.

### 6 Decision weights

As a response to the Allais paradox and similar seemingly incoherent revealed preferences Tversky and Kahneman (1979) introduced the concept of *decision weights*, to fill the role of degrees of belief understood as calculators of expected utility.

“In prospect theory, the value of each outcome is multiplied by a decision weight. Decision weights are inferred from choices between prospects much as subjective probabilities are inferred from preferences in the Ramsey-Savage approach.

However, decision weights are not probabilities: they do not obey the probability axioms and they should not be interpreted as measures of degree or belief.

Consider a gamble in which one can win 1,000 or nothing, depending on the toss of a fair coin. For any reasonable person, the probability of winning is **.**50in this situation. This can be verified in a variety of ways, e.g., by showing that the subject is indifferent between betting on heads or tails, or by his verbal report that he considers the two events equiprobable. As will be shown below, however, the decision weight π(0.50) which is derived from choices is likely to be smaller than 0.50. Decision weights measure the impact of events on the desirability of prospects, and not merely the perceived likelihood of these events. The two scales coincide (i.e., π p = p) if the expectation principle holds, but not otherwise.”

These decision weights are precisely our stake size sensitive degrees of belief. We have shown that they do, contra Tversky and Kahneman, obey the probability axioms relative to any reference class. We have argued that “the probability of winning for any reasonable person,” on the toss of a fair coin is *not* 0.50, if “probability” is interpreted as rational degree of belief, because of the effect of stake size variation, and the greater probability at high stakes that the actual frequency will be less that 0.5 given that the bets will terminate when all the money is gone.

That a subject is indifferent between betting heads and tails does not show that their degree of belief is 0.5 according to the Ramsey-Savage approach unless it is assumed that degrees of belief are Stake Inversion Invariant. All it shows that their degree of belief in heads at positive stakes is equal to their degree of belief in tails at positive stakes, and this could be any probability. This is consistent with the Stake Size Variation Principle and the axioms of probability. It is also consistent with a verbal report that the two events are equally probable. It also allows that the expectation principle *always* holds for rational degrees of belief. Furthermore, it explains the actual preferences of the subjects in Tversky and Kahneman’s trials without recourse to a personal utility function.

### 7 Stake size variant degrees of belief and what is commonly meant by “probability”

On the other hand, stake size variable degrees of belief, although they do the job of probability in terms of expected utility calculation and conform to the axioms of probability, cannot be said to be what most people mean by *probability*, since it is clearly true that when no money is mentioned, people will *say* that the probability of heads on an even coin is ½. In other cases where the issue of the probability of well defined stochastic processes like dice and lotteries comes up, people have a fairly consistent view of what the probability of an event is without mentioning stakes. In what follows we’ll call this probability the “true probability”.

One possible interpretation of the true probability is the convergent expected frequency in a hypothetically infinite set of trials. However, this is quite an arcane mathematical concept and is an unlikely candidate for a common understanding. A reason against taking the true probability to have this interpretation is it is not sufficiently general. Although it applies very well to dice and coins and other repeatable events where there is a clearly defined reference class; it does not have a clear application in the majority of probabilistic judgements, where the event in question is not clearly repeatable, and reference classes to which it may plausibly belong are likely to be vague. For example, while it is clearly meaningful to talk about the true probability of a particular candidate winning a specific election, it is not so clearly meaningful to talk about an infinite series of such elections, let alone the convergent relative frequency of the candidate winning in the limit. It is not obvious what is to be repeated, or in other words what the reference class is supposed to be. It should be recognised that the majority of our probabilistic judgements are of this kind. In such cases, people can still make calculations of expected utility, without necessarily being able to give a clear reference class. In these cases we suggest that the evidential probability will be relative to the interests of the subject. But we recognise that one can still claim that there is a true probability in these cases, which is objective in the sense that it is independent of any particular set of interests. We can capture this objective sense of true probability using the concept of fair odds.

Given the hypothesis that people do use stake size variable degrees of belief to calculate expectation, it is likely that degrees of belief are more or less consciously accessible. We can then have the concept of fair odds*.* The fair odds are the odds at which neither party to a bet has an advantage. Although it requires a degree of empathetic intelligence not available to every human, it is possible at least for some people to see the transaction event from various perspectives, and therefore calculate the expectation on a bet from both sides. Given this we can calculate the true probability without need of a hypothetical reference class.

A bet between S1 and S2 on *p* can be represented as a pair of prospects: S1 {B if *p*, - C if *~p*} and S2 {-B if *p*, + C if *~p*}. If S1 and S2 have the same evidence for p then they should have the same values for db(*p* | E)(B + C), and the same values for db(*p* | E)(-B - C).

The expected utility of S1 will be:

- C + dp (*p* | E)(B + C) (B + C)

And the expected utility of S2 will be:

C + dp (*p* | E)(-B -C)( –B – C)

If these two figures are equal, then the bet will be fair and neither S1 nor S2 will have an advantage and would be indifferent as to what side of the bet they would take. We can then derive the fair odds by assuming that the expected utility A = 0 for both prospects in a situation when you have no other option. In this case the “true probability” will be (A + C)/(B + C), and this should fit the natural meaning of probability in a wider range of cases than the convergent frequency in the limit.

It might be helpful to illustrate with and example. Suppose db(*p* | E)10 = 0.65 and db (*p* | E)-10 = 0.75. Under these conditions, anyone with evidence E should consider a prospect {3 if *p*, -7 if *~p*} of equal value to a prospect {-3 if *p*, +7 if *~p*}. This is because:

EU {3 if *p*, -7 if *~p*} = - 7 + db(*p* | E)10 10 = - 0.5

EU {-3 if *p*, +7 if *~p*} = 7 + db(*p* | E)-10 -10 = -0.5.

This means the *“true probability”* of *p* | E = (0 + 7)/(7 + 3) = 0.7.

Our stake size sensitive degrees of belief not only give the true probability, but also carry the information that at stake size 10, the bid- ask *spread* of uncertainty is db(*p* | E)-10 – db(*p* | E)10 = 0.1. This will reflect properties of the evidence to do with sample size, relevance and variance.

In cases of non degenerate probabilities it should be easily within the realms of current statistical methods to calculate the *spread* of uncertainty using a combination of confidence intervals and probability density functions. However, the hope is that once the Stake Size Variation Principle is accepted, there is no need for the fiction of a “true probability” and the mathematics can be greatly simplified.

The true probability is only of interest in situations in establishing fair odds for zero sum bets. However, in nature the odds are set by reality, and we can’t always choose to bet for or against. For example, what is the probability that global warming will cause London to flood by 2040, given that carbon emissions remain constant? We can see this as a bet: we bet the pleasure of emitting carbon at the current rate against the destruction of London. There is only one side of this bet we can take, because we can’t swap the creation of London if we win the bet, for the inconvenience of continuing to emit carbon at the current rate if we lose. So the *fair odds* are irrelevant, and we should stack our odds in favour of London flooding when evaluating the prospect of continuing to emit carbon, since we won’t get a second bet if we lose.

### 8 How much evidence justifies certainty?

“We sometimes have enough evidence to establish what the result of an experiment would be without actually doing the experiment.” (Williamson, T, 2007 p. 141)

The central question can now be posed: how much evidence is required for certainty relative to a stake size? The Stake Size Variation Principle has the consequence that the value of evidence needed for certainty is equal to the value of knowledge that that certainty would bring, which is equal to the stake size. This is in effect a stake sensitive rule of succession. Rules of succession have occupied the attention of the greatest philosophers of probability, from Laplace to Ramsey. Rules of succession are usually expressed as giving the evidential probability in the result of a binary experiment N + 1, given the results of the N experiments so far conducted. If SN out of the N experiments have been successful, we can say that the results so far give a success rate SN/N. Laplace’s rule of succession is then that the probability that experiment N + 1 will be successful is equal to (SN + 1)/(N + 2).

A problem inherent in Laplace’s rule of succession is that no amount of evidence can ever make the result certain. Given a certainty condition for knowledge, this means that we can never come to know anything through enumerative induction. Another related problem proven by Ramsey (1991) is that Laplace’s rule of succession has a bias towards a probability ½. Both these problems are shared by classic Bayesian conditionalization from a prior probability. To put these problems decision theoretically, Laplace’s rule requires you to have a non zero degree of belief in a proposition for which you have no evidence whatsoever and for which you have considerable evidence against. For example, if 3 out of 3 people had died after ingesting the red pill, and no one has ever eaten a red pill without dying, then you should have a degree of belief of 1/5 that the next person to eat a red pill will not die. But the price of £100 for a prize of £500 seems a bit steep for a bet on an event that has never happened before. What is worse, if we project forward into a much larger trial, we should bet at even odds that more than 20% of people will not die after eating the red pill, a prediction that clearly goes beyond our evidence. In general, a consequence of Laplace’s rule of succession is that for any inductively held belief, there is some odds at which you should view it to be advantageous to bet on an event the like of which has never happened before. So although Laplace’s rule of succession may appear in one light to be epistemically conservative and risk averse, since it makes it impossible to be certain of anything even when there is a massive amount of positive confirmation; if we reverse the stakes Laplace’s rule becomes epistemically liberaland *novelty loving*, prescribing positive degrees of belief in propositional functions for which there is no positive evidence and a massive amount of disconfirmation. The same is true of any monotonic Bayesian updating rule.

With the Stake Size Variation Principle, we can give a more general rule of succession that aims at predicting the frequency of success in the next M results based on the frequency of success in the N previous experiments. The Stake Size Variation Principle is more flexible since it allows reference classes made up of events with varying stake sizes, giving more value to the more important results. To see how the Stake Size Variation Principle compares with Laplace’s rule, we can take a repeatable event that is presumed to always have the same stake size and scale the stake size to the unit 1. The value of evidence gained from N such experiments will be 1 each, with a total of N. If the binary results given the experiment are *p* | E and *~p* | E, then K*p* + K*~p* will equal N; so KP = N. Supposing the next M experiments likewise have a stake size of 1 each, then the stake size on the next M experiments will equal M. Since we are interested in knowledge we will only consider cases where K*p* = KP; in other words, when all N experiments have resulted in *p*& E. In this case the rule of succession is that PN(*p* | E)M = N/M if N < M; or 1 if N > M. In other words, as long as N is higher than M you can be certain of the inference *p* | E, but if M exceeds N, then your degree of belief in *p* | E in the class of M should only be equal to N/M. To return to the quotation at the beginning of this section, we have enough evidence to establish that the results of N experiments E will be *p* when we have conducted N experiments E all of which have resulted in *p*.

For example: suppose a naturalist discovers a bird of a species yet unseen in a large unexplored jungle. He names it the “bluebird” on account of the bright blue plumage of his single specimen. Although he names it the “bluebird”, the blueness of the plumage does not constitute an identifying criterion. He goes on to collect a large sample of bluebirds which appear to be quite common. He collects 1000 birds all of which are blue. So Kbluebird/Kbluebird = 1 and Kbluebird = N. Therefore PK(blue | bluebird)N = 1, for N ≤ 1000 bluebirds. This means, according to the above principle, he can claim with certainty that any number of bluebirds under 1000, will all be blue. This is to say that P(blue | bluebird)M = 1, whenever M ≤ 1000 bluebirds. Given that the stake size on each individual bet is commensurable with the cost of gathering the evidence, then he should bet at any odds for, and no odds against.

This can be argued for in the following way: suppose the naturalist were to claim to his friends in London that out of a thousand bluebirds, every single one would be blue. They take him up on his wager. It seems intuitively obvious to me that in order to settle this bet, it would be sufficient to take a random sample of a thousand bluebirds and check to see whether they were all blue. If they were, then the naturalist has won his bet, but if any one was coloured differently, then he would have lost his bet. If this is a reasonable way of settling the bet, then it means that *he already has evidence sufficient to settle the bet*, as long as he restricts his claim to a thousand birds. On the other had, if he was a little more reckless, he might claim that *all* bluebirds are blue. Let us suppose that, given the population density and the size of the jungle, it would be reasonable to estimate a total population of 1 000 000 bluebirds. To claim that all 1 000 000 birds were blue would be to go beyond his evidence. According to our formula, his degree of belief for what would be a bet at a positive stake size of 1 000 000 times the cost of a single specimen of bluebird, would be very low, at 1/1000. And so it should be. This system gives us the flexibility to be de facto certain at small stakes of any particular prediction issuing from a well supported generalisation, whilst having a very low degree of belief in the generalisation itself. Whilst being certain that if I randomly select at raven it will be black, I am fairly confident that if I offered enough money, somebody somewhere would be able to produce a raven that is not black. The Stake Size Variation Principle allows me to have these two attitudes, without being forced to consider it advantageous to bet against a randomly selected raven being black at some odds.

### 8.1 Level of certainty

What may appear paradoxical about this way of characterising certainty is that it means that one can be rationally certain in something that is false. N/M gives the level of certainty. Just so long as N/M > 1, then it is rational to have a degree of belief 1. But, given the fallibility of induction, it is always metaphysically possible that success rate in the evidence prescribes certainty, so SN/N = 1, but that the number of successes SM in the relevant reference class M is less than M, so SM/M < 1 (when M is not a subset of N). Here we must fall back on evidential externalism and distinguish between *rational certainty*which is internal to the evidence, and *evidential certainty*which has the further requirement that the evidence class must match the object class. While it is reasonable to be certain that *p* | E in M on the evidence that SN/N = 1 and N > M; it is only evidentially certain that *p* | E in M on the evidence that SN/N = 1 if *p* | E is objectively certain in both the evidential reference class N *and* the object class M. For evidential certainty, both SN/N and SM/M must equal 1, in other words (SN + SM)/(N + M) must equal 1.

So for knowledge, not only must the subjective degree of belief be 1 and the rational degree of belief be 1, but also the evidential probability relative to the relevant reference class must also be 1. This is in effect equivalent to a truth condition for knowledge, but it is not so severe as a truth condition, since it allows for knowledge in cases where there is no truth or falsity, in other words in hypothetical or counterfactual cases. So our naturalist of the previous section *knows* that, hypothetically, if he randomly selects a thousand bluebirds, they will all be blue. But let us suppose that out of all the bluebirds in existence there exists a single mutant with bright green feathers. Just so long as the naturalist never actually comes across this single anomaly, then the objective degree of belief that any bluebird he selects will be blue will be 1, and his degree of belief 1 will serve him better than any other degree of belief in terms of predictive accuracy. The existence of the unobserved green blue bird does not defeat his knowledge.

Let us suppose that he offers to bet his fellows in London that out of a thousand bluebirds randomly selected, all will be blue; he pays the expenses of the expedition and gives them the thousand specimens if he is wrong. But if he is right, they pay, and he gets to keep the specimens. Now let us suppose they don’t take him up on this wager, since the wager has convinced them that he is right. In this situation, I claim that he knows that any bluebird randomly selected will be blue, up to a thousand. We can forget about the metaphysically controversial possibility that the expedition *would* have gathered the one anomaly.

But supposing he *actually did* come across the single green bluebird. We can call this an *epistemic shock*. This phrase I am borrowing from Dorothy Edgington (2004, p.5), the phenomena is familiar enough. In Alan Hajek’s(2003) words, “probability 0 does not imply impossible”. Or to put it another way, “propositions accorded a probability 1 are liable to be false”. (Levi, 1967, p. 209). We do not have the monotonic constraints of Bayesian updating, so there is no logical problem here. His updated value of knowledge K*p*/KP will simply be (N + SM)/ (N + M). The evidential requirement for certainty makes sense since you can never through epistemic shock lose more than the value of the knowledge upon which you based your certainty. The value of knowledge is given simply by N times the unit stake size. So at certainty, as long as M doesn’t exceed N, then you will not lose more than N by evaluating the bets at M. However, if M exceeds N, then your degree of belief in S becomes N/M, and you will still never lose more than N. For example, let us suppose our naturalist still had his thousand bluebird specimens and took the wager with his London friends. Now suppose he lost dramatically, and found 1000 non blue bluebirds and no blue bluebirds. The amount he would lose in this wager would be 1000 bluebirds, which he already had. Also, his degree of belief in blue | bluebird would now be at most ½, which is no better than a guess, so in effect he has not only lost his specimens, but also their epistemic value. Had he bet on a million bluebirds being blue, the stake size variation principle would have only allowed him to bet at odds of 1 : 999. In which case, again he could only possibly lose the thousand birds he already had. And his new degree of belief would remain at 1/1000, so once again he has not lost more epistemically than he had to start with. The Stake Size Variation Principle has prevented him from losing more than he had. At no point can his betting behaviour ever exceed his experience.

### 8.2 The greater the certainty, the less impact epistemic shock has on degree of belief.

Epistemic shock is when a rational degree of belief 1 is assigned to a falsehood. Since inductive inference is fallible, this is always possible. After an epistemic shock, a rational agent will no longer be certain, but have a partial degree of belief. We can measure the impact of epistemic shock as the difference between certainty and the post shock zero stakes degree of belief. This will be:

1 – (N + SM)/(N + M).

Given that certainty is measured by N/M when N exceeds M, it follows that no epistemic shock can be greater than 0.5. It also follows that the greater the certainty, the less the maximum impact of epistemic shock on degree of belief.

This generalises into partial belief updating, and thus shows the connection between level of certainty and resilience in the face of new evidence. The difference between high levels of certainty and low levels of certainty do not show up in the odds that you will accept or the expected advantage, but will show up in the extent to which your degree of belief will be updated due to the outcome of the bet. This has a direct interpretation in terms of the maximum stake size at which you will accept those odds. When generalised to mixed evidence, the epistemic shock will be continuous and will be:

SN/N – (SN + SM)/(N + M).

This enables a distinction between say a degree of belief ½ in a coin toss over a range of say ten tosses, where N/M can be assumed to be infinite, and so the epistemic shock will be infinitesimal; and a degree of belief ½ in an unknown proposition based on the principle of indifference, where the epistemic shock over a trial of 10 will be potentially ½ either way. As such, the Stake Size Variation Principle has a lot in common with Good’s (1950)*weight of evidence* measure and Popper’s (1974) severity of test measure. However, the stake size variation measure is superior to both of these since it applies equally well to certainty and it makes clear the link between certainty and behaviour in terms of stake size variation.

### 9. Conclusion

A long standing tension in epistemology comes from two superficially opposing intuitions about knowledge: the certainty intuition and the fallibility intuition. The certainty intuition is that only beliefs held with certainty are knowledge. The fallibility intuition is that any belief, however well supported by evidence or reasoned argument, is possibly false. The first says that all knowledge is certain. The second says that nothing is certain. The absurd conclusion seems to be that nothing is known.

The certainty intuition is that certainty is required for knowledge. The certainty required for knowledge is not only a measure of conviction, but also a measure of evidential support. Using a long standing and respectable method of enumerating probability in terms of betting odds, we can define certainty more precisely as an evidential and subjective probability of 1. Subjective certainty is the willingness to bet at any odds for and at no odds against. It is also the acceptance of a bet as settled. Evidential certainty is thus the evidence required to settle a bet. The certainty intuition is thus explained. If you know that you will win a bet, then you will be willing to bet at any odds for, and no odds against. If you have evidence sufficient to settle a bet, then you know that you will win the bet.

The fallibility intuition stems from the fact that inductive inferences are ampliative, in that the premises and the conclusion contain more information than the premises alone. Hume pointed out that the relationship between the premises and the conclusion in an inductive inference is not one of necessity. In any inductive inference, it is *possible* that the premises are true, but that the conclusion is false. The age old sceptical move is to move from the premise that it is possible that a belief is false, to the conclusion that it is not certain that it is true.

Late twentieth century contextualist solutions to scepticism constituted real progress in this conflict. A general strategy for resolving this sceptical problem is to observe that what is considered possible can vary considerably with non evidential features of the context. All that is necessary for knowledge is that there is no *relevant* possibility of falsehood. We can capture the gist of this epistemological advance by giving a measure of what counts as a relevant possibility in betting terms. A relevant possibility is one to which the subject does or should assign odds at the stake size set by the context. Given this definition we can see that the contextualist solution is not actually necessary, since we already define certainty as a probability 1, which means that no situation compatible with *~p* has any non zero probability. The sceptic can just be seen to be conflating certainty with necessity. It is perfectly consistent to claim that it is possible that I will burst into flames in the next three minutes while simultaneously claiming that it is certain that I will not. Necessity entails certainty, but certainty does not entail necessity.

Kant observed in the Critique of Pure Reason that it is often the case that a person will be certain at low stakes, but become less than certain at high stakes. This observation has had many contemporary proponents when applied to knowledge ascriptions. Both contextualists about knowledge and Interest Relative Invariantists take it as given that a subject with the same evidence can know that *p* at low stakes, but fail to know that *p* at high stakes.

Sticking with our betting interpretation of certainty, this means that at low stakes, subjects can assign zero probability to ~*p* | E, while at high stakes they assign a non zero probability to ~*p* | E. Without altering our definition of certainty and without abandoning the intuition that certainty is necessary for knowledge, the sceptical problem is thus solved. This is intuitively plausible if we consider actual betting practices, and actual burden of evidence arguments. Evidence that suffices to settle a trivial matter, may not be sufficient to settle the same matter were it to become a lot more significant. Evidence sufficient to fine a man may not be sufficient to hang him. Evidence sufficient to claim that there are weapons of mass destruction in Iraq in a private conversation, are not sufficient to claim that there are in the context of justifying an international war.

If the evidential probability *p* | E can be less than 1 at high stakes and yet 1 at low stakes, then it follows that the same principle, the Principle of Stake Size Variation, must apply to all evidential probabilities, not just certainty and knowledge. The stake size variation of probability assignments was first noticed in the eighteenth century by Daniel Bernoulli. We demonstrated that Bernoulli’s risk measurement formula is consistent with the normative claim that probability varies with stake size. However, the matter is confused by a perhaps more dominant non epistemic meaning of “probability”. We define probability in terms of the calculation of expectation. This definition has clear consequences in terms of choice over action. Since certainty is related to action in the same way, it is this definition of probability that is in use when the claim is made that knowledge depends on probability 1. However, from Bernoulli to Tversky and Kahneman, there appears to be a stable meaning of probability that is independent of stake size. This can be given one of two approximate definitions. It is either the limiting frequency in a hypothetically infinite sequence of trials, or it is the fair odds given the evidence, i.e. the odds at which the rational person would be indifferent which side of a bet he took. Both definitions work very well for probabilities close to ½ , but don’t work at all for probability 1.

We have shown that both meanings of probability are reducible to an expectation definition of probability. We pointed out that the limiting frequency varies according to the selection of a reference class. Once the reference class is specified, then the objective degree of belief can be simply defined as the per unit value of all the bets in that reference class as a proportion of the stake size. This gives us a definition of probability that is completely general, mind independent and does not rely on metaphysical fictions of infinite sequences of repeatable events. Using this definition of probability we demonstrate that probability depends on stake size, because the size of the reference class will depend on how soon the player will run out of money. We also show that the uncertainty, which is the degree to which it is expected that the actual value of a sequence of bets will deviate from their expected value, increases with relative increases in stake size. This gives an epistemic rationale for the Stake Size Variation Principle.

The reality of the Stake Size Variation Principle has been known about for centuries, but has been modelled incorrectly using a logarithmic utility function. Ironically, this model has its etymology in D. Bernoulli’s risk measurement paper. We showed than the logarithmic utility function explanation for stake size variation in expected utility calculations fails to fit the data. Specifically it fails to fit the prevailing preferences exhibited by rational test subjects in the Allais problem, the Ellsberg paradox, and the reflection effect. Furthermore, someone who had a logarithmic utility function to money would exhibit inconsistent behaviour, since they would consider the same item to have a different intrinsic value when buying it and when selling it. The Stake Size Variation Principle explains all these preferences accurately.

We observe that Tversky and Kahneman came up with the concept of decision weights that fulfil the same function as our rational degrees of belief. However, Tversky and Kahneman gave no normative argument for these decision weights, and in fact explicitly denied that they conform to the axioms of probability. We, on the other hand, have shown that the Stake Size Variation Principle can hold, while degrees of belief remain coherent relative to a stake size. Because of the buy below, sell above nature of asset pricing, someone conforming to the Stake Size Variation Principle will never be Dutch Booked.

Finally we consider how we can enumerate the level of evidence necessary for a probability 1. This is in terms of the ratio between the value of evidence, measured as the sum of the past stake sizes, and the value of knowledge, in terms of the sum of the future stake sizes. We consider the possibility of a false belief held with evidential probability 1. We term this an epistemic shock. It falls short of knowledge because, while the evidential probability appears to a reasonable agent to be 1, the evidential probability relative to the specified reference class is necessarily less than 1. We can measure the severity of epistemic shock in terms of the size of the change in degree of belief caused by the shock.

We can now resolve the original conflict. It is true that certainty is a necessary condition for knowledge. It is also true that it is possible that any belief held on inductive grounds is false. What is not true is that this mere possibility is enough to undermine certainty. To undermine objective certainty, the possibility must be actual at least once in the relevant reference class. So all the sceptic can say is that we might not know anything on inductive grounds. But unless the sceptic is prepared to bet against us at some odds, then we can simply ignore him as an unrealistic dreamer and unreasonable doubter. If he does bet against us, then we know that we will win, and we have the experience to lend authority to our words.

# References and Bibliography

Allais, M., (1953), "Le comportement de l’homme rationnel devant le risque: critique des postulats et axiomes de l’école Américaine”, Econometrica 21, 503-546.

Allais, P. M. (1979) ‘The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American School’, in Allais, M. and Hagen, 0. (eds), *Expected Utility Hypotheses and the Allais Paradox****,*** pp. 27-145, Dordrecht: Reidel.

Armendt, Brad (2008) “Stake Invariant Belief” *ActaAnalytica* vol. 23 : 1 (46)

Bernoulli, Daniel (1954) “Exposition of a New Theory on the Measurement of Risk” in *Econometrica*, Vol. 22, No. 1. (Jan., 1954), pp. 23-36.

Chevalier**, Nicolas and** Blaye**, Agnès**, (2006) “ False-Belief Representation and Attribution in Preschoolers: Testing a Graded-Representation hypothesis”, Current psychology letters [Online], 18, Vol. 1, online since 26 April 2006, connection on 04 mars 2010. URL : http://cpl.revues.org/index233.html

Descartes, Rene. (1984/85). John Cottingham, Robert Stoothoff and Dughold Murdoch (eds. and trans.), *The Philosophical Writings of Descartes,* 2 vols. Cambridge: Cambridge University Press.

De Rose, K. (2004): *The Problem with Subject-Sensitive Invariantism*, Philosophy and Phenomenological Research 68, 346-350.

Dretske, Fred. (1970). "Epistemic Operators", *The Journal of Philosophy*, 67, pp. 1007-23

Dretske, Fred. (1981). *Knowledge and the Flow of Information*, Oxford: Blackwell

Edgington, Dorothy (1996) “Do Conditionals have truth conditions?” in Jackson ed. *Conditionals* OUP.

Edgington, Dorothy. (2004): “Two Kinds of Possibility” *The Inaugural Address* Supplementary to the Proceedings of the Aristotelian Society, Vol. 78. Issue 1. July 2004, Wiley Interscience.

[Ellsberg, Daniel](http://en.wikipedia.org/wiki/Daniel_Ellsberg) (1961), "Risk, Ambiguity, and the Savage Axioms", [*Quarterly Journal of Economics*](http://en.wikipedia.org/wiki/Quarterly_Journal_of_Economics) 75 (4): 643–669

Fantl, J. and McGrath, M. (2002): *Evidence, Pragmatics, and Justification*, Philosophical Review 111, pp. 67-94.

Frank, J & Garcia, P (2009): *Bid-Ask Spreads, Volume, and Volatility : Evidence from Livestock Markets* Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2009 AAEA & ACCI Joint Annual Meeting, Milwaukee, Wisconsin, July29, 2009.

Friedman, M and Savage, L. J. (1948) “The Utility Analysis of Choices Involving Risk” *Journal of Political Economy* 56, pp. 279 -304.

Gillies, Donald, (1990) “The Turing-Good Weight of Evidence Function and Popper's Measure of the Severity of a Test” *British Journal for the Philosophy of Science.*41, 143-146

Good, I.J. (1950) *Probability and the Weighing of Evidence* Charles Griffin Co. London

Gregoriou, A., Ioannidis, C & Skerratt, L (2005) “Information Asymmetry and the Bid-Ask Spread : Evidence from the UK” *Journal of Business Finance & Accounting, Vol. 32, No. 9-10, pp. 1801-1826, November 2005*

Hajek, Alan (2003) “What Conditional Probability Could Not Be.” *Synthese* **137**: 273-323

Hajek, Alan (2007) "The Reference Class Problem is Your Problem Too",*Synthese* **156**: 185-215.

Hawthorne, J and Bovens, L. (1999). “The Preface, The Lottery and The Logic of Belief”. *Mind* vol. 108. 430. April 1999. Oxford University Press

Hawthorne, J. (2004): *Knowledge and Lotteries*, Oxford: Oxford University Press.

Hume, David. (1825) *Essays and Treatises on Several Subjects* Edinburgh, Printed for Bell and Bradfute; [etc., etc.]

Jeffrey, R. (1970) “Dracula meets Wolfman: Acceptance vs. Partial belief” in *Induction, Acceptance and Rational belief*, Ed., Marshall Swain D. Reidel Publishing company/Dordrecht-Holland

Kahneman, D., and A. Tversky, (1979), “Prospect theory of decisions under risk,” *Econometrica*, 47, 263—291.

Kant, E.*A Critique of Pure Reason* (1781), Translated by Norman Kemp Smith, (1961) London Macmillan.

Kaplan, Mark (1996) *Decision theory as Philosophy.* Cambridge University Press.

Kelly, Jr. J. L. (1956) “A New Interpretation of Information Rate”. *The Bell System Technical Journal, July 1956; 917-918*

Kornblith, Hilary (1999) “Knowledge in humans and other animals” *Philosophical Perspectives,* ***13,*** *Epistemology.*

Kornblith, Hilary (2002) *Knowledge and it place in Nature* Clarendon Press, Oxford

Kyburg, Henry (1961) *Probability and the Logic of Rational Belief,* Wesleyan University Press.

Kyburg, Jr., H. E. (2007). Bayesian inference with evidential probability. In Harper, W. and Wheeler, G., editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg*, Jr., pages 281–296. King’s College, London.

Levi ,I. (1967) *Gambling with Truth* New York: Knopf

Levi, Isaac, ‘The Demons of Decision’, *The Monist*70, (1987), pp. 193 - 211.

Lewis, David (1994). “Humean Supervenience debugged”, *Mind*, New Series, Vol. 103, No. 412.(Oct., 1994), pp. 473-490.

Lewis, David (1996) ‘Elusive Knowledge’ *Australian Journal of Philosophy* 74, 4 pp. 549- 67.

Maher, Patrick, (1992) ‘Diachronic Rationality’, *Philosophy of Science* 59, 120 - 41.

Makinson, David (1965) “The Paradox of the Preface.” *Analysis* 25

Mian, Shehzad, (1995) “Bid Ask Spread and ownership structure.” *Journal of Financial Research* December

Mill, John Stuart. (1859) *On Liberty* Batoche Books, Kitchener 2001.

Millikan, Ruth. (1989) "Biosemantics" *The Journal of Philosophy* 86**.**6 (1989) pp. 281-297.

Millikan, Ruth. (1993) *White Queen Psychology* M.I.T.

Nozick, R (2001) *Knowledge and Scepticism* in *Epistemology* Blackwell eds. E. Sosa and J. Kim 2001

Peirce, C. S. (1949) *Chance, Love and Logic*, Peter Smith

Plato (1989) *Collected Dialogues* eds. Hamilton, E and Cairns, H. Princeton University Press

Popper, Karl (1974) *The Logic of Scientific Discovery,* Hutchinson & Co (Publishers) London, Seventh Impression

Poundstone, William (2005) *Fortune’s Formula,* Hill and Wang, New York

Ramsey, F. P. (1931). “Truth and Probability”. In *Foundations* ed. Braithwaite.

Ramsey, F. P. (1929) “General Propositions and Causality”, and (1926) *Truth and Probability* in (1990) *Philosophical Papers* ed. D.H. Mellor C.U.P.

Ramsey, F. P. (1991) “Weight or The Value of Knowledge.” pp.285-287, in Galavotti, M.C. ed. *Notes on Philosophy Probability and Mathematics* Bibliopolis

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley. Repr. New York. Dover. 1972.

Schiffer, S (2007) **“**Interest-Relative Invariantism”in *Philosophy and Phenomenological Research* vol. 75 issue 1 July (p 188-195)

Schulz, Maurice (2009) “What Should I believe” The Philosophy of Probability II Conference, June 8th to June 9th, LSE, CPNSS, supported by BSPS.

Shaffer, J (2006) “The Irrelevance of the Subject: Against Subject Sensitive Invariantism.” *Philosophical Studies* 127:87—107

Skyrms, Brian (1984) *Pragmatics and Empiricism* New Haven, CT: Yale University Press,

Skyrms, Brian (1987) “Dynamic Coherence and Probability Kinematics” *Philosophy of Science,* **54*,*** pp 1-20.

Stalnaker (1984) *Inquiry* Cambridge, Mass.: MIT.

Stanley, J. (2006): *Knowledge and Practical Interests*, Oxford: Oxford University Press

Sturgeon, S. (2008): “Reason and the grain of belief.” *Nous*: Oxford University Press

Strawson, P. F. (1971) *Introduction to Logical Theory* London: Methuen.

Weinberg, J., Nichols, S. and Stich, S. (2001), “Normativity and Epistemic Intuitions.”  *Philosophical Topics,* 29, 429-460

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. Child Development, 72, 655-684.

Wheeler, G and Williamson J. (2011) *Evidential Probability and Objective Bayesian Epistemology* In Prasanta S. Bandyopadhyay and Malcolm Forster (eds): *Handbook of the Philosophy of Statistics, Elsevier*

Williamson, J (1999) *Countable Additivity and Subjective Probability* in The British Journal for the Philosophy of Science **50** (1999), 401 - 416

Williamson, T (2000) *Knowledge and its Limits* Oxford University Press p. 218.

Williamson, T (2005) “Contextualism, Subject Sensitive Invariantism and Knowledge of Knowledge” in *The Philosophical Quarterly*, Volume 55, Number 219, April 2005, pp. 213-235(23)

Williamson, T. (2007): *The Philosophy of Philosophy*. Blackwell.

1. “One of the most important milestones in preschool development is the formation of a theory of mind (ToM) at about 4 years of age. ToM encompasses both the understanding that people may have different mental representations of the world that guide their behaviors, and the attribution of mental states (such as beliefs, intentions, or desires) to others and self” Chevalier and Blaye 2006 p.1 [↑](#footnote-ref-1)
2. Evidential probability is a term we are introducing by definition and is simply the degree of belief one ought to have given the evidence. It is an important thesis that evidential probability so defined can vary with what is at stake. We do not mean evidential probability in Kyburg’s sense (Kyburg 2007, Wheeler and J. Williamson, 2011). [↑](#footnote-ref-2)
3. According to Ramsey’s system of subjective utility, the interval between utilities A and B is equal to the interval between utilities C and D if the Subject is indifferent between a bet on *p* that results in A if *p* and D if *~p* and a bet that results in B if *p* and C if *~p* where *p* is an ethically neutral proposition believed to a degree ½ . An ethically neutral proposition believed to degree ½ is one such that a Subject is indifferent between A if *p,* B if ~*p* and B if *p*, A if ~*p*, but prefers A to B for certain. However, this assumes that the Subject’s degree of belief in *p* remains the same across bets at different stake sizes. This assumption may have seemed harmless to Ramsey, and it is necessary to get Ramsey’s interval scale of subjective utility. But it means that the utility defence of stake size invariance is circular. [↑](#footnote-ref-3)
4. It has been suggested that the Gettier intuition, that people in this kind of case do not know, is culturally specific. See Millikan (1993) and for experimental philosophy results see Weinberg, J., Nichols, S. and Stich, S. (2001) [↑](#footnote-ref-4)
5. All references to Socrates are from Plato (1989) [↑](#footnote-ref-5)
6. It might be argued that Descartes’ use of “probable” here is not commensurable with the contemporary mathematical concept of probability. However, it is clear that Descartes is using “probable” to mean less than certain, and that is good enough for our purposes. [↑](#footnote-ref-6)
7. See Ramsey in his discussion of Keynes (1926 pp. 161-162): “there is a one-one correspondence between probability relations and the degrees of belief which they justify.” [↑](#footnote-ref-7)
8. Although this proof involves a kind of packaging principle, it is not the packaging principle that normally offends. It is hard to argue that two items are ever *more* valuable than twice the value of one item. Otherwise you would get “2 for the price of 2” offers that people would find appealing. The proof could go through without this principle, but it would be much more complicated. [↑](#footnote-ref-8)
9. Again we accept that there may be cases where some criteria is agreed upon to settle the bet but which does not provide conclusive evidence. For example, bettors on a horse race may agree to accept the verdict of the judges as final, even if the judges verdict is controversial. In this case, the bettors are in fact betting on the judges verdict, and only indirectly betting on the actual outcome of the horse race. [↑](#footnote-ref-9)
10. There maybe some debate about the truth value of such a statement if no coin is actually flipped at this time. In my opinion the statement then merely fails to express a proposition. In this case the assertability of the conditional falls back on the normative belief in the propositional function. If it is known to be false that a coin was flipped at 11/02/2010 in the actual world, then the conditional becomes a counterfactual. This accounts for the reluctance to say that it is straight forwardly true. The inclination is to say that if the coin had have been flipped on this table top at noon on 11/02/2010 then the probability that it would have landed heads would have been ½. This is merely because the time and place of the coin flipping is irrelevant to the degree of belief in the propositional function. [↑](#footnote-ref-10)
11. An objection to the above is that we are assuming that it is rational to be indifferent between a set of bets and the *median* return; whereas perhaps it is more rational to be indifferent between a set of bets and the *mean* return. We will not answer this objection here, except to say that the expectation derived from the objective probability is the mean return across all possible *sequences* of bets, not the mean return *within a sequence.* Whenever N is high, the mean return within most sequences will be lower than the mean return across sequences. Given that any particular person is going to be concerned with only one sequence of potentially infinite length, the mean return within this sequence is going to be more relevant than the mean return across sequences. The mean return within any particular sequence is certain to be less than the mean return over allsequences. So it is rational to adjust one’s degree of belief according to stake size. [↑](#footnote-ref-11)
12. By “market price” we mean the expectation derived by taking the sum of the products of the degrees of belief of the outcomes by their market prices. So for example, if we say that a subject values a chance which he believes to degree ½ to win a prize of a million pounds at less than its market price, we merely mean that the subject values this chance at less than £500 000. [↑](#footnote-ref-12)
13. On the assumption that inverted stakes are negative and therefore of a different size, then Stake Inversion Invariance is a special case of Stake Size Invariance. However to give this assumption full treatment would be too distracting so I have treated them as different assumptions in this passage. In the rest of the thesis I will assume that Stake Size Invariance entails Stake Inversion Invariance for the purposes of brevity. [↑](#footnote-ref-13)
14. Following the probabilist conventions, the subject of beliefs is designated by capitalised “You”, this is designed to keep in mind that we are talking about a measure of personal belief, not an objective property of an event. [↑](#footnote-ref-14)
15. I do not mean to imply that Skyrms, or anyone else for that matter, seriously proposes eliciting degrees of belief using this game for anything more than pedagogic purposes. The point is rather that stake variant degrees of belief can be only be Dutch Booked if they are elicited in this way. [↑](#footnote-ref-15)
16. The argument here is that the Dutch Book argument relies on the agent to intelligibly give a price at which she is indifferent between buying and selling the bet. Now either this price is given in money terms, in which case, the price won’t reflect her degree of belief, or it is given in utility terms, in which case the bookie is not being offered anything intelligible. [↑](#footnote-ref-16)
17. There is a conceptual problem in assigning a probability for x = ∞ and a value for F(T) when x = ∞. We can solve this problem if we suppose that ∑x = 1 - ∞ fxi( ½xi ) converges at some point y so that y < 1. We can then assume that the probability that the bettor will never run out of money = 1 – y. The value for F(T)on an infinite run could conveniently be supposed to be ½, in which case it will be necessarily true that Ex F (T) < ½, since there will always be a probability of ½W that F(T) = 0. It could be argued that F (T) when N = ∞ is actually greater than ½, since it will included all possible runs where T is always greater than H, but not all runs where H is always greater than T. But even if we assume that F(T) > ½ in an infinite run, then this still will not have much effect on Ex F(T) so long as ½ W ≥ y, since in these cases the probability that F(T) = 1 is counter balanced by the probability that F(T) = 0. We can safely assume that ½ s ≥ y for all W where the difference between W and the stake size is small enough to make any decision worth worrying about. If W = ∞ then the bettor would never run out of money and F(T) = ½ . [↑](#footnote-ref-17)
18. See ch. 2 part 3.3 “The Kelly Criterion”. [↑](#footnote-ref-18)
19. We will assume that zero is both positive and negative. [↑](#footnote-ref-19)