From Relative Frequencies
to Bayesian Probabilities

John Wilcox

THE UNIVERSITY OF AUCKLAND

A thesis submitted in fulfilment of the requirements for the degree of Masters of Arts in Philosophy, the University of Auckland, 2016.

Abstract

It is uncontroversial that evidence regarding frequencies should constrain probabilities or degrees of belief. What is controversial is the question of how this should be done. The *random-worlds method* provides some insight on this question. However, the method by itself faces problems in accounting for rational inferences from samples and in accommodating the uncertainty that agents occasionally have about relevant relative frequencies. One potential response to these problems is to seek alternative probability measures to accommodate such inferences and uncertainty. After surveying two such measures and various problems for them, I find this response wanting. I then offer another response in the form of a theory about rational inferences from samples, one which places an emphasis on the role of intuition in interpreting the probabilistic implications of evidence. The theory is nevertheless consistent with formal methods of statistical analysis in many contexts (such as objective Bayesian analyses of random samples). In accordance with the theory, one may use evidence from samples to form probability distributions about the relevant relative frequencies in a population. I then sketch out how the resulting distributions can be integrated with the insights from the random-worlds method à la the theorem of total probability. This, then, provides an approach to constraining probabilities given evidence about relative frequencies.

Acknowledgements

I thank various people for assisting in the production of this thesis. First and foremost, I thank my supervisors, Brendon, Jeremy and Fred. I am very grateful to Brendon for taking a philosophy student on as one of his own and for introducing me to the world of statistics and its real-world applications. It has been a delight to find someone with a shared passion for probability and for conversing about it. I am very grateful to Jeremy for his patient and painstaking guidance and for acquainting me more thoroughly with the technical world of logic and formal epistemology. Fred has also been kind and helpful in assisting with numerous queries and providing feedback. Additionally, I am grateful to have benefited from conversations with Rachael Briggs, Mark Colyvan, Alan Hájek, Kevin Knuth, Susanna Rinard and Tom Yates.

I also sincerely thank the University of Auckland for its generous scholarship supporting the production of this thesis and also for its funding for research-related travel.

Many thanks also to my fellow graduate lab companions for putting up with (I am told) my excessively loud and fast typing, for giving feedback on my ideas and for occasionally letting me use their names for the fictional characters in this thesis. Ben also wanted a special mention in the acknowledgements section too; so there it is.

Contents

1. Introduction

"I intend to live forever. So far, so good."
– Attributed to Steven Wright

There is something both humorous and philosophically significant about this quote. Some inferences from relative frequencies are reasonable whereas others are not. The relative frequency with which I have lived every day since my birth has so far been 1. But this does not legitimise the inference that I will probably continue to live every day since my birth for the rest of eternity.

In this sense, an examination of how evidence about frequencies might justify various beliefs is tremendously important. The justificatory status of a vast quantity of our beliefs depends on their connection to evidence regarding frequencies. For example, one is typically justified in believing in the general reliability of certain kinds of testimonies and the myriad of propositions which they affirm. We might say that this is because of the observed frequency of similar testimonies that have been reliable in the past. Likewise, suppose one is justified in believing that they are safe to fly in a plane which they have boarded. We might say that this is because of their belief that the relative frequency of dangerous malfunctions among flights is low. We could make a similar connection between frequencies and the justification for a plethora of other beliefs. Arguably, then, we make decisions and even risk our lives in many cases on the basis of judgments underpinned by frequency information. For these reasons, the connection between frequencies and belief is both an interesting and important topic.

From a philosophical perspective, the connection between frequencies and degrees of belief is also of particular interest. A position known as subjective or permissive Bayesianism is popular in formal philosophy, one which has few normative constraints on permissible prior probability distributions. Regardless, these Bayesians often, if not always, advocate constraints on degrees of belief that appeal to empirical relative frequencies or "objective chances" (on some understanding of that concept). There is a case to be made that objective chances are closely connected (if not identical) to relative frequencies of some sort.[1] These constraints arguably are the means by which the evidence proffered by the world impinges on the probabilities of the subjective Bayesian, thereby conferring some objectivity on them. Objective or impermissive Bayesians, like Jon Williamson, likewise advocate constraints by which frequencies or chances should impact subjective probabilities.[2] Hence, frequencies are important to Bayesians of all stripes.

The topic is also relevant to other disciplines, including law, medicine and many other fields. This is illustrated by the court case of a particular drug-smuggler who had his prison sentence extended. This extension was based on the grounds that he *probably* smuggled a certain quantity of drugs given that many *other smugglers* like him had frequently smuggled that quantity.[3] This decision was appealed and vacated partly on the grounds that the statistical evidence was not sufficiently "specific" to the smuggler. Considerable debate ensued about the probative value of statistical evidence, evidence which can have significant real-world implications in numerous parts of society.

For these reasons then, it is important to consider the question of what inferences from evidence about frequencies are rational and what inferences are not. In answer to this question, then, we need a theory about how to assign values to probabilities given evidence about frequencies. If probability is, to use Joseph Butler's oft-repeated words, "the very guide of life", then we might say that such a theory is a good meta-guide to obtaining this reliable guide of life.[4]

Theories of this sort have been articulated by various scholars, including Hans Reichenbach, Henry Kyburg, Choh Man Teng and John Pollock. Reichenbach outlines a theory whereby the observed frequency of *A*s in a sequence of observed *B*s is, under certain circumstances, regarded as an estimate of the limit of the frequency were the sequence to be

[1] For instance, one might understand the objective chance of an experimental set-up to produce a certain outcome to be the propensity of the set-up to produce that outcome where propensities are in turn understood by appeal to relative frequencies (perhaps of the hypothetical sort). For a concise discussion of relationships between propensities and relative frequencies as well as other concepts of probability (such as Lewis's best systems account), see Alan Hájek, "Interpetation of Probability," *The Stanford Encyclopedia of Philosophy* (accessed May 8, 2015), http://plato.stanford.edu/archives/win2012/entries/probability-interpret/.

[2] Jon Williamson, *In Defence of Objective Bayesianism* (Oxford; New York: Oxford University Press, 2010).

[3] Mark Colyvan, Helen M. Regan and Scott Ferson, "Is it a Crime to Belong to a Reference Class," *Journal of Political Philosophy* 9, no. 2 (2001): 168-181.

[4] Joseph Butler, *The Analogy of Religion, Natural and Revealed to the Constitution and Course of Nature*, 20th ed. (New York: Mark H. Newman & Co., 1851), 30. Alan Hájek appeals to this idea of a guide to the guide of life in a different context in an unpublished draft. See Alan Hájek, "Symmetry is the Very Guide of Life," *The Pennsylvania State University*, accessed May 3, 2016, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.318.2402&rep=rep1&type=pdf.

extended indefinitely.[5] Since he identifies such limits with probabilities, he therefore provides a theory by which evidence about observed frequencies determines probabilities.[6] Kyburg and Teng also spell out their system of *evidential probability* which defines probability in terms of relative frequencies among models consistent with a knowledge base. They draw on Chebychev's inequality to estimate the relative frequency of *A*s among a population of *B*s based on the characteristics of a sample.[7] Pollock articulates the concept of a *nomic probability* and draws on classical confidence interval methods as a means for estimating these probabilities based on the characteristics of samples.[8] All of these authors regard these probabilities that are founded on frequency data as "legislative for rational belief", to use Kyburg and Teng's words.[9]

However, none of these authors are Bayesians. I understand orthodox Bayesianism in this context to be the view according to which probabilities are (at least in some contexts) understood as the degrees of belief of some agent.[10] Furthermore, on this view, such probabilities should conform to Andrey Kolmogorov's axioms of probability and should be updated via some rule of conditionalisation upon the receipt of evidence.

Neither Reichenbach, Kyburg, Teng nor Pollock endorse all of these components. None of them, for example, work with an interpretation of probabilities as degrees of belief.[11] Reichenbach interprets probabilities as limiting relative frequencies. Kyburg, Teng and Pollock similarly identify or equate probabilities with the frequency with which a proposition is true among a set of logical models or physically possible worlds. Nevertheless, they bear some resemblances to Bayesian thought. Pollock, for instance, endorses axioms for conditional probabilities which resemble Kolmorgov's axioms and Kyburg and Teng's theory explicitly agrees with Bayesian conditionalisation in certain cases.[12] Despite these resemblances, they are not Bayesian.

---

[5] Hans Reichenbach, *The Theory of Probability*, 2nd ed. (Berkeley: University of California Press, 1949).

[6] Reichenbach's views are perhaps more nuanced than various commentators indicate. He articulates his theory of probability with respect to both hypothetical limits and "practical limits" in the actual world, that is, "sequences that behave, in a finite length…, in a way comparable to a mathematical limit". Reichenbach, *The Theory of Probability*, 347-348 and 447.

[7] Henry Kyburg and Choh Man Teng, *Uncertain Inference* (New York: Cambridge University Press, 2001).

[8] John Pollock, "The Theory of Nomic Probability," *Synthese* 90, no. 2 (1992): 263-299.

[9] Kyburg and Teng, *Uncertain Inference*, 226.

[10] Michael Strevens, "Bayes, Bayes' Theorem, Bayesian Approach to Philosophy of Science," vol. 1 of *Encyclopedia of Philosophy*, ed. Donald M. Borchert, 2nd ed. (Detroit: Macmillan Reference USA, 2006), 496.

[11] However, Pollock does work with the concept of a "mixed physical-epistemic probability" which somewhat resembles a Bayesian degree of belief.

[12] John Pollock, "Probable Probabilities (with Proofs)," accessed December 6, 2015, http://johnpollock.us/ftp/PAPERS/Probable%20Probabilities%20with%20proofs.pdf and Kyburg and Teng, *Uncertain Inference*, 263. Throughout this thesis, I occasionally refer to Pollock's work that has remained

However, one theory in the literature does stand out as at least quasi-Bayesian, namely, the *random-worlds method* articulated by computer scientists Fahiem Bacchus and his colleagues. Like the theories of Reichenbach, Kyburg, Teng and Pollock, their theory outlines how probabilities are to be constrained by information regarding frequencies.

Their approach might be interpreted as Bayesian for a few reasons (rightly or wrongly). One is that their focus is on degrees of belief or, as they and others say, *subjective probabilities*.[13] Furthermore, they explicitly mention that they "use the Bayesian approach".[14] In this sense, these degrees of belief gain their values by the use of some characteristically objective Bayesian techniques – the use of an indifferent probability distribution which is then updated by conditionalisation.[15]

Yet their approach does not obviously fit into the typical Bayesian story. For one, although they use prior and posterior probability distributions, they do not explicitly interpret these distributions as representing degrees of belief (or indeed anything else). Instead, these distributions are used as a means of *defining* degrees of belief in terms of the frequency with which propositions are true among the models that are consistent with a knowledge base. They therefore only offer a static constraint on degrees of belief, a constraint specifying that if one has a certain knowledge base at a given time, then their degrees of belief should be a certain way at that time. Additionally, the probabilities of interest are computed entirely on the basis of a set of categorical beliefs, the principle of indifference and axioms of probability. There is no appeal to, say, the subjective intuition or considerations of simplicity that might determine the probabilities of typical Bayesians, particularly of the subjectivist sort.

Regardless, if one is looking for a relatively comprehensive Bayesian theory of how degrees of belief should be constrained by information about relative frequencies, the closest candidate in the literature is Bacchus et al.'s theory.

Furthermore, the theory provides a wealth of insight as to how to do this. For example, it tells one both *that* and *why* they should equate their degree of belief with statistical statements about some reference classes but not others and should regard two properties as probabilistically independent unless given reason to think otherwise. It also outlines what are putatively reasonable ways for assigning values to degrees of belief in other situations where our intuitions about the appropriate value may not be so clear.

---

unpublished since his death. This is because, although unpublished, this work is both insightful (in my opinion) and reflective of his later thinking.

[13] Bacchus et al., "From Statistical Knowledge Bases," 75.

[14] Ibid., 77.

[15] Ibid., "From Statistical Knowledge Bases," 99.

For these reasons, a Bayesian might find in Bacchus et al.'s theory a useful account of how one should assign values to degrees of belief on the basis of frequency information.

Or not. Indeed, the theory only provides guidance under conditions which are often so unrealistic that it might make the theory irrelevant to many practical situations. The method only specifies what the degrees of beliefs should be *if* one is *certain* of the statements in the knowledge base (or prepared to accept such statements as if they are certain). Examples of such statements are those of the form "this *a* is a *B*" and "the frequency with which *B*s are *A*s is *r*" where *a* is some object, *A* and *B* are some predicates and *r* some real number in the unit interval $[0, 1]$.

This supposition of (quasi-)certainty makes the situations envisaged by the method atypical of many actual situations in which agents are curious about the degrees of belief that they should have. For example, suppose a doctor has found that 90 of 100 of their patients with a particular set of symptoms have had a particular ailment. On the basis of this sample of patients, she is 80% confident in the proposition that approximately 90% of the people in her locality with those symptoms have that ailment. But the doctor is not certain of the relative frequency so as to accept this proposition as if they are certain. Yet surely her confidence about the uncertain relative frequency should somehow affect her degree of belief that the next patient that she sees with those same symptoms will have that ailment. How, then, should her confidence do this? The random-worlds method by itself is silent here. It only specifies that *if* one accepts that the relative frequency is such-and-such, *then* the relevant degree of belief is so-and-so.

Given that the random-worlds method arguably has numerous favourable features and provides some useful guidance, can we salvage this guidance and integrate it into other realistic contexts where the agent is uncertain about the relevant relative frequencies?

This thesis argues affirmatively, albeit via a somewhat long-winded route. I argue that the method can be supplemented with a theory of so-called *inverse inference* that acknowledges the role of intuition in interpreting the significance of sample evidence. The theory accords with both the subjective Bayesian thought that inference involves intuitive and non-formal elements and the objective Bayesian thought that not all such elements are rationally permissible. However, this theory also accommodates formal and established methods of statistical inference, such as objective Bayesian data analysis of random-samples. Yet it also accommodates some of the commonplace inferences that agents make about relative frequencies on the basis of evidence (particularly from so-called *non-probability samples*) thus allowing them to thereby have their own probability distribution over the possible relative frequencies of interest. I then argue that such distributions can be integrated with the insights from the random-worlds method via the theorem of total probability to thereby constrain degrees of belief.

1.1. Outline

So that is my thesis. Here is how I argue for it.

The second chapter discusses some important topics, concepts and terminology. Crucial in this respect is the formal language utilised by the (hypothetical) agent in making various inductive inferences. A taxonomy of these inductive inferences is presented to map out the subject matter of this thesis. In this regard, the end of the chapter examines the degrees of belief, credences or subjective probabilities which are the outputs of such inferences.

In Chapter 3, I outline the random-worlds method and, in Chapter 4, I then explore how the random-worlds method may be evaluated before delineating some putatively favourable features which it possesses. Here, I argue that while the method is insightful, it is also problematic. In particular, it faces difficulties in accommodating important cases of uncertainty about relative frequencies (like the aforementioned case of the doctor) and validating rational inferences from samples.

The fifth and sixth chapters outline and assess one response to these difficulties, one that turns to alternative formal measures for inductive inference. I describe two of these measures in Chapter 5. In Chapter 6, I then evaluate these measures and survey some problems confronting them. I ultimately conclude that we currently lack a general, objective and formal method of inference to replace the measure used in the random-worlds method, particularly where non-probability samples are concerned.

The seventh chapter then articulates a theory about certain kinds of inferences from samples, especially in the context of non-probability samples. This theory delineates two dimensions along which the rationality of such inferences are assessed. It argues that certain intuitions may be trusted in making these inferences. In accordance with the theory, individuals may, on the basis of sample evidence, form probability distributions over the space of possible relative frequencies in a population.

Chapter 8 then examines how such distributions may be integrated with the prescriptions of the random-worlds method to thereby constrain degrees of belief. In particular, it does this by appealing to the spirit of the theorem of total probability.

Having outlined the structure of the thesis, a word is in order to clarify what the reader can expect of it. It should be clear that this thesis concerns debates in both philosophy and statistics (as well as certain other disciplines). This is because branches of these disciplines are closely connected through their common interest in reasoning in the face of uncertainty. Such a connection caused philosophers Colin Howson and Peter Urbach to claim that the branch of statistics known as *statistical inference* in fact "naturally belongs to philosophy of science."[16] Indeed, the thesis's philosophical and statistical focus follows in the footsteps of prominent philosophers, including Reichenbach, Pollock and Kyburg. Furthermore, that the thesis treats subjects in both disciplines is to be expected since the thesis falls under purview of both the philosophy and the statistics departments of the University of Auckland. Consequently, the thesis tries to navigate two disciplines in a way that is accessible to both and neglectful of neither. Given this background, I hope that the thesis is not unduly criticised by the philosophically inclined as too statistically oriented nor by the statistically inclined as too philosophically oriented.

[16] Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, 3rd ed. (Chicago: Open Court, c2006), 91.

2. The Epistemological Backdrop

To help us to understand the random-worlds method and related theories, it would be helpful to locate it in the context of a wider set of topics and concepts. This is particularly the case since this thesis is not just written for a philosophical audience, but also a wider audience who may be less familiar with some of the topics that follow.

2.1. Language and Semantics

To explore formally how information about frequencies can constrain degrees of belief, we will need a formal language and some logical notation. I have chosen notation based on its elegance and similarity to other notation found in the literature. I have also included an appendix with a translation key for the notation.

The topic of language is important in this thesis and in the study of logic more generally. Logicians utilise formal languages that represent the world which some agent (whether actual or hypothetical) is interested in reasoning about. The choice of language is significant since it delineates the categories which are deemed important and relevant to reasoning in the agent's situation. In Chapter 6, we shall see that, according to some theories of inductive inference, the choice of language can also bear on the values of probabilities.

Let us now survey the formalism used in this work.

2.1.1. The Basics of a First-Order Language

In this thesis, we will use the language of Bacchus et al., albeit occasionally with different symbolism.[17] The language, symbolised as $\mathcal{L}^{\approx}$, expresses both statistical and first-order information, that is, information in first-order logic.

While it is neither possible nor desirable to examine in-depth the typical syntax and semantics of first-order logic here, I will informally explain some notation so that non-specialist audiences can have a rough understanding of the notation.

First-order logic involves various symbols known as constants, variables and predicates.

[17] Bacchus et al., "From Statistical Knowledge Bases," 95.

To understand these symbols, it is useful to understand the semantic concept of *domain objects*. Domain objects are represented with lower-case ds and the *domain* is the set of domain objects $\{d_1, d_2, \ldots, d_N\}$ where N is the size of the domain. The domain is what the language is *about*.

The language used in the random-worlds method can contain so-called *constants* which are represented with lower-case letters a, b, c ... and which are linguistic items used to denote the non-linguistic domain objects. (Note that I am following the logical tradition of counting persons and events as "objects" too.) Suppose that the agent is interested in reasoning about whether his friend Aaron went to a funeral given that his friends Brendon and Claire went. The agent may then have a language with three constants denoting the respective friends or objects: a, b and c for Aaron, Brendon and Claire respectively. If we suppose the agent is only interested in reasoning about these three people and no other objects, then we suppose that the domain has only three objects, each of which is uniquely named with a constant.

The language also contains *variables*, that is, lower-case letters toward the end of the alphabet x, y, z .... Variables are syntactic items which serve two purposes in logic. First, they are used to denote a particular object that is unspecified. For instance, a logician might ask a student to suppose that "x went to the funeral". Second, they are used to, in Stewart Shapiro's words, "express generality" as when a logician might endorse a proposition such as, "For all *x*, if *x* is the agent's friend, then *x* went to the funeral" (this is one way of saying that all of the agent's friends went to the funeral).[18]

The language contains various *predicate symbols*, represented with upper-case letters, $A, B, \ldots, F, G,$ .... Predicate letters correspond to sentences when combined with constants or variables. For example, a language may contain the predicate F where Fx is the sentence "x went to the funeral" and Fc is the sentence "Claire went to the funeral." Predicates can also express relations between objects. For example, if Gxy expresses the proposition that x is ganglier than y, then Gab expresses a proposition about the relation between Aaron and Brendon, namely, the relation of Aaron being ganglier than Brendon.

The language of Bacchus et al. also allows the use of function symbols, but these are not of particular importance or relevance here, so I will omit an explanation of them.

First-order logic also involves various quantifiers ($\forall$ and $\exists$) and logical connectives ($\supset$, &, $\vee$ and $\equiv$). An informal translation of these and all the other logical symbols in this thesis can be found in the appended glossary.

[18] Stewart Shapiro, "Classical Logic," *The Stanford Encyclopedia of Philosophy*, accessed April 20, 2016, http://plato.stanford.edu/archives/win2013/entries/logic-classical/.

2.1.2. Statistical Quantifiers

An important feature of the logic of Bacchus et al. are statistical quantifiers involving *proportion expressions*.

Following Bacchus et al., we call symbols of the form $\%x(\alpha(x))$ *unconditional proportion expressions* where α is any formula mentioning the variable x.[19] Such an expression is interpreted as a rational number in the interval $[0,1]$ that denotes the proportion of domain elements satisfying some formula α (i.e. the proportion of objects for which the formula is true). For example, the expression $\%x(Fx)$ where the formula Fx stands for "x is went to the funeral" refers to the proportion of domain objects that went to the funeral. Unconditional proportion expressions can also contain more than one variable, so that, for example, $\%xy(Gxy)$ denotes the proportion of pairs of domain objects that stand in the "ganglier than" relation.

Again following Bacchus et al., we call symbols of the form $\%x(\alpha(x)|\beta(x))$ *conditional proportion expressions* where α and β are any two formulas mentioning the variable x.[20] This represents the proportion of domain elements satisfying the formula $\beta(x)$ that also satisfy $\alpha(x)$. We can understand these proportion expressions (and, in a sense, unconditional proportion expressions) as being relative frequencies. The formula in place of $\beta(x)$ is often called the *reference formula* which corresponds to *reference class* of objects which share a *reference property*. The formula in place of α is often called the *target formula* which corresponds to *target* or *attribute class* of objects which share a *target property*. An example of a conditional proportion expression is the statement that 50% of humans are male. Formally, we can express this with the symbol $\%x(Mx|Hx) = 0.5$ where the reference formula Hx stands for "x is a human" and the target formula Mx stands for "x is a male." Here, the reference class is the set of all objects with the reference property of being a human and the target class is the set of all objects with the target property of being a male.

Note that the class of proportion expressions also includes expressions denoting rational numbers and sums and products of other proportion expressions (such as, for example, $2\%x(Ax) + 3/5$).[21]

[19] Bacchus et al. use the notation of the form $||\psi||_X$ for unconditional proportion expressions. I have used different notation which, I believe, more closely resembles typical first-order logic notation.
[20] Bacchus et al. use the notation of the form $||\psi|\theta||_X$ for conditional proportion expressions.
[21] Bacchus et al., "From Statistical Knowledge Bases," 95.

From here on, the segment of the proportions expressions which specifies the binding of the variables will be omitted, so a formula such as $\%x(\alpha(x)|\beta(x))$ will simply read as $\%(\alpha(x)|\beta(x))$. This omission will be harmless as the context will indicate the binding of the variables.

2.1.3. Approximate Equalities and Inequalities

Bacchus et al.'s formulas for proportion expressions are articulated through the use of approximate equalities (and inequalities). Hence, they typically use a statement such as $\%(\alpha(x)|\beta(x)) \approx 0.8$ in place of one such as $\%(\alpha(x)|\beta(x)) = 0.8$. This is a major innovation of their logic.

The use of this language reflects the fact that strict equalities may often inappropriately imply that the number of objects satisfying, say, $\beta(x)$ has to be a multiple of some number (which is 5 in the case of $\%(\alpha(x)|\beta(x)) = 0.8$) when information supporting this implication is lacking.

The degree of approximation is specified by a tolerance factor τ or, in other words, a degree of slack for the approximation. The tolerance factors are intended to account for "measurement error, sample variations, and so on."[22] A formula such as $\%(Hx|Jx) \approx_i 0.8$ states that the proportion of jaundiced patients with hepatitis is within some tolerance factor τ_i of 80% where Jx stands for "x has jaundice" and Hx stands for "x has hepatitis".[23] The indices on the approximation symbol indicate that the tolerance may vary in different contexts; it might be 0.01 in one context but 0.05 in another. Strict equalities and inequalities like $=$ and $\leq$ are a special case where the tolerance factor is 0.

Nevertheless, Bacchus et al. state that the tolerance factor is "very small, but unknown" and hence approximations are formalised with a vector of tolerances $\vec{\tau} = \{\tau_1, \dots, \tau_m\}$ such that $\tau_i > 0$.[24] We will discuss these tolerance vectors more in Chapter 3.

Approximate inequalities of the forms $\preccurlyeq$ and $\succcurlyeq$ are also used in place of strict inequalities of the forms $\leq$ and $\geq$.

Subscripts on the approximate equalities or inequalities which specify the tolerance will be omitted unless doing so could cause confusion.

[22] Bacchus et al., "From Statistical Knowledge Bases," 95.
[23] Ibid., 77.
[24] Ibid., 96.

Proportion formulas then relate proportion expressions to each other via (approximate) equalities and inequalities. They sometimes state the (possibly approximate) value of proportions expressions. For example, $\%(Hx|Jx) \approx 0.8$ is a proportion formula.

2.1.4. A Recursive Definition of $\mathcal{L}^{\approx}$

Following Bacchus et al., we can provide an informal recursive definition of the language $\mathcal{L}^{\approx}$.[25] $\mathcal{L}^{\approx}$ consists of sets of terms, proportion expressions and formulas. Let ϕ be a finite first-order vocabulary of predicate, function and constant symbols and χ be a set of variables.

The set of *terms* in $\mathcal{L}^{\approx}$ is the least set containing χ and the constant symbols in ϕ which is closed under function application (meaning that if f is a function symbol with arity *r* and $t_1, \dots, t_r$ are terms, then $\mathrm{f}(t_1, \dots, t_r)$ is also a term).

The set of *proportion expressions* in $\mathcal{L}^{\approx}$ is the least set which:

a) includes the rational numbers,
b) includes *proportion terms* of the form $\%X(\alpha|\beta)$ and $\%X(\alpha)$ for formulas $\alpha, \beta \in \mathcal{L}^{\approx}$ and a finite set of variables $X \subseteq \chi$ and
c) is closed under addition and multiplication.

The set of *formulas* in $\mathcal{L}^{\approx}$ is the least set which:

a) includes *atomic formulas* of the form $At_1 \dots t_r$ where A is a predicate in ϕ of arity *r* and $t_1, \dots, t_r$ are terms,
b) includes *proportions formulas* of the form $\zeta \preccurlyeq_i \zeta'$ and $\zeta \approx_i \zeta'$ where ζ and ζ' are proportion expressions and i is a natural number and
c) is closed under conjunction, negation and first-order quantification.

[25] Bacchus et al., "From Statistical Knowledge Bases," 96.

2.1.5. Semantics and Models

The semantics for a logic specifies an interpretation of the formal language or the conditions under which statements in the language are true.

The semantics for the language $\mathcal{L}^{\approx}$ is largely the same as the standard semantics for first-order logics. While it is neither possible nor desirable to thoroughly outline and explain the standard semantics here, I will make a few comments, some of which are intended for a non-specialist audience.

An important concept in semantics is that of a *model* over the domain $\{d_1, \ldots, d_N\}$. Bacchus et al. call this "a possible world."[26] The set of all models form the *space of possible outcomes* or the *sample space*. Models specify *denotations* for symbols in the language. In the standard semantics for first-order logic, a denotation for a constant or predicate symbol is the set of domain objects which that symbol applies to. Models, then, just specify which symbols apply to which domain objects. So, for instance, suppose we have a language with one predicate *F* to symbolise going to the funeral and three constants *a*, *b* and *c* symbolising the names Aaron, Brendon and Claire. For a domain of three objects $\{d_1, d_2, d_3\}$, there are 216 models constructible from this language.[27] Any formula in the language will be true or false only relative to a model. For example, consider these four models (or possible outcomes) from the 216 models constructible from the language and domain:

M_1: "*F*" denotes $\{d_1\}$, "*a*" denotes d_1, "*b*" denotes d_2 and "*c*" denotes d_3.
M_2: "*F*" denotes $\{d_2\}$, "*a*" denotes d_2, "*b*" denotes d_1 and "*c*" denotes d_3.
M_3: "*F*" denotes $\{d_2, d_3\}$, "*a*" denotes d_1, "*b*" denotes d_2 and "*c*" denotes d_3.
M_4: "*F*" denotes the null set {}, "*a*" denotes d_1, "*b*" denotes d_2 and "*c*" denotes d_3.

The atomic formula *Fa* is true according to the models M_1 and M_2 since *F* and *a* apply both apply to one particular object in each case. According to these models, then, Aaron went to the funeral. *Fa* is not true in M_3 or M_4, however, since *F* and *a* do not both apply to one

[26] Bacchus et al., "From Statistical Knowledge Bases," 78.

[27] The calculation for this is as follows. Each constant of a language denotes one of the *N* objects in the domain. So the number of models for a language containing only constants is N^C where *C* is the number of constants. Each one-place predicate in the language refers to one of 2^N sub-sets of the domain. For a language where there are *k* predicates and no other symbols, there are $(2^N)^k$ models. When the language contains only *C* constants and *k* one-place predicates, then the number of models is $(N^C)(2^{Nk})$. In our example, there are three domain objects and the language contains three constants and one one-place predicate; hence, the total number of models is $(N^C)(2^{Nk}) = (3^3)(2^{(3)(1)}) = 216$.

particular object in either model. Note that a set of models for a language can be isomorphic; M_1 and M_2 are examples of this since they only differ by the permutation of d_1 and d_2. Such models will generate the same truth values for the formulas in a language. Hence, in both models, the true formulas are the atomic formulas Fa, $\sim Fb$ and $\sim Fc$ as well as certain other complex formulas like $Fa \,\&\, \sim Fb$ and proportion formulas like $\%(Fx) = \frac{1}{3}$. Hence, a probability measure over the sample space translates into a probability measure over all the statements expressible in the language.

That concludes our examination of the formal language used in this thesis.

2.2. A Taxonomy of Inductive Inferences

Ultimately, language is important because it is the means by which the agent represents the world in order to makes inferences about it. We will now consider the inductive inferences that are of interest in this thesis.

It is useful here to distinguish various types of inductive inference. According to the below taxonomy, each type of inference is characterised by the ascription of probabilities to certain types of propositions on the basis of certain types of evidence. The kinds of inference vary by virtue of the various types of propositions and types of evidence which they involve.

These inferences often involve populations or samples. A population is a non-empty set of all of the objects satisfying some formula. For example, the population of Wellingtonians is the set of all objects satisfying the formula Wx where Wx stands for "x is a Wellingtonian". Sometimes, however, a population is simply the set of all domain objects; the context will make it clear when this is the case. A sample is a non-empty subset of a population that is of interest for one of two reasons: 1) the proportion of objects in the sample satisfying a target formula α is not known but is to be estimated or 2) the proportion of objects in the sample satisfying α is known and this serves as evidence for estimating the extent to which other objects satisfy α or a another formula that denotes a similar property.[28] The sample need not be one that is selected randomly from the population.

We will see examples of these concepts in the below taxonomy of inductive inferences:

[28] This conception of a sample also resonates with the conception of a sample in Kyburg and Teng, *Uncertain Inference*, 175-176.

- A *direct inference* is an inference from evidence about a population (or set of populations) to probabilities about a sample. It assigns probabilities to the possible proportions of a sample satisfying a target formula given some evidence about the proportion of objects in some population(s) that satisfy the formula. For example, suppose the agent knows that 80% of the population of Wellingtonians smoke (i.e. that $\%(Sx|Wx) = 0.8$ where the target formula Sx stands for "x is a smoker"). She then randomly samples three Wellingtonians. Given her knowledge, she assigns a probability of 0.008, 0.096, 0.384 and 0.512 respectively to the propositions that 0/3, 1/3, 2/3 and 3/3 of the sampled objects smoke (i.e. satisfy Sx). In this thesis, we will typically consider a special case of direct inference which we can call *singular direct inference*. This is where the sample consists of one object and one is estimating whether it does or does not satisfy the target formula given some evidence about the proportion of objects in the relevant population(s) that satisfy the formula. In other words, one is estimating the proportion of the sample satisfying the formula where the possible proportions are only 0/1 or 1/1. A case of singular direct inference is when the agent knows that $\%(Sx|Wx) = 0.8$ and she randomly samples one Wellingtonian to assign a probability of 0.8 to the proposition that the sampled Wellingtonian smokes. Direct inference can also involve multiple reference classes or populations, as we shall see in Chapter 4.

- A *predictive inference* is an inference from evidence about a sample to probabilities about another non-overlapping sample. More specifically, it assigns probabilities to the possible proportions of a sample satisfying a target formula given evidence about the proportion of objects in another sample satisfying that formula or one denoting a similar property. This includes *singular predictive inference,* a special case in which the agent is inferring the probability that just one sampled object satisfies the target formula. An example of singular predictive inference is when one infers that the next sampled philosopher will probably be weird since all of the philosophers in another sample of philosophers proved to be weird.

- An *analogical inference* is an inference about the probability that a set of objects are similar in some respect on the basis of evidence about them sharing other similarities or dissimilarities. In this respect, one might infer that one object satisfies some formula α since it is similar in various respects to another set of objects which also satisfy α. For example, suppose one is trying to determine the probability that a chimpanzee sometimes feels pain or, in other words, that it satisfies the formula Px where Px stands for "x sometimes feels pain". The agent infers that the chimpanzee probably does satisfy Px because it is neurologically, physiologically and behaviourally similar to humans who do sometimes feel pain (i.e. satisfy Px). Hence, the agent's inference is that humans and this chimpanzee are similar in respect of sometimes feeling pain on the basis of the evidence that they share neurological, physiological and behavioural similarities.

- An *inverse inference* is an inference from evidence about a sample to probabilities about a population. It assigns probabilities to the possible proportions of a population satisfying a target formula given evidence about the proportion of objects in a sample satisfying that formula or another formula representing a similar property. For example, suppose the agent randomly samples five balls from an urn of 20 balls. The sampled balls turn out to be green and satisfy Gx where Gx stands for "x is green". On this basis of the sample from the population of balls in the urn, the agent assigns probabilities to the propositions that 6/20, 7/20,…,19/20 or 20/20 of the population are green (i.e. satisfy Gx). Each of these propositions is representable with a proportion formula of the form $\%(Gx|Bx) = f$ where Bx stands for "x is a ball in the urn" and *f* is some rational number in the unit interval $[0, 1]$.

- A *universal inference* is an inference from evidence about a sample to a high probability for a universal generalisation for a population. A universal generalisation is a technical concept in logic. These generalisations are typically symbolised with formulas of the form $\forall x(Fx)$ and $\forall x(Fx \supset Gx)$ for some predicates F and G. A universal generalisation corresponds to a proportion formula stating that all of the objects in a particular population satisfy a target formula.[29] (Note that, importantly, a universal generalisation need not be about an infinite number of objects.) A canonical

[29] When considering generalisations of the logical form $\forall x(Fx \supset Gx)$, this is true only if the population has at least one member satisfying the antecedent formula, otherwise a universal generalisation such as "All *F*s are *G*s" can be trivially true by virtue of nothing being an *F* whereas a unit valued conditional proportion formula cannot.

> example of a universal inference is when one infers that probably all emeralds are green on the basis of sample of emeralds, all of which were found to be green.

These inferences are not all mutually exclusive since, for example, an inverse inference can involve a universal inference. Note also that this taxonomy resembles the definition of various inferences given by others, especially Rudolf Carnap.[30] Regardless, these other scholars construe or label these kinds of inference somewhat differently. Kyburg and Teng, for example, construe direct inference as being an inference to a "nonprobabilistic conclusion".[31] Pollock also calls inverse inference *statistical induction* and universal inference *enumerative induction*.[32] Statisticians use the term *parameter estimation* to refer to inverse inference (as well as distinct but closely related kinds of inference).

The random-worlds method is a theory of direct inference, but the other forms of inductive inference, and the challenges they face, will be relevant later in this thesis.

Each characterisation of the aforementioned inferences alludes to the putative probabilistic nature of induction. But an important question arises as to how these probabilities are to be interpreted. Various interpretations are possible, and the most well-known scholars on direct inference (Reichenbach, Kyburg, Pollock and Bacchus et al.) all utilise very different conceptions of probability in their accounts of induction. Nevertheless, one conception of probability is of particular importance in this thesis, that is, the *subjective interpretation of probability* as *credence*. Credences are the topic of the next section.

---

[30] Rudolf Carnap, *Logical Foundations of Probability* (London: Routledge and K. Paul, 1951), 207.
[31] Kyburg and Teng, *Uncertain inference*, 175.
[32] John Pollock, *Nomic Probability and the Foundations of Induction* (New York: Oxford University Press, 1990), 21 and 36.

2.3. Credence Functions and Subjective Probabilities

Credences are the degrees of belief people have towards propositions.[33] For example, an agent may have a high credence in the sun rising tomorrow morning but a low credence in seeing an alien spaceship in their lifetime.

The aim of this thesis is to explore how evidence about frequencies should constrain the credences of an agent in realistic contexts.

Credences are modelled by functions from propositions to the unit interval [0,1]. These are called *credence functions*. So if the function $P(.)$ models the agent's credences, then the number that it assigns to the sentence expressing the proposition that the sun will rise tomorrow is close to 1 and the number that it assigns to the sentence expressing the proposition that the agent will see an alien spaceship is close to 0. Note that we will represent credences with sharp numerical credence functions, that is, functions that assign numerically exact values like 0.6000.... This contrasts to using an "indeterminate probability" approach à la Isaac Levi.[34] On the indeterminate probability approach, the agent's credence for a proposition is representable with not just one credence function, but rather with a *set* of credence functions spanning some interval of values that are assigned to the proposition by different functions in the set.

I will assume that sharp credence functions are appropriate models for credences. The reason for this is partly because debates concerning the merits and weaknesses of indeterminate or other representations of credences are too large to be given sufficient attention in this thesis.[35] But it is also partly because sharp representations of credences are the current orthodoxy in both formal epistemology and Bayesian statistics; so it should not be too objectionable that this thesis accords with orthodoxy.[36] In doing this, however, I do not

[33] I have chosen to use the philosophical term "credence" instead of "degree of belief" or "degree of confidence" in this thesis. The reason for this is that I sometimes wish to speak of an agent having a low or indifferent credence in some proposition in a way which may be misleading if "credence" is substituted by "degree of belief". For example, an agent may have a low credence of 0.01 that they will see an alien spaceship in their lifetime. Yet it seems misleading to say that they have a low degree of "belief" in this proposition since there is arguably a sense in which they have no degree of belief in the proposition at all. Rather they just have a high degree of *disbelief* in the proposition or a high degree of belief in its negation.

[34] Isaac Levi, "Imprecision and Indeterminacy in Probability Judgment," *Philosophy of Science* 52, no. 3 (1985): 390-409.

[35] For a discussion of indeterminate probabilities, as well as a list of arguments for taking them seriously, see Alan Hájek and Michael Smithson, "Rationality and Indeterminate Probabilities," *Synthese* 187, no. 1 (2012): 33-48. Certain others have argued that agents do not or should not have indeterminate probabilities, an example being Adam Elga. See Adam Elga, "Subjective Probabilities should be Sharp," *Philosophers' Imprint* 10, no. 5 (2010): 1-11.

[36] That this is orthodoxy in formal epistemology appears to be affirmed in a number of places. For example, Bradley Seamus speaks of the "precise attitudes" that "orthodox probability requires" and its "insistence that

suppose that credences are sharply graded entities (thus corresponding to what Levi would call *determinate probabilities*), but rather that sharp credence functions are at least *adequate idealisations* for credences in this thesis.

It is occasionally useful to distinguish between the agent's "prior" credences and "posterior" credences. The distinction concern that change between an earlier and later credal state upon the receipt of new evidence. The agent's credences before receiving the evidence are her prior credences and the credences thereafter are her posterior credences. These are modelled by a prior credence function $P(.)$ and a posterior credence function $P'(.)$. The difference between P and P' reflects what one has learned from the evidence. For example, suppose the agent doubts that she will get a particular job, so we can say that $P(Ja) = 0.1$ where Jx stands for "x will get the job" and a stands for the agent. Suppose that the agent then receives a phone call from the potential employer who promises her that, surprisingly, she will get the job. She then updates her credence so that she is now confident that she will get the job, so $P'(Ja) = 0.99$.

Sometimes, however, it is useful to refer to the agent's credences without reference to prior or posterior states relative to the receipt of some evidence. In this case, we will simply use $P(.)$ to refer to this person's credences without intending to imply that it is prior to the receipt of some particular evidence or to some updated credence function.

Credences are also known as *subjective probabilities*. The interpretation of probabilities as credences dominates formal epistemology.[37] In the thesis, I sometimes use the terms 'credence' and 'probability' interchangeably since this practice of interchangeability is common in relevant debates and some concepts are also more naturally describable and comprehensible with one term rather than another (such as a "prior probability distribution" rather than a "prior credence distribution"). This interchangeability also reflects the common assumption that credence functions obey the standard axioms of probability. These are the following:

(A1) All probabilities are between 1 and 0, i.e. $0 \leq P(\alpha) \leq 1$ for any α.

(A2) Logical truths have a probability of 1, i.e. $P(L) = 1$ where L is any logical truth.

states of belief be represented by a single real-valued probability function." Seamus Bradley, "Imprecise Probabilities," *The Stanford Encyclopedia of Philosophy*, accessed June 21, 2016, http://plato.stanford.edu/archives/sum2015/entries/imprecise-probabilities/. Hájek and Smithson also note that "the ideal Bayesian agent" is one that "assigns perfectly sharp credences to all propositions." Alan Hájek and Michael Smithson, "Rationality and Indeterminate Probabilities," 34.

[37] Debates about whether the subjective interpretation of probability or any other interpretation of probability is plausible are beyond the scope of this thesis.

(A3) Where α and β are two mutually exclusive formulas, the probability of α or β $(\alpha \vee \beta)$ is the sum of their respective probabilities, i.e. $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$.[38]

Actual agents cannot realistically conform to these axioms, but there are putatively good arguments as to why such conformity is an ideal to which rational agents should strive.[39]

Unfortunately, however, there is ambiguity in the adjective "subjective" that is used to describe a (prior) probability distribution.[40] It can refer to the fact that the distribution is interpreted as the credences of some subject. Alternatively (or additionally), it can refer to the lack of constraints forcing the distribution to take some form (such as the principle of indifference). The first meaning corresponds to *subjectivism* as a semantic claim about the meaning of probability statements.[41] The second meaning corresponds to *subjectivism* as a normative claim about what credences are rationally permissible.[42] Some accept subjectivism in its semantic sense while rejecting subjectivism in its normative sense, an example being objective Bayesians such as E.T. Jaynes.[43] In this thesis, I call subjectivism in its normative sense *subjective Bayesianism* and refer to the probabilities of subjectivism in its semantic sense simply as subjective probabilities (or just probabilities) without thereby implying a relative paucity of constraints on such probabilities.

[38] William Talbott, "Bayesian Epistemology," *The Stanford Encyclopedia of Philosophy*, accessed July 22, 2014, http://plato.stanford.edu/archives/sum2011/entries/epistemology-bayesian/.
[39] There are some critics of the aforementioned orthodox axiomatisation of probability in terms of unconditional probabilities. See, for example, Alan Hájek, "What Conditional Probability Could Not Be," *Synthese* 137, no. 3 (2003): 273-323. Yet again, this is another topic that is beyond the scope of this thesis.
[40] Mike Titelbaum makes this point in Mike Titelbaum, "Fundamentals of Bayesian Epistemology: Chapter 5," *Fitelson,* accessed May 8, 2015, http://fitelson.org/bayes/titelbaum_ch5.pdf.
[41] See this construal of subjectivism in Hájek, "Interpretations of Probability."
[42] See this construal of subjectivism in Vincenzo Crupi, "Confirmation," *The Stanford Encyclopedia of Philosophy*, accessed May 5, 2016, http://plato.stanford.edu/archives/fall2015/entries/confirmation/.
[43] E.T. Jaynes, *Probability Theory: The Logic of Science* (New York: Cambridge University Press, 2003).

3. The Random-Worlds Method

This chapter examines the main features of the random-worlds method. We have already outlined a version of the language that is employed in the method. What remains to be seen is what we might call the probabilistic features of the method.

In their 1996 article, Bacchus et al. focus on a particular sort of credence that I will call a *random-worlds credence* (note, however, that they use the term "degree of belief" instead of "credence").[44]

While they do not specify anything about the "agent" who is presumed to have such credences, they do assume that "all" of the agent's knowledge is encapsulated in a knowledge base KB.[45] They also do not explicitly define or outline the features of the KB. Regardless, I will assume that a KB is a finite set of propositions which the agent is certain of or is prepared to treat as if she is certain of. I will refer to the KB sometimes as a set of propositions and sometimes as if it was one conjunctive proposition. Hence, I will sometimes use logical notation of the form $KB \equiv (\alpha \,\&\, ... \,\&\, \beta)$ meaning that KB is logically equivalent to what is a conjunction the formulas α, β and perhaps others in place of the ellipsis.

On Bacchus et al.'s account, the values of the random-world credences are obtained by use of the principle of indifference and by an unconventional kind of conditionalisation (which I shall shortly explain). That this is so is explicitly affirmed in various places, including their statement that they "give semantics to degrees of belief by considering all worlds of [domain] size N to be equally likely, conditioning on KB, and then checking the probability of [a formula] φ over the resulting probability distribution."[46] However, Bacchus et al. provide no interpretation of the distributions (such as, for example, specifying that they are long-run frequencies or credences). They only define the random-worlds credence in terms of the non-interpreted posterior distribution or, as we shall see later, the limit(s) of the distribution as the domain size goes to infinity and the tolerance for any relevant approximate connectives goes to zero.

Let us examine the principle of indifference and conditionalisation in more depth. To simplify the presentation that follows, let us assume that the sample space of possible

[44] Bacchus et al., "From Statistical Knowledge Bases."

[45] Ibid., 77.

[46] Ibid., 99. See also the quote, "Combining our choice of possible worlds with the principle of indifference, we obtain our prior distribution. We can now induce a degree of belief in [a formula] φ given KB by conditioning on KB to obtain a posterior distribution and then computing the probability of φ according to this new distribution." Ibid., 78.

outcomes are certain sentences constructible from a language rather than models. This diverges from Bacchus et al.'s logic in which the sample space are models, but this divergence will be useful for understanding the basic ideas behind indifference and conditionalisation before relating these ideas back to the logic.

The principle of indifference is typically formulated as instructing one to regard each possible outcome as equally probable in the absence of evidence to favour one outcome over another.[47] For example, suppose the agent is before three doors where each door may or may not conceal a prize. Suppose we have a language, then, with only one predicate Ax where Ax stands for *x* conceals a prize and where each door is uniquely named with a constant a_i where a_i is the *i*-th door. Then the sample space of possible outcomes O_i is as follows:

O_1: Aa_1 & Aa_2 & Aa_3
O_2: Aa_1 & Aa_2 & $\sim Aa_3$
O_3: Aa_1 & $\sim Aa_2$ & Aa_3
O_4: Aa_1 & $\sim Aa_2$ & $\sim Aa_3$
O_5: $\sim Aa_1$ & Aa_2 & Aa_3
O_6: $\sim Aa_1$ & Aa_2 & $\sim Aa_3$
O_7: $\sim Aa_1$ & $\sim Aa_2$ & Aa_3
O_8: $\sim Aa_1$ & $\sim Aa_2$ & $\sim Aa_3$

Supposing the agent has no reason to favour one outcome over another, the principle of indifference states that these outcomes have equal and sharp probabilities. Bacchus et al. imagine such a uniform distribution to exist prior to receiving the information in KB and, hence, they call it a "prior distribution".[48]

They attempt to provide some justification for their use of the principle of indifference. They assume that KB represents everything that the agent knows. The agent then has a prior probability distribution which reflects a state of complete ignorance. Since the agent in this state has no *a priori* reason to prefer one outcome over another, this justifies the principle of indifference's prescription that all outcomes are assigned the same sharp probability.[49] They

[47] Various versions of the principle of indifference can be found in Christopher Meacham, "Impermissive Bayesianism," *Erkenntnis* 79, no. 6 (2014): 1191, Michael Strevens, *Tychomancy* (Cambridge, Mass.: Harvard University Press, 2013), 32-3 and Jonathan Weisberg, "Varieties of Bayesianism," in *Handbook of the History of Logic*, vol. 10, edited by Dov M. Gabbay and John Woods (Amsterdam; Boston: Elsevier, 2010), 505 and 509.
[48] Bacchus et al., "From Statistical Knowledge Bases," 78.
[49] Ibis.

restrict their use of the principle, however, by claiming that it is a useful method of determining credences only “in certain contexts, and in particular, in the context where we restrict our attention to a finite collection of worlds.”[50]

The posterior probabilities are obtained by conditionalising the distribution on the knowledge base. In the Bayesian landscape, standard conditionalisation can be understood as the process by which an agent, upon becoming certain of some new evidence that is represented by e, changes her (subjective) probability distribution to equal her prior probability distribution *conditional* on e (where e is the strongest proposition that the agent becomes certain of). In technical notation, then, conditionalisation is when $P'(h) = P(h|e)$ for any hypothesis h and any consistent statement e where $P(.)$ and $P'(.)$ are the agent’s prior and posterior credence functions respectively. $P(h|e)$ is frequently computed with Bayes’s theorem, an instance of which is such:

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)}$$
$$\text{where } P(e) > 0.$$

Bacchus et al. do not define what they mean by “conditioning”, but they do associate it with “the Bayesian approach”.[51] This suggests that “conditioning” is the prescription that the posterior distribution given KB is equal to the uniform prior probability distribution *conditional* on the set of propositions in KB (or a limit or set of limits for such distributions, as we shall see).

Let us illustrate how this works by building on the example concerning the three doors. Suppose the agent is told that there is a prize behind one of the three doors and so she accepts into her knowledge base the statement that $\%(Ax) = \frac{1}{3}$. She then wants to know what her credence should be in Aa_1, the proposition that the prize is behind door 1. According to conditionalisation, her credence should be equal to the uniform probability distribution conditional on KB. For illustrative purposes, we can suppose (unrealisatically) that the KB is, in this case, equivalent to $\%(Ax) = \frac{1}{3}$, so $KB \equiv \%(Ax) = \frac{1}{3}$. Although Bayesians typically compute the prior conditional probability $P(h|e)$ with Bayes’s theorem, it can also be

[50] Bacchus et al., “From Statistical Knowledge Bases,” 78.
[51] Ibid.

computed with Kolmogorov's ratio formula from which Bayes's theorem can be derived.[52] Use of this latter equation is simpler in this context. Therefore:

$$\mu(Aa_1|KB) = \frac{\mu(Aa_1 \& KB)}{\mu(KB)}$$

if $\mu(KB) > 0$ where $\mu(.)$ is the random-worlds prior probability measure.[53]

Recall that each of the outcomes in the below sample space have an equal measure.

O_1: Aa_1 & Aa_2 & Aa_3
O_2: Aa_1 & Aa_2 & $\sim Aa_3$
O_3: Aa_1 & $\sim Aa_2$ & Aa_3
O_4: Aa_1 & $\sim Aa_2$ & $\sim Aa_3$
O_5: $\sim Aa_1$ & Aa_2 & Aa_3
O_6: $\sim Aa_1$ & Aa_2 & $\sim Aa_3$
O_7: $\sim Aa_1$ & $\sim Aa_2$ & Aa_3
O_8: $\sim Aa_1$ & $\sim Aa_2$ & $\sim Aa_3$

Note that $\mu(KB) = \frac{3}{8}$ since $\%(Ax) = \frac{1}{3}$ is true in three of the outcomes (O_4, O_6 and O_7) where each outcome has a measure of $\frac{1}{8}$. Additionally, $\mu(Aa_1 \& KB) = \frac{1}{8}$ since the conjunction $Aa_1 \& \%(Ax) = \frac{1}{3}$ is true in only one outcome (O_4). Consequently:

$$\mu(Aa_1|KB) = \frac{\mu(Aa_1 \& KB)}{\mu(KB)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}$$

\

[52] Alan Hájek calls this ratio equation the *ratio analysis of conditional probability* and he critiques it in Hájek, "What Conditional Probability Could Not Be," 273-323.

[53] The function $\mu(.)$ is used instead of a credence function $P(.)$ since, as mentioned, Bacchus et al. do not explicitly interpret $\mu(.)$ as an agent's credence. Even if $\mu(.)$ were a credence function, it would presumably be no real agent's credence function since Bacchus et al. appeal to the dynamic process of conditionalisation here as a hypothetical means for calculating a static constraint on an agent's credences given that they have a particular knowledge base.

Note how $\mu(Aa_1|KB)$ can be calculated, then, by simply counting the number of outcomes for which Aa_1 & KB is true relative to the total number of outcomes for which KB is true. Hence:

$$\mu(Aa_1|KB) = \frac{|\boldsymbol{O}(Aa_1 \,\&\, KB)|}{|\boldsymbol{O}(KB)|}$$

if $|\boldsymbol{O}(KB)| > 0$ and where $\boldsymbol{O}(\alpha)$ is the set of all outcomes such that α is true for any α and $|\boldsymbol{O}(\alpha)|$ is the cardinality of that set.

After conditionalising on the KB, the resulting posterior distribution $\mu'(.)$ is equal to $\mu(.|KB)$. Obtaining the value of the posterior distribution is then simply a matter of counting outcomes.

In this sense then, the credences of the posterior distributions for random-world credences can be ascertained by counting outcomes. However, the random-worlds method is more complicated than the above (simplified) presentation because the possible outcomes are models rather sentences in a language.[54]

The calculation for random-worlds credences is also more complicated since one typically does not know the domain size N or the value for the tolerance in relevant approximate equalities such as those of the form $\%(\alpha|\beta) \approx f$. Bacchus et al. state that one often only knows that N is large and that the degree of approximation is small. Consequently, they define the random-worlds credence function through use of limits for N and for a vector of tolerances $\vec{\tau} = \{\tau_1, \dots, \tau_m\}$ such that $\tau_i > 0$ (a concept which we saw in Sub-Section 2.1.3.). (If, however, the domain size is known, they state that the credence can be calculated using the known domain size rather than the limit.)[55] Their definition is essentially that the credence for a formula α given KB is equal to the proportion of worlds in which the statement α & KB is true relative to the total number of worlds in which KB is true. Taking into account the limits, we can say formally:

[54] Given the language in the example and a domain of three objects, the number of possible outcomes qua sentences constructible from the language is eight whereas the number of possible outcomes qua logical models constructible from the language and domain is 216. We can see, then, how visually illustrating conditionalisation and indifference is easier when the sample space is the set of such sentences instead of the corresponding models.

[55] Bacchus et al., “From Statistical Knowledge Bases,” 99, fn. 12.

$$\lim_{\vec{\tau}\to\vec{0}} \lim_{N\to\infty} P_{Random-Worlds_N^{\vec{\tau}}}(\alpha|KB) = \frac{|\boldsymbol{W}_N^{\vec{\tau}}(\alpha \,\&\, KB)|}{|\boldsymbol{W}_N^{\vec{\tau}}(KB)|}$$

if $\left|\boldsymbol{W}_N^{\vec{\tau}}(KB)\right| > 0$ and where $\boldsymbol{W}_N^{\vec{\tau}}(\alpha)$ is the set of all worlds of domain size N such that α is true for any α and $\left|\boldsymbol{W}_N^{\vec{\tau}}(\alpha)\right|$ is the cardinality of that set.[56]

If the number of worlds in which KB holds is zero, then the probability is not well-defined. However, the above definition may not be well-defined even if the number of worlds in which KB holds is not zero. This may be because the limit oscillates between values. They account for such cases by appealing to infimums, supremums and eventual consistency, but these details need not bother us here. Regardless, sometimes the random-worlds credence is not well-defined for certain knowledge bases.[57]

To simplify the notation, let us denote $\lim_{\vec{\tau}\to\vec{0}} \lim_{N\to\infty} P_{Random-Worlds_N^{\vec{\tau}}}(\alpha|KB)$ with $P_{Random-Worlds}(\alpha|KB)$.

The random-worlds method may be confusing because it resembles two distinct and conflicting ways of understanding the relationship between credences and categorical belief. Categorical belief is a notion of belief that is non-gradational – you either have the belief or you do not. There are no strengths of belief. One way of thinking reduces categorical belief to credence.[58] Accordingly, credence is epistemologically primitive, thus entailing that credence is not defined in terms of categorical belief. Here, then, the agent is taken to believe a proposition if the credence reaches a certain threshold of confidence.[59] For example, one might say that one categorically believes a proposition *p* iff their credence in *p* is 1 (or some similarly high value in the unit interval). This is known as the *Lockean thesis*. This way of thinking is associated with a Bayesian approach in epistemology. Another way of thinking, however, reduces credences to categorical beliefs by defining credences via counting the models which are consistent with the set of categorical beliefs.[60]

The random-worlds method resembles ideas associated with both approaches. Like the latter way of thinking, it defines credences by counting the models that are consistent with a

---

Bacchus et al., "From Statistical Knowledge Bases," 99. Bacchus et al. instead use the notation $\Pr_n^{\vec{\tau}}(\varphi|KB) = \frac{\#worlds_n^{\vec{\tau}}(\varphi \wedge KB)}{\#worlds_n^{\vec{\tau}}(KB)}$.

[57] Ibid., 100.

[58] However, there is at least one other way of reducing quantitative to qualitative belief, one which regards quantitative beliefs as being just qualitative beliefs about quantitative statements such as those expressing propositions about probabilities.

[59] Franz Huber, "Formal Representations of Belief," *The Stanford Encyclopedia of Philosophy*, accessed January 31, 2016, http://plato.stanford.edu/entries/formal-belief/.

[60] I thank Jeremy Seligman for bringing this way of thinking and its relevance to my attention.

set of sentences in KB, a set that is essentially a categorical belief set. However, it also resembles ideas associated with the former way of thinking. In particular, the random-worlds method resembles *objective Bayesianism*, a position which typically recommends a prior probability distribution that is constrained by the principle of indifference (or a related principle known as Maximum Entropy) and updated in accordance with Bayesian conditionalisation.[61]

Nevertheless, the random-worlds method does not resemble typical Bayesian thought in a few respects. For one, in cases where the domain size is presumably large but unknown, the random-worlds credence is defined as the *limit* (or limits) of a set of posterior probability distributions, rather than just one distribution. Furthermore, the prior and posterior distribution are not interpreted as credences. The conditionalisation envisaged by the random-worlds method is best seen as *hypothetical*. The random-worlds method does not actually envisage that a real agent has a uniform *prior* probability distribution over all the models constructible from a language and that this is then updated by conditionalising on a KB to obtain a *posterior* distribution. The method does not specifically tell the agent how to update their credences from one time to another later time. In this sense, their definition of credences is compatible with a non-Bayesian belief-revision theory of updating. Instead, the random-worlds method articulates a *static* or *synchronic* constraint on credences; it specifies only that if the agent has a set of beliefs or propositions as her KB at *a given time*, then she should have a specific credence *at that same time*. Additionally, contra the credences of the random-worlds method, many Bayesians (particularly subjective Bayesians) hold that accepted statements in a KB and the principle of indifference do not suffice determine all of one's credences. Some credences may just be, for example, hunches or determined via subjective judgments of simplicity.

That concludes our summary of the essential elements of the random-worlds method.

[61] A vociferous proponent of objective Bayesianism was E.T. Jaynes who promoted both conditionalisation and the principle of maximum entropy (which entails the principle of indifference). See Jaynes, *Probability Theory.*

4. Evaluating the Random-Worlds Method

So we have examined the basics of the random-worlds method. Ultimately, though, we are interested in the central question of how evidence about relative frequencies should constrain credences. My thesis's contention is that the random-worlds method provides at least part of a good answer to this question, although it can also be supplemented with the proposals presented in chapters 7 and 8 to extend the method's applicability. To argue this contention, it will be necessary to outline some important features of the method and to also evaluate its merit. This, then, is the focus of this chapter.

4.1. Evaluative Criteria

To assess whether the random-worlds method is a good answer to the question of how to constrain credences, we need to explore the criteria for what a good answer is.

So what, then, are touchstones for evaluating the worth of such an answer, or indeed any normative epistemological theory?

Three putative touchstones are salient in the literature, each of which correspond to a particular position.

To illustrate these positions, I will sometimes discuss them in the context of whether to accept a particular belief rather than in the context of whether to have a particular credence or accept an epistemological theory that somehow constrains beliefs or credences. Regardless, these basic points can be generalised to these latter contexts.

One position is what we may call *pragmatism*. This is the position that a normative epistemological theory is acceptable to the extent that it is practically rational to accept the theory. Practical rationality may be cashed out differently, but we can say here that an agent is practically rational to the extent that she acts in a way that is in her interests. Dutch book arguments are a hallmark example of pragmatic arguments.[62] A Dutch book is a combination of wagers which will guarantee financial loss for the agent if accepted by her. Dutch book arguments assert that, if one does not follow some particular norm, then they are susceptible

[62] Susan Vineberg, "Dutch Book Arguments," *The Stanford Encyclopedia of Philosophy*, accessed May 6, 2016, http://plato.stanford.edu/archives/spr2016/entries/dutch-book/ and Alan Hájek, "Dutch Book Arguments," in *The Oxford Handbook of Rational and Social Choice,* eds. Paul Anand, Prasanta Pattanaik and Clemens Puppe (Oxford: Oxford University Press, 2009), 173-195.

to accepting a Dutch book and so they should therefore accept the norm. Such arguments have been offered in support of various Bayesian norms, including the norms that credences should conform to axioms of probability and some principle of conditionalisation.

However, pragmatism is problematic. By the very nature of belief as purporting to be about what is true, we should believe something only to the extent that it appears to us to reflect the truth. Yet there is often no connection between what is true and what is practically rational to believe (to some degree). Consider people who have faced the prospect of torture and execution for their beliefs and are offered life on the condition that they recant them.[63] We can easily imagine a situation like this in which the agent has the greatest pragmatic motivation for believing or disbelieving a given proposition, yet is unable to do as such and rightly so. After all, it seems absurd to say, for example, "I now think that the universe is geocentric because, if I do, I then minimise my expected loss of life." At most, it seems that the agent can reasonably only *act as if* the proposition is false. The agent, for example, may only declare her belief to be false *as if* she disbelieved it on pragmatic grounds despite the grounds failing to warrant *genuine* disbelief. Hence, a normative epistemological theory is best evaluated with respect to non-pragmatic criteria.

Two non-pragmatic positions are salient in this respect.

One is what we might call *structuralism*, the position asserting that a normative epistemological theory is acceptable to the extent that it corresponds to some objectively existing features of the world (features that matter here not because of pragmatic concerns). In essence, then, the position holds that what justifies accepting the theory is our beliefs about the way that the world is structured. An example of structuralism concerns the justification of induction.[64] Consider the classic inductive inference that the sun will continue to rise every day since it has been observed to regularly rise in the past. An induction of this sort might be justified by the belief that nature is structured in a way that is uniform or regular in some respects, including this one; hence, we can appeal to the past to predict the future since nature is and will be uniform in this respect even as it has been in the past.

Yet a problem confronting structuralist criteria for evaluating a normative epistemological theory is that the structures often, if not always, cannot be justifiably believed in (to some degree) without already controversially presupposing the normative theory. For example,

[63] See, for example, the accounts of the Bahá'ís who faced these scenarios in Moojan Momen, ed., *The Bábí and Bahá'í Religions 1844-1944: Some Contemporary Western Accounts* (Oxford: George Ronald, 1981).
[64] See a useful introduction to the controversy in John Vickers, "The Problem of Induction," *The Stanford Encyclopedia of Philosophy*, accessed May 6, 2016, http://plato.stanford.edu/archives/spr2016/entries/induction-problem/#HumInd.

David Hume's problem of induction is a problem precisely because induction cannot be justified by appealing to uniformity in nature without already presupposing the reliability of induction to infer this uniformity (or its continuation).

A third position is what we might call *epistemic intuitivism*. A normative epistemological theory is acceptable to the extent that it accords with our foundational intuitions about what is epistemically rational to believe (to a degree).[65] Loosely speaking, one is epistemically rational to the extent that their beliefs or credences are supported by the evidence.[66] Foundational intuitions are beliefs that are (at least initially) not justified by other beliefs.[67] To further distinguish the view, these intuitions must not be intuitions about the structure of the world or pragmatic considerations. To illustrate a foundational intuition, suppose I truthfully tell you that there is a marble hidden in my hand and that it is either green, red or blue. What should be your credence that it is blue given the evidence of my testimony? A common and intuitive answer is to say that the credence should be $1/3$. For some, the intuitive appeal of this answer is a reason to endorse the principle of indifference insofar as the principle accords with this intuition. However, this intuition is a foundational intuition about what is epistemically rational to believe; there is no appeal to pragmatic benefits in having such a credence nor some way in which the world is structured.[68] One might doubt the other positions because some epistemic norms are putatively rational (an example possibly being the principle of indifference), yet fail to be pragmatically rational or correspond to some justifying-structure in the world.

Much of the literature on direct inference conforms to this third position. As Bacchus et al. note, scholars on direct inference and default reasoning often assess a theory based on whether it accounts for intuitions about what the right inferences are in various cases.[69] These

[65] Nelson Goodman articulates a similar position in Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge, MA: Harvard University Press, 1955). He would also introduce a distinction between intuitions about *instances* of rational inferences and intuitions about *rules* of rational inference, stating that we aim for a sort of reflective equilibrium among intuitions of the two kinds.

[66] Some, such as Susanna Rinard, may have qualms with the concept of two distinct types of rationality – one concerning epistemically rational belief and the other concerning practically rational belief. Nevertheless, presumably these people would accept some distinction between a belief being supported by evidence bearing on its (probable) truth and a belief being accepted on pragmatic grounds which conduce to, say, utility. Epistemic rationality, then, can just be some concept that fills in for the reader's preferred concept which applies to cases of the former sort.

[67] These are just instances of what are known as *basic beliefs* in traditional epistemology. See Noah Lemos, *An Introduction to the Theory of Knowledge* (Cambridge, UK: Cambridge University Press, 2007), 45.

[68] Some might attempt to justify the intuition in terms pragmatic or structuralist considerations, yet I suspect many (like myself) have this intuition without such a justification.

[69] Bacchus et al., "From Statistical Knowledge Bases," 79. As an example, see Kyburg and Teng's statement that the "true test" of his semantics are "whether it yields intuitively plausible results in simple cases, to begin

intuitions are divorced from structuralist or pragmatist considerations (at least in the way that they are presented).

Some putative advantages of intuitivism is that we can know what the intuitions are that speak for or against a theory without already presupposing the theory – contra structuralist criteria – and the intuitions, analytically speaking, intuitively bear on what is epistemically rational to believe – contra pragmatist criteria. A putative disadvantage of epistemic intuitivism is that it confronts the problems facing a position known as *foundationalism* insofar as it does not demand justification for these basic intuitions in terms of other beliefs.[70]

Of course, these positions can be interrelated and a mixture of them may be possible. One might say that a normative epistemological theory is acceptable to the extent that it performs well on the three criteria associated with these positions. One might also think that some criteria relate to others. For example, we could tell a story which ties all three criteria together: the cognitive faculties of humans, and the resulting intuitions about rational epistemic practices, are products of the structures of the world and, furthermore, those faculties and intuitions conduce to pragmatically rational action and belief by virtue of their conformity to the way that the world is structured.

Given that I find epistemic intuitivism to be a reasonable position and that it is embraced by others working on direct inference (at least in appearance), I will carry on the tradition of using it to evaluate normative theories, even though, admittedly, it is under argued for here and is necessarily so for considerations of space.

However, a question arises: if a normative theory such as the random-worlds method is ultimately evaluated by whether it accounts for intuitions about rational epistemic practices, then why not simply let these intuitions guide our epistemic practices without the theory? In other words, why not cut out the middle-man and simply let intuitions be the guides of inductive inference?

My response is that although these intuitions may be in themselves reliable guides, a theory may provide guidance for inductive inferences in problematic cases where our intuitions are initially less decisive or clear. Hence, if a theory validates a number of clear intuitions about rational practices, then this suggests that it likewise provides trustworthy guidance in these problematic cases. Perhaps the prime example of this concerns the problem of what to do when one has evidence regarding relevant but incomparable reference classes

with, and reasonable results in more complex cases in which our intuitions are not so strong." Kyburg and Teng, *Uncertain Inference*, 244.

[70] See a useful introduction to foundationalism in Lemos, *An Introduction to the Theory of Knowledge,* ch. 3.

(which we shall encounter in Sub-Section 4.2.7.) and it is not clear what inductive inferences are rational in such contexts. The random-worlds method can assign *specific and sharp values* to the probabilities of interest where intuitions may not.

For this reason, it is worth seeing whether the intuitions weighing in on the random-worlds method are such that they justify using it as a guide for our inductive inferences, particularly in cases where intuitions are (initially) less helpful.

So how well does the random-worlds method accommodate foundational intuitions about epistemically rational practices?

To answer this question, we need to evaluate the implications and features of the random-worlds method and see whether they deliver intuitive results.

4.2. Favourable Features of the Random-Worlds Method

Let us now consider the various features of the method.

In doing so, I will occasionally present certain results of the method that are entailed by slightly more general theorems proved by Bacchus et al. However, it is important to note that Bacchus et al. say that the results and the original theorems are "not the most general ones possible" and that the conditions stipulated by the theorems can be relaxed (in some unspecified way) while producing intuitively reasonable results.[71] Hence, one should not conclude that the method is often inapplicable merely because the results I present in what follows concern unrealistically simple situations.

4.2.1. Basic Direct Inference

The first feature that we will consider is the method's validation of *basic direct inference.*

To illustrate basic direct inference, suppose the agent sees a news report in which a policeman states that 80 of 100 passengers on the *Midnight Express* train had been fatally injured. Furthermore, the agent has evidence that her friend Aaron caught that train. Therefore, suppose the agent accepts into her knowledge base the statements $\%(Ax|Bx) = 0.8$ and Ba where Bx stands for "x was a passenger on the *Midnight Express,*" Ax stands for

[71] Bacchus et al., "From Statistical Knowledge Bases," 102.

"x was fatally injured" and a stands for Aaron. Suppose the agent has no other information bearing on the probability of Aa, including information about any of the other passengers.

So what should $P(Aa)$ be given the agent's acceptance of $\%(Ax|Bx) = 0.8$ and Ba?

We probably have at least two intuitions about this case. First, if $(\%(Ax|Bx) = 0.8) \in KB$, then $P(Aa) = 0.8$. Second, $P(Aa)$ would increase as the known value of $\%(Ax|Bx)$ increases and it would decrease as $\%(Ax|Bx)$ decreases. So, for instance, if $\%(Ax|Bx) = 0$, then $P(Aa) = 0$ and if $\%(Ax|Bx) = 1$, then $P(Aa) = 1$.

The random-worlds method validates these intuitions by prescribing credences that accord with them. Bacchus et al. prove a theorem to this effect. Suppose $KB \equiv (\beta(c) \,\&\, \%(\alpha(x)|\beta(x)) \in [f, g] \,\&\, KB')$ where KB' represents other statements in the knowledge base, $\alpha(x)$ and $\beta(x)$ are formulas mentioning the variable *x* and $\beta(c)$ is the formula $\beta(x)$ where the constant c takes the place of the variable *x*. Bacchus et al.'s theorem entails that if the KB is as such, then $P_{Random-Worlds}(\alpha(c)|KB) \in [f, g]$ (given that the credence exists).[72] This is only on certain restricting assumptions which effectively ensure that $\beta(c)$ encapsulates all the relevant information about *c*.[73] For one, KB' cannot have any statements about *c*. Also, c cannot appear in α or β. To illustrate why this is so, suppose that we define the symbols so that $\beta(x) \equiv ((Cx \,\&\, x \neq c) \vee \sim Cx)$, $\alpha(x) \equiv Cx$ and $KB \equiv (\beta(c) \,\&\, \%(\alpha(x)|\beta(x)) = 0.5)$.[74] In this case, a naive account of direct inference could prescribe that $P(\alpha(c)|KB) = 0.5$, but $\beta(c)$ holds only if $\sim\alpha(c)$ holds, and so it should really be the case that $P(\alpha(c)|KB) = 0$.

We can then see how the random-worlds method validates the two intuitions in the example case. If we instantiate the knowledge base so that $KB \equiv (Ba \,\&\, \%(Ax|Bx) = 0.8 \,\&\, KB')$, then $P_{Random-Worlds}(Aa|KB) = 0.8$. (Note that $\%(Ax|Bx) = 0.8$ is equivalent to $\%(Ax|Bx) \in [0.8, 0.8]$ and likewise for $P_{Random-Worlds}(Aa|KB) = 0.8$.) Furthermore, the value of the random-worlds credence varies with the known proportion of injured passengers, so $P_{Random-Worlds}(Aa|KB) = \%(Ax|Bx)$. This then validates the two intuitions for the illustrative case.[75]

[72] A statement of the theorem is found in Bacchus et al., "From Statistical Knowledge Bases," 105-6. Proof can be found in Bacchus et al., "From Statistical Knowledge Bases," 133-4.

[73] Bacchus et al., "From Statistical Knowledge Bases," 106.

[74] Ibid.

[75] Although many would share these two intuitions, I should avoid giving the impression that they are universally held since there is at least one dissenter, philosopher Isaac Levi. At least at one point in his career, he thought one needed to know that the means for selecting the object from the reference class is known to be unbiased. We will consider this issue in more detail in Chapter 6.

Let us see how their system guarantees this result.[76] This will be useful as it will also give an example of the kind of proof method used by Bacchus et al. for establishing other theorems. Suppose that these restricting assumptions hold and that $KB \equiv (\beta(c)\ \&\ \%(\alpha(x)|\beta(x)) \in [f,g]\ \&\ KB')$. Suppose KB is satisfiable for a domain of fixed size *N*. Let us now partition the models that satisfy KB into clusters whereby two models are in the same cluster iff they agree on the denotation of all symbols in the vocabulary except *c*. For example, for any predicate F and any domain object d such that *c* does not denote *d*, two worlds are in the same cluster iff both worlds agree on whether F does or does not denote *d*. Now all the worlds satisfy KB and hence the formulas $\beta(c)$ and $\%(\alpha(x)|\beta(x)) \in [f,g]$. The worlds within a cluster differ only in respect of which object *c* is assigned to. In a cluster, c denotes one of the objects satisfying $\beta(x)$ and the fraction of such objects that satisfy $\alpha(x)$ is given by $\%(\alpha(x)|\beta(x)) \in [f,g]$. Hence, the fraction of the equiprobable worlds in a cluster in which $\alpha(c)$ is true is given by $\%(\alpha(x)|\beta(x)) \in [f,g]$. Therefore, the weighted average of the probabilities within all of the clusters gives us $\mu(\alpha(c)\ |KB) \in [f,g]$ where μ is the random-worlds prior probability function. This holds irrespective of the choice for *N*. Thus, if $KB \equiv (\beta(c)\ \&\ \%(\alpha(x)|\beta(x)) \in [f,g]\ \&\ KB')$ and the restricting assumptions hold, then $P_{Random-Worlds}(\alpha(c)|KB) \in [f,g]$ (provided that the random-worlds credence exists).

Let us now note some of the other features of the random-worlds method.

4.2.2. Independence and Rational Monotonicity

Probabilistic independence is an important concept in inductive reasoning. As Pollock notes, probability practitioners typically assume two propositions to be probabilistically independent unless there is reason to think otherwise.[77] For example, the probability of me having arthritis should be regarded as independent of whether or not Barack Obama has a cute kitten. While this might provide an intuitive idea of independence, two formulas α and β are formally said to be independent iff it is the case that $P(\alpha\ \&\ \beta) = P(\alpha)P(\beta)$ (although

[76] This proof is provided in Bacchus et al., “From Statistical Knowledge Bases,” 133-4.
[77] John Pollock, “Probable Probabilities (with Proofs).”

authors occasionally express independence using a slightly different equation whose terms are conditional probabilities).[78]

The intuition that, unless given reasons to think otherwise, two predicates should be regarded as independent is validated by Bacchus et al.'s method.

Bacchus et al. prove a theorem which specifies a class of cases in which two predicates are treated as independent.[79] Suppose that there are two knowledge bases KB_1 and KB_2 and two formulas α_1 and α_2 and that these are in pairwise separate languages. In other words, the language of KB_1 and α_1 and the language of KB_2 and α_2 are entirely different: there is no basic symbol (predicate or constant) used in both. Bacchus et al. prove a theorem entailing that the two are independent so that $P_{Random-Worlds}(\alpha_1 \,\&\, \alpha_2|KB_1 \& KB_2) = P_{Random-Worlds}(\alpha_1|KB_1)P_{Random-Worlds}(\alpha_2|KB_2)$. They also show that this result holds if the languages share a single constant.[80]

The random-worlds method also validates a version of what is called *rational monotonicity*. Bacchus et al. define rational monotonicity in the context of default reasoning, a type of defeasible inference from premises to conclusions, i.e. an inference that purports to be rationally compelling but not deductively valid.[81] Informally, a system of default reasoning has the property of rational monotonicity when it allows one to conclude α from the premise $(KB \,\&\, \beta)$ if one can conclude α from the premise KB and cannot conclude $\sim\beta$ from the premise KB.[82] Bacchus et al. also provide examples of rational monotonicity whereby adding a predicate to the reference formula in a probability does not impact the relevant probability. Let us consider one of these examples. Suppose the agent knows that (approximately) all penguins cannot fly and that Tweety is a yellow penguin. Formally, let us say that the agent's knowledge base includes that statements $Pt \,\&\, Yt$ and $\%(Fx|Px) \approx 0$ where Fx stands for "x can fly", Px stands for "x is a penguin", Yx stands for "x is a yellow" and t stands for Tweety. The random-worlds method entails that $P_{Random-Worlds}(Ft|Pt \,\&\, Yt \,\&\, \%(Fx|Px) \approx 0) = 0$.[83]

[78] Ibid.
[79] Bacchus et al., "From Statistical Knowledge Bases," 118-119 and 140-141.
[80] Ibid., 118.
[81] For more on defeasible reasoning and default logic, see Robert Koons, "Defeasible Reasoning," *The Stanford Encyclopedia of Philosophy*, accessed June 21, 2016, http://plato.stanford.edu/archives/spr2014/entries/reasoning-defeasible/.
[82] Bacchus et al., "From Statistical Knowledge Bases," 87.
[83] Ibid., 112.

This is a fairly intuitive result.[84] Indeed, in real life, any object from a reference class will presumably have some property (or combination of properties), such as being yellow and bearing the name "Tweety", that the reference-class as a whole must lack. Otherwise, there would be no simple or complex property to distinguish the object from other objects in the reference class. Yet if direct inference from a reference class statistic is to be useful, such differences must be treated as irrelevant to the probability of interest unless given reason to think otherwise.

So independence and rational monotonicity are intuitively reasonable features of the random-worlds method.[85]

4.2.3. The Reference Class Problem

So we have explored how the random-worlds method validates basic direct inference and the assumption of the probabilistic independence of predicates. Yet there are many cases in which the agent possesses evidence about relative frequencies that are putatively relevant to what her credences should be, but which are not usefully guided by just the aforementioned features of the method.

This is because in many cases, the object of interest is known to (at least probably) belong to two or more reference classes which are known to (probably) have different statistics for the target formula of interest. For example, suppose Sally smokes but exercises religiously. In this sense, she is a member of two reference classes, the class of all smokers on the one hand and the class of all fitness fanatics on the other. Suppose the relative frequency of eventually developing cancer among smokers is 0.3 whereas the relative frequency of eventually developing cancer among the fanatics is 0.05. What, then, is the probability that Sally will eventually develop cancer? Basic direct inference is unable to help us here since there are

[84] I personally have qualms with inferring a subjective probability of zero here from an approximate statistical statement which is consistent with the possibility that there is at least one penguin that flies which could be Tweety. Regardless, this is a fairly minor qualm with Bacchus et al.'s derivation of strict equalities for probability statements from approximate equalities for statistical statements, not for their validation of rational monotonicity as such.

[85] Rational monotonicity is similar to what Pollock calls *nonclassical direct inference*, the gist of which is that the addition of a predicate to a reference formula in a probability does not affect the probability unless given reason to think otherwise. According to Pollock, probabilists often reason in accordance with nonclassical direct inference in practice. Pollock notes that, according to the probability calculus, if one endorses the defeasible assumption of statistical independence, then they must endorse the principle of nonclassical direct inference (albeit with a minor qualification), and vice versa. Yet Pollock claims that people often only have intuitions in favour of the former, but not the latter. Pollock, "Probable Probabilities (with Proofs)," 11-13.

relevant statistics for multiple reference classes. So we have the problem of trying to determine from which class, if any, does Sally *inherit*, to use a term in the literature, her probability of eventually developing cancer.

This is widely-known as *the reference class problem.* However, specialists on the topic actually distinguish various kinds of reference class problems, a number of which concern different types of possible tension between statistics for distinct reference classes.[86]

Here, we will examine the random-worlds treatment of four types of tension, namely, the tension between statistics for:

a) A more specific and a less specific reference class
b) A more precise reference class and a less precise reference class.
c) A reference class that is too specific and a reference class that is too general, and
d) Two incomparable reference classes.

4.2.4. Specificity

The first kind of tension that we will consider is between statistics for a more specific reference class and a less specific reference class. A reference class $\boldsymbol{B'}$ is more specific than another class $\boldsymbol{B}$ iff $\boldsymbol{B'} \subseteq \boldsymbol{B}$ where $\boldsymbol{B}$ is the set of all objects satisfying Bx and $\boldsymbol{B'}$ is the set of all objects satisfying $B'x$. (From here on, bolded predicate letters or formulas will refer to the sets of objects satisfying the predicate or formula.)

The preference for using statistics for more specific reference classes is quite intuitive (although there are plausibly certain exceptions to it to be discussed in the next sub-section). For instance, let Bx stand for "x was a passenger on the *Midnight Express"* and let $B'x$ stand for "x was a passenger on carriage 3 of the *Midnight Express*". Note that $\boldsymbol{B'} \subseteq \boldsymbol{B}$ since the set of all of the passengers who were on carriage 3 is a subset of the set of all of the passengers who were on the train. Now suppose the agent knows that Ba and $B'a$ where a stands for Aaron. Suppose that the agent also knows that $\%(Ax|Bx) = 0.8$, $\boldsymbol{B'} \subseteq \boldsymbol{B}$ and $\%(Ax|B'x) = 0.05$ where Ax stands for "x was fatally injured". From which statistics should the object denoted by a inherit its probability of satisfying Ax? Intuitively, it seems

[86] See, for example, Paul Thorn, "Three Problems of Direct Inference" (Doctoral thesis, The University of Arizona, 2007) and Alan Hájek, "The Reference Class Problem is your Problem too," Synthese 156, no. 3 (2007): 563-585.

that the agent should set their credence so that $P(Aa) = \%(Ax|B'x) = 0.05 \neq \%(Ax|Bx) = 0.8$. This widely held intuition finds expression in what is known as the *principle of the narrowest reference class*.[87]

The random-worlds method validates this preference for using statistics for the more specific reference classes.

Bacchus et al. prove a theorem showing that the random-worlds approach satisfies several desiderata, including this preference. The theorem employs the concept of a minimal reference class $\boldsymbol{\beta'}$ with respect to a target class $\boldsymbol{\alpha}$.[88] Suppose that the KB contains a number of statistical statements for different reference classes. The minimal reference class $\boldsymbol{\beta'}$ is the one such that all other reference classes in KB are either larger or disjoint from it. This means that these other classes are either super classes of the minimal one (like the class of birds is for the class of sparrows) or are completely non-overlapping (like the classes of sparrows and fish which do not overlap). This can be represented with logical notation by saying that all other reference classes $\boldsymbol{\beta}$ are such that either $\forall x(\beta'(x) \supset \beta(x))$ or $\forall x(\beta'(x) \supset \sim\beta(x))$.

Here is a less general statement of the theorem for cases in which the statistical statement for the minimal reference class involves a strict equality. Let KB be a knowledge base such that:

a) $KB \vDash \beta'(c)$
b) For any statement of the form $\%(\alpha(x)|\beta(x))$ for any reference class $\boldsymbol{\beta}$, either $KB \vDash \forall x(\beta'(x) \supset \beta(x))$ or $KB \vDash \forall x(\beta'(x) \supset \sim\beta(x))$
c) The symbols in the target formula α in KB are only on the left hand side of the statistical statements mentioned in condition b
d) c does not appear in the formula $\alpha(x)$
e) $KB \vDash \%(\alpha(x)|\beta'(x)) = f$

Then $P_{Random-Worlds}(\alpha(c)|KB) = f$ (assuming that the limit exists).

Proof of the more general theorem is found in Bacchus et al.[89]

While also enabling the random-worlds method to validate an intuitive preference for more specific reference classes, note also that the theorem has the putatively desirable

[87] Williamson, *In Defense of Objective Bayesianism*, 13 and Reichenbach, *Theory of Probability*, 449.
[88] Bacchus et al., "From Statistical Knowledge Bases," 110.
[89] Ibid.,134-6.

consequence of validating the practice of ignoring statistics for irrelevant reference classes. This entails, for example, that the statistics about the prevalence of warm-bloodedness among, say, pencils are treated as irrelevant for determining the probability that a sparrow is warm-blooded.

4.2.5. Precision

In discussing the preference given to statistics for more specific reference classes, I alluded to an important exception. This exception concerns cases of a particular tension between statistics for two (or more) classes. In such cases, the tension arises because each statistic specifies that the relative frequency of objects satisfying some formula α among the relevant reference class lies in some interval where one of the intervals is more precise than the other. Technically speaking, an interval $[f, g]$ in this context is more precise (or, as some say, *accurate*) than another $[f', g']$ when it is the case that $[f, g] \subseteq [f', g']$ or, in other words, $[f', g']$ includes $[f, g]$. So this tension would arise if the knowledge base includes, say, statements of the form $\%(\alpha(x)|\beta'(x)) \in [f', g']$, $\%(\alpha(x)|\beta(x)) \in [f, g]$, $\beta(c)$ and $\beta'(c)$ where $[f', g']$ includes $[f, g]$. In such cases, a question arises: what should be the value of $P(\alpha(c))$? There is some consensus that $P(\alpha(c))$ should lie in the more precise interval $[f, g]$ (although various authors articulate more or less the same point with different interpretations of probability).[90]

This preference for more precise statistics is intuitive, especially as it constitutes a solution to what Paul Thorn has called the *problem of uninformative statistics*.[91] The problem is that for many non-trivial statistics of the form $\%(\alpha(x)|\beta(x)) \in [f, g]$ with reals $f, g \in (0,1)$ and some predicate β such that the agent knows $\beta(c)$, there is some more specific class $\boldsymbol{\beta'}$ such that the only knowledge that the agent has about $\boldsymbol{\beta'}$ is that $\beta'(c)$ and that $\%(\alpha(x)|\beta'(x)) \in [0,1]$. For example, the agent in the *Midnight Express* case knows that $\%(Ax|Bx) \in [0.8, 0.8]$ and Ba, but they also know that $\%(Ax|B''x) \in [0,1]$ where $\boldsymbol{B''}$ is the set of passengers on the train who were born on Aaron's birthday. Suppose further that the agent's knowledge does not entail that $\%(Ax|B''x)$ is in a particular interval more precise than $[0, 1]$. The agent's knowledge that $\%(Ax|B''x) \in [0,1]$ follows merely from the tautology that a proportion expression must be in the unit interval. The problem, then, is that

[90] Bacchus et al., "From Statistical Knowledge Bases," 136-138 and Kyburg and Teng, *Uncertain Inference,* 213.
[91] Kyburg and Teng, *Uncertain Inference,* 213. Among other places, Thorn uses this term for the problem in Paul Thorn, "Two Problems of Direct Inference," *Erkenntnis* 76, no. 3 (2012): 299-318.

if statistics for the more specific class are to be preferred for assigning probabilities *in general*, then one must embrace the uninformative conclusion merely that $P(Aa) \in [0,1]$. This problem can be generalised to many contexts. Kyburg and Teng regard the preference for more precise reference classes as a way of avoiding this problem. This is because the statistic for a more specific reference class is used for determining a probability *unless* the more specific statistic is for an interval which is less precise than a statistic for a less specific super-class. Consequently, this preference for statistical precision constitutes an important exception to the preference for reference class specificity.

Bacchus et al. present and prove a theorem which purports to validate this preference for more precise statistics.[92] Suppose $KB \equiv (\&_{i=1}^{n}(\%(\alpha(x)|\beta_i(x)) \in [f_i, g_i] \,\&\, \beta_i(c)) \,\&\, KB')$ where c is some constant of interest, $\alpha(x)$ and $\beta_i(x)$ are some formulas mentioning x and KB' denotes the rest of the knowledge base. Suppose also that $KB \vDash (\forall x(\beta_i(x) \supset \beta_{i+1}(x)) \,\&\, {\sim}\%(\beta_i(x)) \approx_1 0)$ for all i and that no symbol in $\alpha(x)$ appears in KB' or any formula $\beta_i(x)$. Suppose also that there is some j, such that $[f_j, g_j]$ is the most precise interval. Bacchus et al.'s theorem entails that if the knowledge base is as such, then $P_{Random-Worlds}(\alpha(x)|KB) \in [f_j, g_j]$ (if the limit exists).

4.2.6. Overly Specific and Overly General Reference Classes

Another kind of tension that is less discussed in the philosophical literature concerns the tension between statistics for reference classes which are too specific or too general in some sense. To illustrate this tension using Bacchus et al.'s example, consider a knowledge base containing the propositions that 90% of birds chirp, the set of all magpies is a subset of the set of all birds, 20% of moody magpies chirp and Tweety is a magpie.[93] Thence, suppose $KB \equiv (\%(Cx|Bx) = 0.9 \& \boldsymbol{M} \subseteq \boldsymbol{B} \,\&\, \%(Cx|Mx \,\&\, Nx) = 0.2 \,\&\, Mt \,\&\, KB')$ where Cx stands for "x chirps", Bx stands for "x is a bird", Mx stands for "x is a magpie", Nx stands for "x is moody", t stands for Tweety and KB' contains no statements that are relevant to the random-worlds credence for Ct. Bacchus et al. claim that, in this case, theories of direct inference usually disregard the information about $\%(Cx|Mx \,\&\, Nx)$ since it is not known

[92] Bacchus et al., "From Statistical Knowledge Bases," 136-138.
[93] Ibid., 115.

that Tweety is a moody magpie (i.e. that $Mt \,\&\, Nt$).[94] Scott Goodwin suggests that one should not ignore this information because of various considerations, including the possibility that Tweety could be moody.[95] Furthermore, Bacchus et al. claim that "ignoring the second statistic [$\%(Cx|Mx \,\&\, Nx) = 0.2$] in effect amounts to assuming that magpies generally are not moody."[96] The random-worlds approach validates Goodwin's suggestion by prescribing a credence of less than 0.9 to Ct given the information that $\%(Cx|Mx \,\&\, Nx) = 0.2$ (although Bacchus et al. do not provide a general theorem stating something to this effect).

4.2.7. Incomparable Reference Classes

In Bacchus et al.'s opinion, the most important case of tension concerns the tension between statistics involving *incomparable reference classes*.[97] In standard mathematical terminology, two sets are incomparable if neither is a subset of the other.

Let us illustrate this tension using a modified example from Pollock. Imagine that a patient named Bernard tests positive on two unrelated tests of a particular disease. The relative frequencies of the disease among patients who test positive are 0.7 and 0.75 for the tests. We can represent these frequencies with the proportion formulas $\%(Dx|\, T_1x) = 0.7$ and $\%(Dx|T_2x) = 0.75$ where Dx stands for "x has the disease" and T_ix stands for "x tested positive on test *i*". Furthermore, suppose the agent knows that Bernard tested positive on both tests and that $\%(Dx|T_1x) = 0.7$ and $\%(Dx|T_2x) = 0.75$. Suppose, though, that the agent is completely ignorant about the value of $\%(Dx|T_1x \,\&\, T_2x)$. In this case, the class corresponding to T_1x and the class corresponding to T_2x are the incomparable reference classes.

A problem arises because it may not be clear how the agent should assign probabilities in such cases. The agent cannot simultaneously assign incompatible values such as 0.7 and 0.75 to the probability that Bernard has the disease given that he is a member of the incomparable

[94] Ibid., 115.
[95] Scott D. Goodwin, "Second Order Direct Inference: a Reference Class Selection Policy," *International Journal of Expert Systems Research and Applications* 5 (1992): 185-210. I was unable to access this paper, so I am relying on Bacchus et al.'s brief mention of Goodwin's suggestion in Bacchus et al., "From Statistical Knowledge Bases," 115.
[96] Bacchus et al., "From Statistical Knowledge Bases," 115.
[97] Ibid., 116.

classes. Thorn calls this the *problem of competing statistics* while Pollock calls this the problem of *sparse probability knowledge*.[98] How, then, should the probability be determined given the agent's knowledge base in this case?

Unfortunately, there is much less consensus on how to appropriately use the relevant statistics here than in other cases of tension between reference classes. Reichenbach seems to prescribe ignoring the statistics for all of the incomparable classes and focusing on acquiring statistics for the most specific common class: "The logician can only indicate a method by which our knowledge may be improved. This is achieved by the rule: look for a larger number of cases in the narrowest common class at your disposal."[99] Kyburg and Teng, in contrast, prescribe associating the probability with the interval of values that spans from the lowest relevant frequency value to the highest.[100] So, in this case, they would simply state that the evidential probability of Db given the agent's knowledge base is the interval $[0.7, 0.75]$ where b stands for Bernard.[101]

Bacchus et al. and Pollock have approaches that are similar to each other, but markedly different from the approaches of Reichenbach, Kyburg and Teng.

Bacchus et al.'s method validates a well-known means for combining independent pieces of evidence which is called *Dempster's rule of combination*.[102] Given the assumption that c satisfies formulas of the form $\beta_i(x)$ for a set of reference classes $\{\beta_1, .., \beta_n\}$ and the assumption that the overlap between the classes is small relative to the size of the classes, Bacchus et al. interpret proportion formulas of the form $\%(Ax|\, \beta_i(x)) \approx f_i$ as each providing evidence of weight f_i for the proposition Ac (where Ax is a one-place predicate). They present a theorem stating that the random-worlds method provides the rule of combination as a solution to certain cases involving incomparable reference class statistics.

$$\frac{\prod_{i=1}^{m} f_i}{\prod_{i=1}^{m} f_i + \prod_{i=1}^{m}(1 - f_i)}$$

[98] Pollock, "Probable Probabilities (with Proofs)," 1 and Thorn, "Three Problems."
[99] Reichenbach, *The Theory of Probability*, 375.
[100] Kyburg and Teng, *Uncertain Inference*, 251.
[101] I carefully use the words "is the interval" because, for Kyburg and Teng, evidential probabilities are just intervals of the form $[f, g]$ where f and g are reals in $[0, 1]$ such that either $f = g$ or $f \neq g$. See Kyburg and Teng, *Uncertain Inference*, ch. 9.
[102] Bacchus et al., "From Statistical Knowledge Bases," 116.

However, the theorem states that this result holds given four conditions: 1) $f_i \in (0,1)$ for each formula $\%(Ax|\, \beta_i(x)) \approx f_i$ mentioned in the knowledge base, 2) none of the formulas $\beta_i(x)$ mention Ax or c, 3) the intersection of the classes $\boldsymbol{\beta}_i$ is known to consist of exactly one member that is denoted by c and 4) the knowledge base contains no other information.

This delivers an intuitively reasonable result in the case of Bernard. Let us idealise the knowledge base of the agent trying to assess the probability that Bernard has the disease. We can then instantiate the knowledge base in the above theorem so that the agent's knowledge base is such that $KB \equiv (\%(Dx|\, T_1x) = 0.7\ \&\ \%(Dx|T_2x) = 0.75\ \&\ T_1b\ \&\ T_2b\ \&\ \boldsymbol{T_1} \cap \boldsymbol{T_2} = \{b\})$. Then, as per Bacchus et al.'s theorem, $P_{Random-Worlds}(Db|KB) = 0.875$. Note that the agent's credence that Bernard has the disease is higher given knowledge of both of the positives than what it would be if the agent only had knowledge of one of the positives. This is an intuitive result. Both results of the test would individually make it probable that Bernard had the disease (with a probability of 0.7 or 0.75). Knowledge that one had positive results on both tests should, if anything, make it more probable that one has the disease; after all, surely this would be more worrying for the patient, if anything. Note that neither Kyburg, Teng or Reichenbach's theories get this result; Kyburg and Teng merely associate the probability with $[0.7, 0.75]$ and Reichenbach prescribes an agnosticism about the probability of Bernard having the disease given knowledge of the two positive tests.

Regardless, the rule of combination is surely not the final word on treating tensions with incomparable reference classes. This is because, as Pollock emphasises, base rate information (if present) should play a role in probability judgments about such tensions.[103] Consider the case of Bernard again, albeit with different statistics. Suppose the base rate for the disease among the domain of humans is low, so $\%(Dx) = 0.1$. Suppose further that the relative frequencies of the disease among those who obtain a positive result on the two tests are each 0.4, so $\%(Dx|\, T_1x) = \%(Dx|\, T_2x) = 0.4$. Suppose the agent knows that $\%(Dx) = 0.1$, $\%(Dx|\, T_1x) = \%(Dx|\, T_2x) = 0.4$, T_1b and T_2b where b stands for Bernard. If one applies the rule of combination formula above to determine the probability of Bernard having the disease, the probability is actually *decreased* by the two positives rather than increased. Pollock claims that such a result is absurd.[104] Indeed, a positive result on one of the tests would raise the probability of Bernard having the disease above what it originally was given

[103] Here, Pollock works with nomic probabilities concerning proportions among physically possible objects instead of statistical statements about the actual world like $\%(Dx|T_1x)$.
[104] Pollock, "Probable Probabilities (with Proofs)," 17.

only the base rate statistics (from 0.1 to 0.4).[105] It is then counterintuitive that getting a positive on another test would somehow lower this probability below 0.4 when the additional test would by itself similarly raise the probability above what it originally was.

Note that this does not constitute a counter-example to the random-worlds method's approach to the problem. The above theorem, and indeed the only proven theorem, specifying the method's approach to the problem makes the assumption that one does not have statistical knowledge of the base rate (recall condition four for the result). This suggests that more research is needed about how the random-worlds method copes with tensions between incomparable reference classes given knowledge of the base-rate. Because of the similarities between Pollock's account of probability and the random-worlds method, a useful source of insight as to the possible behaviour of the method may be Pollock's *Y-function*, a treatment of this tension that I find both very intuitive and facsinating.[106] Interestingly, Pollock's Y-function raises the probability of Db to 0.8 in the case in which the agent's knowledge base contains the statements $\%(Dx) = 0.1$, $\%(Dx|\, T_1x) = \%(Dx|\, T_2x) = 0.4$, T_1b and T_2b.[107]

4.2.8. Summarising the Insights of the Random-Worlds Method

So we have seen a number of favourable features of the random-worlds method. These features delineate the intuitive insights that it provides about what inductive inferences are rational in cases where our intuitions were readily clear and in others where our intuitions are perhaps initially less clear. These cases concern:

- Basic direct inference

[105] Bear in mind that Pollock is not adhering to a subjective interpretation of probability when he makes claims about the probability of Bernard having the disease. Regardless, the same points are arguably relevant to rational credence too.

[106] Joseph Halpern, one of the authors of the 1996 article "From Statistical Knowledge Bases to Degrees of Belief," suspected that the Y-function reflects the behaviour of the random-worlds method. Halpern forwarded to me (January 25, 2016) an email exchange between himself and John Pollock in which Halpern states, "Indeed, although I haven't checked carefully, my guess is that our approach gets the Y principle in general. Our theorem covered only Dempster's Rule because, at the time, we were just trying to relate our approach to Dempster's Rule." (Email communication dated September 23, 2006.)

[107] Of course, Pollock articulates the Y-function in accordance with his theory of nomic probability which is significantly different to the random-worlds method in some respects, particularly with its use of relative frequencies among physically possible worlds rather than actual relative frequencies. Regardless, random-worlds method is structurally similar to Pollock's account of nomic probability, as Pollock and Bacchus et al. have explicitly acknowledged.

- Probabilistic independence
- Rational monotonicity
- Specificity
- Precision
- The treatment between of the tension between overly specific and overly general reference classes
- Incomparable reference classes

The insights proffered by the random-worlds method, then, are its judgements on these kinds of cases.

4.3. Problems for the Random-Worlds Method

So we have explored the random-worlds method and some considerations favouring it. However, the method is not without its problems.

Bacchus et al. acknowledge a number of these. For example, they acknowledge that the random-worlds method faces computational difficulties and also encounters problems regarding reasoning about causality and time.[108] Nevertheless, they do not regard these problems as reasons to outright reject the method and they hint at potential solutions for them.

However, I wish to focus on another problem. The problem is that the random-worlds method does not give a promising account of how to generally accommodate uncertainty about proportion formulas. According to the method, a credence is calculated only in relation to a knowledge base; if a proportion formula features in the knowledge base, then the calculation of the credence treats the formula as if the agent was certain of it.

But we often are not even virtually certain of non-trivial and relevant proportion formulas on the basis of our evidence, yet the evidence still seems relevant to how we should constrain our credences. Suppose, for instance, that the agent hears a TV news report that all of the passengers on a particular train were fatally injured, but the agent is only 70% confident that the report said this train was the *Midnight Express* as opposed to some other train (perhaps because she was somewhat distracted at the time of partially hearing the report). If we suppose that she has no other relevant evidence, then surely this frequency evidence should

[108] Bacchus et al., “From Statistical Knowledge Bases,” 125-6 and 130-1.

constrain her credence for the proposition that her friend Aaron was fatally injured since she is certain that he was on the train. But surely the relative frequency of fatal injury on the train that is probably affirmed by the news report should not be accepted as knowledge so that she then has a credence of 1 that Aaron was fatally injured. After all, the agent is 30% confident, so to speak, that the news report did not concern the *Midnight Express* and so she should be open to the possibility that Aaron was not fatally injured. How then does one proceed?

One might claim that Bacchus et al.'s appeal to approximations intend to accommodate our uncertainty in cases like this, but this claim is doubtful. Suppose $\%(Ax|Bx)$ is the proportion of passengers on the train who were fatally injured. How might approximations help model the agent's situation above? Well, one suggestion is that the agent's situation should be modelled by her acceptance of approximate formulas such as $\%(Ax|Bx) \approx 1$ or $\%(Ax|Bx) \approx 0.7$. Let us consider problems for this suggestion. The agent clearly does not accept $\%(Ax|Bx) \approx 1$, especially given her openness toward the possibility that none of the passengers on the train were injured. This statistic would also generate an inappropriate credence of 1 that Aaron was fatally injured. Regarding the formula $\%(Ax|Bx) \approx 0.7$, the agent also simply does not have the information to accept this formula. She knows that the proportion of fatally injured passengers was either 1 or (at least approximately) 0; she has no information stating that the proportion of fatally injured passengers was approximately 0.7. Accepting such a statistic may nevertheless generate a reasonable random-worlds credence in the proposition that Aaron was fatally injured in the agent's case (given that she arguably should have a credence of 0.7 that Aaron was fatally injured anyway). But a substantive and tenuous assumption is needed if cases like these can be modelled in a way so that the agent hypothetically accepts statistical statements that she actually does not just to get the right probabilistic results.

The putative inability of the random-worlds method to accommodate the agent's uncertainty about proportion formulas is also highlighted when this uncertainty should arise from inferences from samples. Bacchus et al. write (using our notation so that Bx stands for "x is a bird" and Fx stands for "x can fly"):

> Under what circumstances is a statement such as $\%(Fx|Bx) \approx 0.9$ accepted as knowledge? Although we regard this as an objective statement about the world, it is unrealistic to suppose that anyone could examine all the birds in the world and count how many of them fly. In practice, it seems that this statistical statement would appear in KB if someone

inspects a (presumably large) sample of birds and about 90% of the birds in this sample fly. Then a leap is made: the sample is assumed to be typical, and we then conclude that 90% of all birds fly.[109]

As they note, though, the method "by itself does not support this leap", at least on the way that they represent the procedure of analysing a sample.[110] On this representation, the relative frequency of flying birds in a sample of birds is represented with the expression $\%(Fx|Bx\&Sx)$ where Sx stands for "x is in the sample". They state, however, that the random-worlds method fails to provide an adequate account of sampling on this representation since the method treats the members of the population inside the sample and the members of the population outside the sample as two unrelated populations.[111]

Another obvious but problematic way of representing the inference from a sample does not look promising either. On this approach, the sampled objects are not distinguished from the non-sampled objects via the predicate *S*; instead, the sampled objects merely appear as objects satisfying the predicate of interest in the KB. Suppose, for instance, the agent randomly samples nine balls from an urn which contains 10 balls. The agent is told that any ball could could be either green or red, but is told nothing about the proportion of balls of either colour. Perhaps all of them are red and none are green or perhaps all of them are green and none are red. The agent knows that the balls are each uniquely named with a constant of the form b_i where b_i is the ball labelled with an integer i in $[1,10]$. The agent randomly samples balls from the urn, draws balls 1 to 9 as a result and finds that all of them are green. What colour is the 10th ball to be sampled from the urn? Suppose for illustration that the domain is known to consist of only the 10 objects and that the language contains only the constants known to denote specific objects and the predicates Gx for "x is green" and Rx for "x is red". The sample, then, is represented not with a predicate *S*, but rather with the statement $(Gb_1 \,\&\, .. \,\&\, Gb_9)$ which features in the KB. Then, we can represent the models consistent with the agent's knowledge base as follows:

M_1: Gb_1 & Gb_2 & Gb_3 & Gb_4 & Gb_5 & Gb_6 & Gb_7 & Gb_8 & Gb_9 & Gb_{10}
M_2: Gb_1 & Gb_2 & Gb_3 & Gb_4 & Gb_5 & Gb_6 & Gb_7 & Gb_8 & Gb_9 & Rb_{10}

[109] Bacchus et al., "From Statistical Knowledge Bases," 128.
[110] Ibid.
[111] Bacchus et al., "From Statistical Knowledge Bases," 128.

The random-worlds method regards the two models as equally likely and so $P_{Random-Worlds}(Gb_{10}|KB) = P_{Random-Worlds}(Rb_{10}|KB) = P_{Random-Worlds}(\%(Gx) = 1|KB) = P_{Random-Worlds}(\%(Gx) = 0.9|KB) = 0.5$ where $(Gb_1 \,\&\, ..\,\&\, Gb_9) \in KB$. This is counter-intuitive as the evidence about the sample of green balls completely fails to significantly raise the probability that the 10th drawn ball will be green and that all balls are green.[112]

In this example of one way to represent an inference from a sample, we can see that the random-worlds measure clearly lacks *instantial relevance*. In other words, it lacks the property whereby the observation that a sampled object satisfies a predicate raises the probability (by however much) that other objects in the relevant population will satisfy the same predicate.[113]

Bacchus et al. express the hope that the random-worlds method would be capable of learning statistics from samples on another representation of knowledge about the domain.[114] Yet since the publication of their 1996 article some 20 years ago, neither them nor anyone else that I am aware of has developed this idea for me to see a promising representation; indeed, I think there is no such representation. In any case, Bacchus et al. state that they "do not have a good answer" to the question of when information should be accepted as knowledge.[115]

One way of summarising the difficulty confronting the random-worlds method, then, is that, although it is insightful as a theory of direct inference from proportion formulas about populations, it is inadequate as a theory of inductive inferences from samples or, in other words, inverse and predictive inferences (recall the taxonomy of inductive inferences discussed in Section 2.3).

So the method faces a considerable problem in accommodating the uncertainty that we can have about proportion expressions, uncertainty based on samples or other information (such as information regarding the aforementioned news report). We will consider a response to this difficulty in the next chapter.

---

[112] Carnap makes this criticism of a particular measure that is the analogue of the measure used in the random-worlds method. See Carnap, *The Logical Foundations of Probability*, 565.

[113] Sandy L. Zabell, "Carnap and the Logic of Inductive Inference," in *Handbook of the History of Logic*, vol. 10, eds. Dov M. Gabbay, and John Woods (Amsterdam; Boston: Elsevier, 2010), 281. Sandy Zabell characterises it differently. He phrases the desideratum as requiring that "if a particular type is observed, then it is more likely that such a type will be observed in the future."

[114] Bacchus et al., "From Statistical Knowledge Bases," 129.

[115] Ibid., 128.

5. Some Alternative Measures

As noted in the previous chapter, the measure employed in the random-worlds method faces various problems and, in particular, a difficulty in accommodating rational inverse inferences from samples.

One natural response to this difficulty is to attempt to modify the method by utilising alternative measures which hopefully preserve its insights.

This chapter will briefly outline two measures that stand out as salient candidate replacements for the measure used in the random-worlds method. This is because they are similar to the random-worlds method in that they are also probabilistically uniform or symmetrical, albeit with respect to different classifications of the possible outcomes.

5.1. Carnap's *m** and the Rule of Succession

In in a postscript of *The Logical Foundations of Probability,* Carnap proposed *m**, a measure similar to that of the random-worlds method.[116] *m** was articulated as a measure to underpin Carnap's *c**-function. The *m** and *c** functions are best understood in the context of Carnap's early philosophy of probability. A well-known and established distinction delineated by Carnap was the difference between two types of probabilities imaginatively called $probability_1$ and $probability_2$. $Probability_1$ was said to be a "degree of confirmation" and $probability_2$ was said to be a physical probability or, more specifically, a long-run relative frequency. (Later, however, Carnap instead preferred to speak of "subjective (or personal) probability" (i.e. credence) and "objective (or statistical) probability".)[117] The *c**-function was proposed as an explicatum of $probability_1$, that is, a function which explicates and makes precise and rigorous the vague notions we have about "the inference from the evidence to the hypothesis or, more correctly speaking, the logical relation holding between the evidence and the hypothesis."[118]

[116] Carnap, *The Logical Foundations of Probability,* 563.
[117] Rudolph Carnap, "The Aim of Inductive Logic," in *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress,* eds. Ernest Nagel, Patrick Suppes and Alfred Tarski (Amsterdam: North-Holland, 1966), 304-5.
[118] Carnap, *The Logical Foundations of Probability,* 33.

To understand $m*$, we first need to understand some terminology. For Carnap, a *population* is the set of objects that are of interest in a "statistical investigation" (to use his words).[119] Each object is called an *individual* and each individual is uniquely designated with a naming constant of the form a_i. It is analytically true that any two constants designate different individuals. Each individual satisfies exactly one *Q-predicate.*[120] Q-predicates are defined in term of primitive predicates. The primitive predicates in a language are the fundamental or atomistic predicates in a language. According to Carnap, Q-predicates are defined by "conjunctions in which every [primitive] predicate or its negation occurs."[121] For example, if a language has only three primitive predicates *P*, *R* and *S*, then Px & Rx & ∼Sx is a Q-predicate but Px & Rx is not since the predicate *S* or its negation fails to appear in the latter conjunction.

Put informally, a *structure-description* specifies how many objects satisfy any given Q-predicate. A structure-description corresponds to one or more conjunctive statements that Carnap called *state-descriptions*. A state description state specifies which Q-predicate every individual satisfies. For example, suppose there is a population of four individuals $\{a_1, \ldots, a_4\}$ and four Q-predicates $\{Q_1, \ldots, Q_4\}$ in the language. Then the conjunctive statements Q_1a_1 & Q_2a_2 & Q_3a_3 & Q_1a_4 and Q_1a_4 & Q_2a_2 & Q_3a_3 & Q_1a_1 are state-descriptions specifying which Q-predicate every individual satisfies. The state-descriptions Q_1a_1 & Q_2a_2 & Q_3a_3 & Q_1a_4 and Q_1a_4 & Q_2a_2 & Q_3a_3 & Q_1a_1 belong to the same structure-description. This is because, in both state-descriptions, the same number of objects satisfy a given Q-predicate and the state-descriptions only differ by the permutation of individuals. That is, in both state-descriptions, two individuals satisfy Q_1x, one individual satisfies Q_2x, one individual satisfies Q_3x, no individuals satisfy Q_4x and the state-descriptions only differ by permuting a_1 and a_4.

$m*$ then assigns each structure-description an equal measure and in turn distributes the measure for a given structure-description evenly over the corresponding state-descriptions. To illustrate this, suppose we have a language in which there are only two Q-predicates, Q_1 and Q_2, and constants naming four individuals $\{a_1, \ldots, a_4\}$. We can then construct the following sample space:

---

[119] Carnap's definition of a population is different to the one I give in Section 2.2. since the population I speak of may be a non-empty proper subset of the objects that one is reasoning about, a subset that is united by virtue of every member satisfying some particular formula. Carnap's population is instead the set of *all* objects that one is reasoning about. Carnap, *The Logical Foundations of Probability,* 207.

[120] Note that Carnap's later work did not revolve around Q-predicates.

[121] Carnap, *The Logical Foundations of Probability,* 122.

<u>A Sample Space for *m**</u>

Structure-description	**State-description**	**Measure for each state-description**	**Measure for each structure-description**
$\|\boldsymbol{Q_1}\| = 4$	Q_1a_1 & Q_1a_2 & Q_1a_3 & Q_1a_4	1/5	1/5
$\|\boldsymbol{Q_2}\| = 0$			
$\|\boldsymbol{Q_1}\| = 3$	Q_1a_1 & Q_1a_2 & Q_1a_3 & Q_2a_4	1/20	1/5
$\|\boldsymbol{Q_2}\| = 1$	Q_1a_1 & Q_1a_2 & Q_2a_3 & Q_1a_4	1/20	
	Q_1a_1 & Q_2a_2 & Q_1a_3 & Q_1a_4	1/20	
	Q_2a_1 & Q_1a_2 & Q_1a_3 & Q_1a_4	1/20	
$\|\boldsymbol{Q_1}\| = 2$	Q_2a_1 & Q_1a_2 & Q_1a_3 & Q_2a_4	1/30	1/5
$\|\boldsymbol{Q_2}\| = 2$	Q_1a_1 & Q_1a_2 & Q_2a_3 & Q_2a_4	1/30	
	Q_1a_1 & Q_2a_2 & Q_2a_3 & Q_1a_4	1/30	
	Q_2a_1 & Q_2a_2 & Q_1a_3 & Q_1a_4	1/30	
	Q_1a_1 & Q_2a_2 & Q_1a_3 & Q_2a_4	1/30	
	Q_2a_1 & Q_1a_2 & Q_2a_3 & Q_1a_4	1/30	
$\|\boldsymbol{Q_1}\| = 1$	Q_2a_1 & Q_1a_2 & Q_2a_3 & Q_2a_4	1/20	1/5
$\|\boldsymbol{Q_2}\| = 3$	Q_1a_1 & Q_2a_2 & Q_2a_3 & Q_2a_4	1/20	
	Q_2a_1 & Q_2a_2 & Q_2a_3 & Q_1a_4	1/20	
	Q_2a_1 & Q_2a_2 & Q_1a_3 & Q_2a_4	1/20	
$\|\boldsymbol{Q_1}\| = 0$	Q_2a_1 & Q_2a_2 & Q_2a_3 & Q_2a_4	1/5	1/5
$\|\boldsymbol{Q_2}\| = 4$			

where $|\boldsymbol{Q_i}|$ is the number of objects satisfying Q_ix.

Carnap's c^* is then defined as such for two fomulas α and β:

$$c^*(\alpha, \beta) = \frac{m^*(\alpha \,\&\, \beta)}{m^*(\beta)}$$

The measure avoids some unintuitive consequences of the random-worlds method. Suppose the sample space represents the number of balls in an urn that have some Q-property designated by Q_i and one learns that the randomly-sampled balls a_1, a_2 and a_3 satisfy the formula Q_1x. (Perhaps, for example, we could suppose that Q_1 represents being white and Q_2 represents being non-white.) According to the axioms of probability, the remaining state-descriptions are then assigned the following weights:

Structure-description	**State-description**	**Measure for each state-description**	**Measure for each structure-description**
$\lvert\boldsymbol{Q_1}\rvert = 4$ $\lvert\boldsymbol{Q_2}\rvert = 0$	Q_1a_1 & Q_1a_2 & Q_1a_3 & Q_1a_4	4/5	4/5
$\lvert\boldsymbol{Q_1}\rvert = 3$ $\lvert\boldsymbol{Q_2}\rvert = 1$	Q_1a_1 & Q_1a_2 & Q_1a_3 & Q_2a_4	1/5	1/5

Note that $c^*(Q_1a_4, Q_1a_1 \,\&\, Q_1a_2 \,\&\, Q_1a_3) = \frac{m^*(Q_1a_1 \,\&\, Q_1a_2 \,\&\, Q_1a_3 \,\&\, Q_1a_4)}{m^*(Q_1a_1 \,\&\, Q_1a_2 \,\&\, Q_1a_3)} = \frac{1/5}{(\frac{1}{5}+\frac{1}{20})} = 0.8$, that is, the degree of confirmation afforded to Q_1a_4 by the evidence $Q_1a_1 \,\&\, Q_2a_2 \,\&\, Q_3a_3$ is 0.8. Note how this sample evidence has affected the probabilities of proportion formulas. Given the sample evidence $Q_1a_1 \,\&\, Q_2a_2 \,\&\, Q_3a_3$, the formula $\%(Q_1x) = 1$ has a measure of 0.8 and the formula $\%(Q_1x) = 0.75$ has a measure of 0.2; hence, the sample evidence raised the probability of the formula $\%(Q_1x) = 1$ above its initial measure of 0.2. This is an intuitive result. If three of the four balls in the urn satisfy Q_1x is it reasonable to regard it as more likely that the next ball will also satisfy Q_1x and that all of the balls satisfy Q_1x. With

Carnap's $m*$ function, then, one can make useful inverse inferences from samples. This gives $m*$ an advantage over the measure used in the random-worlds method.

Interestingly, $m*$ also resembles measures that statisticians and probability practitioners sometimes use in practice.[122]

The $c*$-function is an instance of Laplace's rule of succession which, in this context, takes the following form:

$$c^*(Q_i a_{n+1}, e_n) = \frac{n_i + 1}{n + k}$$

where e_n is a sentence stating the individuals in a sample the satisfy or fail to satisfy some particular predicate Q_i, n_i is the number of individuals in the sample satisfying Q_i, n is the total number of individuals in the sample, $Q_i a_{n+1}$ is the statement that the next sampled individual will satisfy Q_i and k is the number of Q-predicates (including Q_i) that the members of the population may possibly satisfy.[123]

For example, consider the case of the urn again. Let $Q_i a_{n+1}$ be $Q_1 a_4$ in order to consider the probability that the next sampled individual a_4 will satisfy $Q_1 x$. e_n in this case is the sample evidence $Q_1 a_1$ & $Q_1 a_2$ & $Q_1 a_3$. $n_i = 3$ since there are three individuals in the sample satisfying Q_1; note also that $n = 3$. $k = 2$ since the members of the population may possibly satisfy only two Q-predicates, Q_1 and Q_2. Therefore, we have the following equation:

$$c^*(Q_1 a_4, Q_1 a_1 \,\&\, Q_1 a_2 \,\&\, Q_1 a_3) = \frac{n_i + 1}{n + k} = \frac{3 + 1}{3 + 2} = 0.8$$

So that concludes our description of $m*$, a measure which has some favourable features by virtue of enabling useful inferences from samples.

However, $m*$, as Carnap proposed it, is not problem free. One problem arises when considering infinite sets of individuals. Carnap, similarly to Bacchus et al., uses limits for $m*$

[122] I say this based on some conversations that I have had with Brendon Brewer.
[123] Ilkka Niiniluoto, "The Development of the Hintikka Program," in *Handbook of the History of Logic*, vol. 10, eds. Dov M. Gabbay, and John Woods (Amsterdam; Boston: Elsevier, 2010), 316. Laplace's rule of succession is often presented in cases where $k = 2$; however, according to Zabell, Laplace also presented a more complex form of the rule where k can be an integer greater than two. Sandy L. Zabell, "Predicting the Unpredictable," *Synthese* 90, no. 2 (1992): 207-8.

as the number of individuals or objects tends to infinity, although this is only in situations with an population of infinite size rather than just a domain with an unknown but presumably "large" size.[124] This limit construction allows for evidence to affect the probability of many propositions, but it entails that a probability of 0 is always assigned to universal generalisations for infinite domains (aside from trivial and tautological generalisations specifying that every individual satisfies the disjunction of the k Q-predicates ($Q_1 x \lor \ldots \lor Q_k x$)). To see an example of how this is so, suppose that the above sample space instead had 20 individuals. Then there would be 21 structure-descriptions representable with the set $\{|\boldsymbol{Q_1}| = 20 \,\&\, |\boldsymbol{Q_2}| = 0, |\boldsymbol{Q_1}| = 19 \,\&\, |\boldsymbol{Q_2}| = 1, \ldots, |\boldsymbol{Q_1}| = 0 \,\&\, |\boldsymbol{Q_2}| = 20\}$. In this case, the universal generalisation or structure-description in which all objects satisfy $Q_1 x$ would have a lower measure at $1/21$. Likewise, if there were 50 individuals, then there would be 51 structure-descriptions and the one in which all objects satisfy $Q_1 x$ would have an even lower measure at $1/51$. We can see, then, that as the population size grows and goes to infinity, the probability of this structure-description tends to 0. Consequently, the limit construction assigns zero-valued probabilities to certain universal generalisations.

This consequence is worrying in several ways, at least if $m*$ and the corresponding $c*$-function are taken to be instructive regarding what credences are rational. For one, it is reasonable to be open minded about certain infinite universal generalisations. An example of this is the generalisation that all macro-level objects in the past, present and future obey a certain theorised law of physics (although perhaps only to a degree of approximation).[125] Furthermore, the probability of such generalisations should be at least raised given some confirming instances of objects conforming to the generalisation, something that the $m*$ and $c*$ functions prohibit. To make matters worse, this raises concerns about how much $m*$ and $c*$ can guide the work of scientists who sometimes seem open toward, if not confident in, universal generalisations in infinite domains. Carnap stated, however, that "the role of universal sentences in the inductive procedures of science has generally been overestimated."[126] Rather, he thought that the laws expressing universal generalisations were "efficient instruments for finding those highly confirmed singular predictions which we need

[124] Carnap, *The Logical Foundations of Probability*, 302.

[125] Examples of such infinite universal generalisations are that all macro-level objects attract each other with a force (perhaps approximately) proportional to their masses, that all macro-level objects (approximately) obey Newton's law of universal gravitation and that all objects obey the law of the conversation of momentum. Physicists may not outright believe that these laws apply to an infinite number of objects, but they may attach some positive probability to it. Brendon Brewer (himself a physicist) thinks there are physicists who do precisely this.

[126] Carnap, "On Inductive Logic," 88.

for guiding our actions."[127] Regardless, A.J. Ayer argues that a probability of 0 for an infinite universal generalisation can sometimes have an odd and unintuitive implication.[128] A universal generalisation is false iff there is at least one instance of an object not conforming to the generalisation. For example, the law that all macro-level objects (approximately) attract each other with a force proportional to their masses is false if there is an instance of, say, two objects not attracting each other with a force approximately proportional to their masses. According to Ayer, $c*$ prescribes a probability of 1 that at least one individual exists in the universe to disconfirm any infinite universal generalisation. Such a probability is often uncalled for and, even if it is called for, it would presumably not be as such *just because* the population of individuals is infinite, a point that $c*$ fails to account for.

In certain cases, however, it can be reasonable for the probability of universal generalisations to be partly dependent on the population size. For example, suppose there is a factory that has produced a certain number of toy cars at a time T_1 and at a subsequent time T_2. At T_1, the factory has produced two cars. At T_2, the factory has produced 2,000 cars. What is the probability that that all of the produced toy cars are black at T_1 and T_2? Some would intuitively feel that the probability of the universal generalisation that all of the cars are black should be lower at T_2 than at T_1. The reason is that the generalisation at T_2 is logically stronger – in a sense, it makes 1,998 more claims than the generalisation at T_1. On this note, Carnap says that this dependence "seems plausible because, the larger [the population size is] is, the more is asserted" by the generalisation.[129] This suggests that sometimes the probability of a generalisation should depend on the population size.

So $m*$ might be a reasonable measure if the population size is finite and known.

Regardless, even if $m*$ could adequately accommodate intuitions in some cases like the toy car one above, it does not accommodate intuitions about universal generalisations in certain other cases, such as when one is reasoning about whether the laws of physics apply (approximately) to a potentially infinite number of objects in space-time.

[127] Ibid., 93.
[128] A.J. Ayer, *Probability and Evidence* (London: Macmillan, 1972), 38.
[129] Carnap, *The Logical Foundations of Probability*, 571.

5.2. Hintikka's Measures

Jaakko Hintikka attempted to solve this problem regarding infinite universal generalisations. To do this, he proposed some measures published in 1965.[130]

The measures make use of the logical concept of a *constituent*. A constituent specifies which of the k Q-predicates in a language are satisfied or not satisfied by at least one individual. The logical form of a constituent is as follows:

$$\pm\exists x(Q_1x)\& \dots \& \pm\exists x(Q_kx)$$

where $\pm$ is either omitted or substituted with a negation symbol $\sim$.

The *Jerusalem system* distributes the probability evenly to the constituents and then evenly over the state-descriptions corresponding to these constituents.[131] The *combined system* distributes the probability evenly to the constituents, then evenly over the corresponding structure-descriptions and then evenly over the state-descriptions corresponding to these constituents.[132] Each consistent universal generalisation in the language corresponds to one constituent or a disjunction of constituents (this is because two or more Q-predicates could, if applied to an individual, agree that the individual satisfies some primitive predicate(s), even though they will disagree with respect to some other primitive predicate(s)). Such correspondences are stated in what are known as *distributive normal forms*.

To illustrate how these measures work, suppose there is a language with four constants naming individuals and only two Q-predicates. We then have the following sample space:

[130] Jaakko Hintikka, "Towards a Theory of Inductive Generalization," in *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*, 2nd ed., ed. Yehoshua Bar-Hillel (Amsterdam: North-Holland Pub Co., 1965), 274–288 and Jaakko Hintikka, "On a Combined System of Inductive Logic," in *Studia Logico-Mathematica et Philosophica in Honorem Rolf Nevanlinna die Natali eius Septuagesimo* (Helsinki: Societas Philosophica Fennica, 1965), 21-30.
[131] Hintikka, "Towards a Theory."
[132] Hintikka, "On a Combined System."

Constituent	$\lvert \boldsymbol{Q_i} \rvert$	State-description	Jerusalem system measure	Combined system measure
$\exists x(Q_1x)\&\sim\exists x(Q_2x)$	$\lvert \boldsymbol{Q_1} \rvert = 4$ $\lvert \boldsymbol{Q_2} \rvert = 0$	Q_1a_1 & Q_1a_2 & Q_1a_3 & Q_1a_4	1/3	1/3
$\exists x(Q_1x)\&\exists x(Q_2x)$	$\lvert \boldsymbol{Q_1} \rvert = 3$	Q_1a_1 & Q_1a_2 & Q_1a_3 & Q_2a_4	1/42	1/32
	$\lvert \boldsymbol{Q_2} \rvert = 1$	Q_1a_1 & Q_1a_2 & Q_2a_3 & Q_1a_4	1/42	1/32
		Q_1a_1 & Q_2a_2 & Q_1a_3 & Q_1a_4	1/42	1/32
		Q_2a_1 & Q_1a_2 & Q_1a_3 & Q_1a_4	1/42	1/32
	$\lvert \boldsymbol{Q_1} \rvert = 2$	Q_2a_1 & Q_1a_2 & Q_1a_3 & Q_2a_4	1/42	1/54
	$\lvert \boldsymbol{Q_2} \rvert = 2$	Q_1a_1 & Q_1a_2 & Q_2a_3 & Q_2a_4	1/42	1/54
		Q_1a_1 & Q_2a_2 & Q_2a_3 & Q_1a_4	1/42	1/54
		Q_2a_1 & Q_2a_2 & Q_1a_3 & Q_1a_4	1/42	1/54
		Q_1a_1 & Q_2a_2 & Q_1a_3 & Q_2a_4	1/42	1/54
		Q_2a_1 & Q_1a_2 & Q_2a_3 & Q_1a_4	1/42	1/54
	$\lvert \boldsymbol{Q_1} \rvert = 1$	Q_2a_1 & Q_1a_2 & Q_2a_3 & Q_2a_4	1/42	1/32
	$\lvert \boldsymbol{Q_2} \rvert = 3$	Q_1a_1 & Q_2a_2 & Q_2a_3 & Q_2a_4	1/42	1/32
		Q_2a_1 & Q_2a_2 & Q_2a_3 & Q_1a_4	1/42	1/32
		Q_2a_1 & Q_2a_2 & Q_1a_3 & Q_2a_4	1/42	1/32
$\sim\exists x(Q_1x)\&\exists x(Q_2x)$	$\lvert \boldsymbol{Q_1} \rvert = 0$ $\lvert \boldsymbol{Q_2} \rvert = 4$	Q_2a_1 & Q_2a_2 & Q_2a_3 & Q_2a_4	1/3	1/3

Hintikka's systems avoid the problem of universal generalisation in cases with a finite number of Q-predicates. (Although a problem would remain for the systems if there was a language with an infinite number of predicates.)[133] Note that with the addition of individuals to the sample space, the probability of the constituents stays the same; for example, no matter how many individuals there are (but assuming that there is at least one), a probability of 1/3 is assigned to the constituent $\exists x(Q_1x)$ & $\sim\exists x(Q_2x)$ and, by implication, the generalisations $\forall x(Q_1x)$ or $\%(Q_1x) = 1$. Hence, a universal generalisation corresponding to some constituents may have a positive probability even as the population size goes to infinity. This is an advantage of the measure.

But it is sometimes also a disadvantage. Recall the toy car case in which the probability of the generalisation should intuitively be partly dependent on the population size. Hintikka's measures are inadequate here. We will discuss this problem again later on with respect to varying background evidence in Sub-Section 6.3.2

In any case, the preceding discussion raises a general worry for the measures. We have seen that the probabilities of (non-trivial) proportion formulas such as $\%(Q_1x) = 1$ are crucially sensitive to the structure of the sample space which represents the individuals in the population. But if the number of individuals in the population is unknown, then the appropriate sample space is also unknown, the set of possibly true proportion formulas is unknown and the appropriate probability measure for those formulas is unknown. This creates a further difficulty in applying the measure in the many cases involving inverse inferences for populations of objects with unknown sizes.

So that concludes my brief survey of some salient, albeit problematic, measures which resemble the measure employed in the random-worlds method. Note that other measures have also been proposed in the literature.[134] Regardless, the preceding ones seem to me to be the most prima facie plausible and relevant alternative measures to the random-worlds measure.

[133] I do not know whether there would actually be any use for such a language, but there might be. For example, one might wish to estimate the height of the next human given a sample of humans. To use a language which does not make any contingent assumptions about the possible height of humans, one might then wish to employ a language with an infinite number of height predicates.

[134] For example, Bacchus et al. consider a measure for what they call the "random-propensities approach". Unfortunately, this approach does not seem particularly insightful to me. Given "Proposition 5" outlined in "From Statistics to Beliefs", this measure validates what is known as the *straight rule*. In this context, the rule is that the probability that an object is an A is equal to the relative frequency of As observed in a sample of the domain (assuming that the object is not in the sample). This rule faces many problems, including the unintuitive consequence that if the sample consists of *just one* object that is an A and so the relative frequency is 1, then the

6. Evaluating the Alternative Measures

We have examined some salient candidates for measures that could replace the one used in the random-worlds method, as well as some problems for them. In this chapter, we will evaluate these measures more thoroughly.

What follows, then, is a discussion of the favourable features of these measures followed by a compendium of problems for them.

6.1. An Auto-Biographical Background

Before evaluating these measures, however, it may be useful to explain this chapter's purpose and (relative) brevity in relation to the background of my research project.

The initial impulse for undertaking research for this thesis was a dissatisfaction with subjective Bayesianism and a desire to find objective constraints on rational credences (on some understanding of those terms). Theories appealing to frequencies were natural candidate constraints, especially since every scholar in the area acknowledges that frequencies have some role in judgments of probability. Some of the principles and theories I then explored were Reichenbach's *straight rule* and his idea of *cross-induction*, Pierre-Simon Laplace's *rule of succession*, Pollock's theory of *nomic probability*, Bacchus et al.'s *random propensities approach*, Kyburg's theory of *evidential probability* and classical frequentist and Bayesian theories that are utilised in statistics. At a later stage in my research, I hoped to develop a theory of Bayesian inverse inference which resembled Carnap's *m**-function and drew on statistical practice. I also sought a role for Hintikka's measures in inverse inference. However, after pursuing this direction and hoping that it would work, the problems posed by varying background evidence, similarity and other topics made me pessimistic of finding an adequate, universally applicable and formal theory of inverse inference, at least any time soon (if at all). While I believe that the aforementioned ideas and theories are helpful in many contexts, I believe even more strongly that there are countless important contexts in which

---

agent should be *certain* with probability 1 that the next observed object from the population will also be an *A*. See the random propensities approach in Fahiem Bacchus, Adam Grove, Daphne Koller and Joseph Y. Halpern, "From Statistics to Beliefs," in *AAAI-92: Proceedings* (Menlo Park, Calif.; Cambridge: MIT Press, c1992), 602-608. For one articulation of the straight rule, see Reichenbach, *Theory of Probability*, 446. For criticism of the rule, see Rudolf Carnap, "A Basic System of Inductive Logic, Part II," *in Studies in Inductive Logic and Probability*, ed. Richard Jeffrey, vol. 2 (Berkeley: University of California Press, 1980), 85-86. Bacchus et al. also articulate some problems for the random-propensities approach in Bacchus et al., "From Statistical Knowledge Bases," 129.

they are not. Consequently, I was lead to the theory that I articulate in Chapter 7. My purpose in noting this background here is to avoid giving the impression that I have lazily and hastily advanced the pessimism in this chapter and the theory in the next merely because I did not make a decent effort to avoid it.

However, the various considerations that lead me to my pessimism are not articulable in a thesis that wants to say something positive in 40,000 words. It would be an arduous task to explore in formal detail, for example, every potential response one might give to each of the problems in this chapter in order to defend the measures and other formal methods that I am familiar with.

Consequently, the discussion that follows may not suffice to vindicate my pessimism in the eyes of the reader, but it aims to provide some motivation for it. For that, I apologise and hope that the reader can see my pessimism as being at least potentially reasonable. Even if they are not sympathetic with my pessimism, the reader can perhaps appreciate chapters 7 and 8 as being a conditional claim of the sort "If my pessimism is justified, then here is what one can do."

Regardless, let us explore various problems for the alternative measures. I will argue that these problems suggest that there is currently no adequate probability measure which is generally suitable for inverse and predictive inference. For this reason, I will not examine in-depth how these alternative measures may or may not be integrated with the ideas from the random-worlds method, particularly in light of the differences between the alternative measures and the method with respect to the use of *Q*-predicates in place of primitive predicates and the focus on types of sentences rather than models. In the next chapter, the thesis will then offer another theory of inverse inference in the absence of a generally suitable for measure for inverse inference. The following discussion, then, aims to be sufficiently concise so as to allow space for this theory in the thesis while also sufficiently elaborate so as to somewhat motivate the contention that there are serious problems facing (the development of) a general method for inverse inference.

6.2. Favourable Features of the Measures

Despite the problems that the measures face, they share some putatively favourable features in common with the random-worlds method.

Each of them is uniform over various classifications of the outcomes. The main difference between the measures, then, is the order of priority for uniformity. Hintikka's measures prioritise uniformity over the constituents which classify sets of outcomes. Carnap's $m*$ prioritises uniformity over the structure-descriptions which classify sets of outcomes. The random-worlds method prioritises uniformity over the models. Uniformity in some form or another is arguably favourable if there is no reason to skew the distribution.

The similarities in uniformity also appear to entail favourable similarities in results. For one, it seems that all of the measures validate versions of basic direct inference by virtue of regarding certain sets of outcomes as equally likely (although I do not have a proof stating something to this effect, nor is it necessary to give one here given that these measures are not the focus of this thesis). For example, suppose we have a language with one primitive predicate, two Q-predicates, Q_1 and Q_2, and four constants $\{a_1, \dots, a_4\}$ designating four individuals. The agent also accepts the statement $\%(Q_1x) = 0.75$. Suppose the agent is then interested in the probability that a_1 satisfies the formula Q_1x. According to the measures of Hintikka and Carnap, the outcomes represented by the below conjunctions are equally probable:

O_1: Q_1a_1 & Q_1a_2 & Q_1a_3 & Q_2a_4

O_2: Q_1a_1 & Q_1a_2 & Q_2a_3 & Q_1a_4

O_3: Q_1a_1 & Q_2a_2 & Q_1a_3 & Q_1a_4

O_4: Q_2a_1 & Q_1a_2 & Q_1a_3 & Q_1a_4

Hence, for Hintikka and Carnap's measures, the probability of Q_1a_1 given $\%(Ax) = 0.75$ is 0.75, thus illustrating an instance of basic direct inference. We have also already seen a proof that the random-worlds method validates basic direct inference in Sub-Section 4.2.1. I suspect that the measures share other favourable features, such as the validation of assumptions of independence. Nevertheless, given the criticisms that I shall make of the measures and the overall direction of the thesis, it is not important to thoroughly explore their favourable properties here.

6.3. Problems for the Measures

And on that note, there are general problems that these measures and certain other theories face.

Each of the problems arises when applying the measures to deal with one or more problematic cases.

A legitimate question arises, then, as to whether Carnap and Hintikka intended these measures to be applied to these cases. I have not seen either Carnap or Hintikka specify exactly to what cases these measures may apply for useful inverse inferences.[135] I have not seen, for example, either of them specify that the measures were intended only for samples that are randomly-selected from populations. Regardless, both of them later eschewed one universally applicable measure for predictive or inverse inference. As we shall see, Carnap ultimately endorsed a system of inductive logic that allowed the values of certain parameters to vary depending on the context and Hintikka lost interest in inductive logic altogether.

Regardless, the following discussion is nevertheless useful insofar as it articulates problems confronting theories of inverse inference in general. It does this by way of considering how the measures face up to various problem cases regardless of whether Carnap and Hintikka attempted to address these cases with their measures. These problems will in turn provide some motivation for considering the alternative theory of inverse inference which I discuss in Chapter 7.

The remainder of this section, then, will review some of these problems and, where possible, briefly outline some potential avenues of response to them.[136]

6.3.1. The Problem of Non-Probability Sampling

The alternative measures appear to possess two putatively reasonable features. One is the validation of basic direct inference (see Sub-Section 4.2.1.). By this, I mean (very loosely speaking) that a probability assignment of the form $\mu(Aa) = f$ given $\%(Ax|Bx) = f$ & Ba is

[135] Regardless, Carnap did affirm that $m*$ could be used for inverse inference, although he did not affirm that he thought $m*$ was a perfect or universally adequate measure. He just saw it as the best measure at the time of writing *The Logical Foundations of Probability*. See Carnap, *Logical Foundations of Probability*, 563 and 570.

[136] I have not discussed certain other problems, such as how the measures may or may not be used when the agent is uncertain about exactly what their evidence is. An example of when one is uncertain of their evidence is when they only know that between 90 to 100 of the objects in their sample satisfy a target formula.

prescribed by the measures where μ is some probability function. Another feature of the alternative measures is that they enable evidence concerning samples from populations to affect the probabilities of proportion formulas for those populations (perhaps on the condition that there are a finite number of individuals, that is).

When applied to everyday situations in which agents make inductive inferences, both features of the measures are susceptible to criticism. Let us consider how this is so.

Levi argues that one is not always rationally obligated to perform a direct inference according to an assignment of the form $\mu(Aa) = f$ given $\%(Ax|Bx) = f$ & Ba.[137] To use Levi's own example, suppose that the agent knows that 90% of Swedes are Protestant and that Petersen is a Swede; this is all that the agent knows about Petersen.[138] Levi sees no compelling reason why the agent should have a credence of 0.9 that Petersen is a Protestant. This is because the agent "does not know whether the way in which Petersen came to be selected for presentation to him is or is not in some way biased in favor of selecting Swedish Catholics with a statistical probability, or chance, different from the frequency with which Catholics appear in the Swedish population as a whole."[139] Consequently, Levi favours the general requirement that one knows that the selection mechanism proffers a chance of this kind that is equal to the frequency.[140] Some interpret him as stating that the kind of direct inference which equates the credence with the relative frequency is legitimate only if the object of interest is selected *at random* from the reference class.[141]

So that is one criticism of applying the measures to samples of an everyday sort where the sampled object is not randomly selected from the reference class.

Yet an analogous criticism one may come from the social sciences. Certain social scientists have objected to various kinds of inferences from so-called *non-probability samples*.

Let us spell out the distinction between probability and non-probability samples, a distinction that will also be useful in the next chapter. In probability sampling, the each member of the population of interest has some positive probability for inclusion in the sample and, furthermore, that probability is known.[142] The most well-known type of probability

---

137 Isaac Levi, "Direct Inference," *The Journal of Philosophy* 74, no. 1 (1977): 9.
138 Ibid., 9.
139 Ibid., 9-10.
140 Ibid., 10.
141 Pollock, *Nomic Probability,* 114 and Roger White, "Evidential Symmetry and Mushy Credence," in *Oxford Studies in Epistemology*, vol. 3, eds. Tamar S. Gendler and James Hawthorne (Oxford: Oxford University Press, 2010), 170.
142 Samuel R. Lucas, "Beyond the Existence Proof: Ontological Conditions, Epistemological Implications, and In-Depth Interview Research," *Quality & Quantity* 48, no. 1 (2014): 393-4. It is not clear to me exactly how critics of non-probability sampling generally understand the "probability" in probability sampling. However, it

sample is a random sample where each member of the population has an equal probability of being included in the sample.[143] To give an example of a probability sample, suppose we have an electoral roll of every resident in a given town. We could then use this as a so-called *sampling frame* from which we could select the objects (in this case, humans) for inclusion in the sample. Each object listed on the frame could be assigned a unique number and then a random number generator can be used to select these units for inclusion according to their number. They might then make various inferences from this sample. This is a typical instance of random sampling.

Contrast probability sampling to what is a second type of sampling known as *non-probability sampling*. Here, some object(s) in the population either does not have a non-zero probability for inclusion in the sample or the probability of inclusion is unknown. A typical form of non-probability sampling is *convenience sampling* whereby the sample is selected by a process that is convenient but fails to satisfy the demands of probability sampling. For instance, the agent might estimate the prevalence of open-mindedness in a town based on a sample of her friends from that town. This is a sample that happens to be conveniently accessible to the agent but has a zero probability of including residents in the town who are not (or could not be) in her circle of friends.

A problem arises in applying the measures to make inferences from the non-probability samples that we encounter in every-day contexts.

Some social scientists appear to disparage inverse inferences from non-probability samples in general. For example, speaking of non-probability samples, sociologist Samuel Lucas states (or maybe even generalises) that it "is well-known that such samples prohibit out-of-sample generalization."[144] Similarly, sociologists Thomas Dietz and Linda Kalof say that, with convenience sampling, "there is no way to know the relationship between the sample and the population and thus no way to speak about the population."[145] The concern here is that the non-probability sampling method may introduce bias into the sample so as to make it unrepresentative of the population that it is extracted from. One might interpret them

probably includes the "probabilities" conferred via the use of random number generators, and perhaps it includes other (quasi-)random processes, like drawing cards from a deck. Whatever the case, these critics probably do not have in mind a subjective interpretation of probabilities, and maybe even they do not know exactly what they have in mind.

[143] Steven K. Thompson, *Sampling*, 3rd ed. (Hoboken, NJ: John Wiley & Sons, 2012), 2.

[144] Lucas, "Beyond the Existence Proof," 394.

[145] Thomas Dietz and Linda Kalof, *Introduction to Social Statistics: The Logic of Statistical Reasoning*, (Chichester, West Sussex; Malden, MA.: Wiley-Blackwell, 2009), 259.

as implicitly qualifying their statements so that they only concern situations in the social sciences. However, there is no indication that they have such a qualification in mind.

These concerns from Levi and the social scientists share something in common. They both would criticise the measures as validating unreasonable inferences in the absence of a probability sampling strategy. In Levi's case, a direct inference according to a probability assignment of the form $\mu(Aa) = f$ given $\%(Ax|Bx) = f$ & Ba is illegitimate unless a is sampled from the population of *B*s via random sampling. In the case of the social scientists, inverse inference is illegitimate unless the sample is selected via probability sampling. According to each, probability sampling of one form or another is required for the relevant inductive inference so as to minimise the potential for bias in obtaining the sample.

Philosophers have responded to these concerns by noting that such requirements are inconsistent with epistemically rational practices in which agents assign probabilities on the basis of evidence. In reply to Levi, Pollock asserts that agents typically consider an object of interest and *then* find an appropriate reference class for estimating the probability that it satisfies some target formula.[146] For example, insurance agencies may consider an individual's claim for health insurance by noting that the individual is a smoker and then using the reference class of smokers to estimate the probability that the individual will incur significant medical costs. Pollocks says that this contrasts to Levi's requirement whereby one starts off with statistics about a reference class and *then* randomly selects the object of interest from it. Pollock further asserts that even though there could be some bias along the lines that Levi suggests in his example regarding Petersen, it is not necessary to rule such a possibility out. Instead, the direct inference provides a *prima facie* reason for accepting the conclusion that Petersen is a Swede. Roger White and Thorn likewise object to Levi, with White also noting that Levi's requirement is in tension with the practices of statistical reasoning which people commonly engage in (including those in insurance companies).[147]

Kyburg and Teng also defend the legitimacy of inverse inferences on the basis of non-probability samples. We can use an adapted example of theirs to illustrate their point. Suppose we toss a coin many times to estimate whether it is unbiased or, in other words, whether the relative frequency of landing heads in the population of all tosses of the coin (past and present) is approximately 1/2 under normal conditions.[148] A sample of tosses is, they would say, perfectly adequate for this purpose, even if the sample is a non-probability

[146] Pollock, *Nomic Probability*, 114-115.
[147] White, "Evidential Symmetry," 170-171 and Thorn, "Two problems," 306, fn. 15.
[148] Of course, this example presupposes Kyburg and Teng's frequentist account of probability, but the example is arguably insightful for non-frequentist accounts too. Kyburg and Teng, *Uncertain Inference*, 255.

one since not all tosses of the coin in the past could have been included in one's sample. They eschew the universal requirement of probability sampling and instead claim that the sample should be regarded as fair unless there are grounds for impugning its fairness.[149]

There are also plenty of other examples in which non-probability samples should affect our credences for proportion formulas, some of which I shall give in Chapter 7.

In any case, I take it as reasonable that inductive inferences can be made both about and from non-probability samples.

6.3.2. The Problem of Varying Background Evidence

One problem facing the measures is that they may be rarely, if ever, useful in situations of actual interest since the agent's background evidence may preclude their fruitful application. Recall how Hintikka's systems may be more applicable when confirming certain laws of physics but *m** may be more applicable when reasoning about the colour of toy cars produced by a factory. This suggests that neither measure is universally applicable.

An example of Reichenbach's may further illustrate this worry.[150] Consider how the continued observation of white swans putatively justified the universal inference to the generalisation that all swans were white. Hintikka's measure, *m** or a modified version of either may seem applicable here. To formalise this inference, one may, for example, then have a (perhaps idealised) sample space involving finitely many individuals (representing swans) and two predicates to symbolise being white or not white. The measures will then distribute the probability over the outcomes in some uniform manner and this will enable the evidence of white swans to support the generalisation that all swans are white. However, Reichenbach asserts that the universal inference that all swans are white was illegitimate in any case since colour was known to not be constant in other species of animals. Hence, the inference to the generalisation, Reichenbach claims, is undermined by an instance of so-called *cross-induction* from the features of other species. Presumably, then, this background evidence warranted lowering the probability for the generalisation. But the extent which it

[149] For Kyburg and Teng, such grounds involve knowledge of impugning statistics for alternative reference classes that the sample belongs to. For example, the agent might have sample of her close friends from Texas and hope to estimate the height of the population of Texans given the sample evidence. But she knows that all of her close friends from Texas are basketball players and that, in general, basketball players are unusually tall among Texans. Hence, she has grounds for impugning the fairness of her sample since it belongs to a reference class of basketball players with statistics for height that are known to be unusual of Texans in general. Kyburg and Teng, *Uncertain Inference*, 270.

[150] Reichenbach, *Theory of Probability*, 430.

should have done so depends on the evidence specific to the case at hand such as the number of other species for which colour is known to not be constant. Consequently, applying the measures over a sample space in the aforementioned manner is insensitive to how this background evidence may vary; the probabilities depend on the language, the population size, the sample of swans and *not* the evidence about other species of animal.

Varying background evidence creates problems in applying the measures in many other cases. Indeed, a similar cross-induction from the mortality of others also arguably undercuts the inference to the agent's own (probable) immortality from the fact that she has lived every day since her birth. (Recall the opening quote in the introduction of this thesis.)

This illustrates the general concern that credences should and often do reasonably differ from context to context while taking into account the varying and often complex background evidence local to the situation at hand. This variation arguably precludes the use of a universally applicable measure.

6.3.3. The Problem of Language Relativity

A further problem arises from the choice of language to delineate the possible outcomes.

Let us use an example to illustrate the problem. Suppose the agent has an urn in front of her that contains one ball of an unknown colour and she has no information that limits the possible colours that it could be. Suppose she is interested in assigning a value to her credence that the ball is green. Here, she could chose different languages, each of which correspond to a different sample space. For example, suppose she chooses a language with one constant *b* that represents the one object or individual and one predicate *G* which stands for greenness. In this case, the relevant sample space of outcomes can be represented as such:

Sample Space 1:

$O_1 : Gb$

$O_2 : \sim Gb$

However, suppose she chooses a different language, one in which there are four predicates $\{Gb,\ Yb, Rb, Sb\}$ where *Gx* stands for "*x* is green", *Yx* stands for "*x* is yellow", *Rx* stands for "*x* is red" and *Sx* stands for "*x* is something else", i.e. a colour aside from the others mentioned. Then, the relevant sample space can be represented as such (taking into account the logical relations between the predicates so that the object must satisfy exactly one of the predicates):

Sample Space 2:

$O_1 : Gb$

$O_2 : Yb$

$O_3 : Rb$

$O_4 : Sb$

This shows how language can influence the set-up of the relevant sample space. Note also that we could construct a tremendous number of different sample spaces with a tremendous number of different languages (consider, for example, the many languages with complex colour predicates such as "*x* is red with blue dots and the number *n* displayed on it in purple"). (However, note that on Bacchus et al.'s account, the sample space does not automatically take into account logical relations between predicates by, say, incorporating the fact that no green thing is also yellow. Instead, these logical relations appear in the knowledge base KB and the measure over the sample space is adjusted given the knowledge that no green thing is also yellow and so forth.)[151]

The problem arises because the choice of language affects certain probabilities in virtue of it affecting the sample space and it is not obvious that there is always one choice of language that is *the* correct one.

Let us see how this choice of language affects probabilities. The random-worlds method, *m** and Hintikka's measures all assign a prior probability of 1/2 to the outcome where the ball is green on the first space, but they assign a probability of 1/4 to that same outcome in the second space. Hence, we have an example in which the choice of language influences the probabilities. Yet the choice of language can also affect probabilities given some evidence. Let us illustrate this with another example using $m*$. Suppose we have two languages, each with three constants naming individuals $\{a_1, a_2, a_3\}$. In the first language L, there are only two Q-predicates $\{Q_1x, Q_2x\}$. In the second language L', there are four Q-predicates $\{Q_1x, \ldots, Q_4x\}$. Suppose that the first two individuals, a_1 and a_2, are observed to satisfy the predicate Q_1x for both languages, so the agent wishes calculate the probability of Q_1a_3 given

[151] Bacchus et al., "From Statistical Knowledge Bases," 126-7.

Q_1a_1 & Q_1a_2.[152] According to $m*$ with L's sample space, the probability of Q_1a_3 given Q_1a_1 & Q_1a_2 is 0.75. On the other hand, according to $m*$ with L''s sample space, the probability of Q_1a_3 given Q_1a_1 & Q_1a_2 is 0.5.[153] Hence, we can see how various probabilities are affected by the choice of language. The sample point holds for certain probabilities of proportion formulas.

This dependence on language is problematic only because the choice of language is problematic. In particular, the choice of language risks being arbitrary in the absence of well-principled reasons for selecting the language. Sometimes the choice of language might be objectively correct, such as a case in which the agent knows that they are truthfully told that only three specific colours could characterise the ball in the urn. But sometimes exactly one correct choice appears to be lacking, an example of which possibly being the case where the agent has no information that limits the possible colours that the ball could be.

Interestingly, however, many probabilities do not depend on the choice of language (at least for the random-worlds method). Indeed, Bacchus et al. prove many results regarding the random-worlds credences that are not dependent on language. One such result concerns basic direct inference. To use an illustrative example, suppose the agent has two languages about a domain of three objects. Suppose that the predicates A and B are known to constitute a partition of predicates for the first language (meaning that each object satisfies exactly one of the predicates in the partition) and that the predicates A, B and C are known to constitute a partition of predicates for the second language. (Perhaps we can suppose that A symbolises the same property in both languages but B and C symbolise two different properties in the second language, the disjunction of which is symbolised with just B in the first language.) Each language has three constants a, b and c which are known to uniquely and specifically denote the domain objects and the languages have no other symbols aside from the aforementioned predicates and constants. We then have the below sets of models which are represented with conjunctive sentences and which are consistent with a knowledge base specifying the denotation of the constants and the logical relations among the predicates. The first set of models corresponds to the first language and the second set of models corresponds to the second language. Suppose that the agent also knows that $\%(Ax) = \frac{2}{3}$. This effectively rules out, probabilistically speaking, the highlighted models in the following sets of models:

[152] Technically, Q_1 would involve at least one primitive predicate in L' that is not in L, but we can harmlessly suppose that, whatever Q_1 means and whatever primitive predicates it involves in the different languages, it applies to both a_1 and a_2.

[153] This result can be obtained using the rule of succession mentioned in Section 5.1.

Set of Models #1:

M_1: Aa & Ab & Ac

M_2: Aa & Ab & Bc

M_3: Aa & Bb & Ac

M_4: Aa & Bb & Bc

M_5: Ba & Ab & Ac

M_6: Ba & Ab & Bc

M_7: Ba & Bb & Ac

M_8: Ba & Bb & Bc

Set of Models #2:

M_1: Aa & Ab & Ac

M_2: Aa & Ab & Bc

M_3: Aa & Ab & Cc

M_4: Aa & Bb & Ac

M_5: Aa & Bb & Bc

M_6: Aa & Bb & Cc

M_7: Aa & Cb & Ac

M_8: Aa & Cb & Bc

M_9: Aa & Cb & Cc

M_{10}: Ba & Ab & Ac

M_{11}: Ba & Ab & Bc

M_{12}: Ba & Ab & Cc

M_{13}: Ba & Bb & Ac

M_{14}: Ba & Bb & Bc

M_{15}: Ba & Bb & Cc

M_{16}: Ba & Cb & Ac

M_{17}: Ba & Cb & Bc

M_{18}: Ba & Cb & Cc

M_{19}: Ca & Ab & Ac

M_{20}: Ca & Ab & Bc

M_{21}: Ca & Ab & Cc

M_{22}: Ca & Bb & Ac

M_{23}: Ca & Bb & Bc

M_{24}: Ca & Bb & Cc

M_{25}: Ca & Cb & Ac

M_{26}: Ca & Cb & Bc

M_{27}: Ca & Cb & Cc

The fraction of models for which Aa is true is $\frac{2}{3}$ in both sets of models. Hence, $P_{Random-Worlds}(Aa|KB) = \frac{2}{3}$ irrespective of which language is used (where the knowledge base contains $\%(Ax) = \frac{2}{3}$ and the other information noted above). Consequently, we can see how certain probabilities do not depend on the choice of language.

The measures, as tools for inverse inference, also confront Nelson Goodman's problem of induction.

The problem can be described as follows. Suppose we have a large sample of n emeralds and find that all of them are green. Also suppose that from this we want to infer that all emeralds are (probably) green. But note that we can construct a language which putatively justifies the inference that they are not all green. Define an object to be grue just in case it is green and first observed before 2050 or is blue and first observed after 2050. (Technically, we should say something like "observed before [or after] 12:00am, January 1, 2050" to accommodate emeralds observed in 2050, but we will just say "2050" for simplicity's sake.) We could represent this in logical notation with the biconditional $G'x \equiv ((Gx\&Tx) \vee (Bx\&{\sim}Tx))$ where $G'x$ stands for "x is grue," *Gx* stands for "x is green," *Tx* stands for "x is first observed before 2050," *Bx* stands for "x is blue" and it is true that $\forall x((Bx \vee Gx) \supset ({\sim}Bx \vee {\sim}Gx))$ (i.e. no blue thing is green and vice versa). All emeralds observed up to now are green and grue since it is prior to 2050. Therefore, we have equally good reasons to believe that all emeralds are green and that all emeralds are grue. But all emeralds are grue only if all emeralds first observed after 2050 are blue and not green. Therefore, we have we have equally good reasons to believe that all emeralds first observed after 2050 are green and not green, but surely this is an unintuitive conclusion. This problem can be generalised to potentially undermine any inference from the features of observed objects to the probable features of unobserved objects.[154] Goodman's problem is a problem for the measures since grueness and other convoluted properties could be treated as primitive predicates in place of greenness to thereby license unintuitive inferences with the help of the measures.[155]

---

[154] In support of this point, see Goodman's "emeroses" example in Goodman, *Fact, Fiction, and Forecast*, 74.

[155] To be fair, I should avoid giving the impression that Carnap was unaware of the problem and did not have a putative solution to it. See his discussion of Goodman's problem of induction and related issues in Carnap, "A Basic System of Inductive Logic, Part I," in *Studies in Inductive Logic and Probability,* eds. Rudolf Carnap and Richard Jeffrey, vol. 1 (Berkeley: University of California Press, 1971), 71-76. In his basic system of inductive logic, Carnap defined primitive predicates in a way that excludes the combination of modalities found in the grue predicate. Modalities, according to Carnap, are *types of properties*. Examples of modalities include shape, colour and height. Grue, according to Carnap, is not a primitive predicate since it denotes a property which combines two modalities (colour and time), one of which is qualitative while the other is temporal and locational (i.e. it locates an object in an absolute position in time or space, namely, pre- or post-2050). Carnap

Hence, we have seen that certain probabilities for the measures are dependent on the choice of language to represent the possible outcomes, a choice that can be problematic in various ways.

6.3.4. The Sampling of Species Problem

The problem of language relativity is related to another problem known in statistics as the *sampling of species problem*.

Although I have not found a clear characterisation of the sampling of species problem in the literature, it is clear that it arises from the agent's uncertainty regarding the number of possible species or categories of object in the population which she is sampling from.[156] Zabell uses the example of estimating the number of zoological species in an area. He states that "if the region is found rich in the variety of species present, the chance of seeing a particular species again may be judged small, while if there are only a few present, the chances of another sighting will be judged quite high."[157] Presumably, then, the problem is that of how to assign probabilities to seeing both previously observed and unobserved categories of object, particularly given the variety of categories present in one's sample.

We can see how this creates a problem for the measures. Let us adapt Zabell's zoological example and suppose that we are trying to determine the proportion of animals that are of a certain species in an area. Recall that the alternative measures are over a sample space that is generated by the Q-predicates in a language, predicates which are in turn formed from primitive predicates. Presumably there should be a primitive predicate for each species in the population – for example, A for ape, B for bull, C for cat and so forth. But if the number and types of species in the population are unknown, then it is unclear what the primitive predicates should be and consequently what the appropriate sample space is. Therefore, it is unclear how to apply the measures.

One potential approach to the problem is to have a sample space constructed from a language with predicates symbolising only the species that are known to exist in the

required the primitive predicates of inductive logic to be of one modality that is not locational, thus disallowing grue predicates. This categorisation of predicates serves as the basis for his disqualification of grue-favouring inferences. Whether his solution succeeds or not is another question that I shall not consider here.

[156] Zabell discusses the problem in "Carnap and the Logic of Inductive Inference," 294, "Confirming Universal Generalizations," *Erkenntnis* 45, no. 2-3 (1996): 280 and "Predicting the Unpredictable," 206.

[157] Zabell." Predicting the Unpredictable," 206.

population at a given time. For example, if one's sample consists only of apes and bulls and they therefore know that these two species exist in the population, then they should have a language and sample space involving just two primitive predicates for the species. But this is problematic because every time a new species is sampled, one needs to reconstruct the language, the sample space and the resulting measure. Furthermore, if the sample space only involves predicates known to be satisfied by a previously sampled animal, then there is no space for the probability that there are other non-sampled species in the population (such as cats or dogs in our somewhat bizarre example).

In response to these problems, one might modify this approach and denote unobserved species with a special predicate such as S to represent a "something else aside from what is in the sample" category, so to speak. But if every sampled object satisfies a predicate aside from S (since the sampled objects have their own predicates that are distinct from S), then there are no objects in the sample supporting the probability that the next sampled object will be something else aside from the observed species since none of sampled objects belong to that category. But surely this probability should be relatively high given a sample with diverse species that suggests that other unobserved species exist in an even more diverse population. If have a sample of 100 animals which each come from a different species, presumably this would make it probable that I would come across something else aside from these species in the future. But since none of the objects satisfy the formula Sx, the probability of observing something else is extremely low on this approach to the problem.[158]

Arguably, then, the problem involves the problematic choice of language, yet it is primarily one about how to assign probabilities to the possibility that unsampled categories of object exist in a population given the observed variation in the categories within a sample.

This problem arises in other contexts where one is trying to estimate the probabilities of interest among an unknown number of possible types of outcome.

Regardless, Zabell defends a putative solution to this problem, and Hintikka and Niiniluoto's later systems also attempted to address this problem too (or something like it). These authors specify a probability function which takes as arguments not only the observed frequency with which objects satisfy any given target formula in a sample, but also the

[158] In fact, if k in the rule of succession seen in Section 5.1. represents the categories of object that may possibly appear in the population (i.e. the 100 observed species and the "something else" category) then the probability of observing something else with the next observation is approximately 0.005 according to the rule.

number of properties or categories of object observed in the sample.[159] However, neither of these attempts to address this problem specifically concern inverse inference and it is not clear that they could be extended to do so.

6.3.5. The Problem of Similarity

In my opinion, though, the most important problem for these alternative measures stems from their putative inability to account for the evidential implications of similarity.

A general theory of inverse inference would ideally be able to account for two kinds of evidential implications arising from similarity.

One kind is the influence that the similarity between a sample and a population has for determining the probability of proportions regarding one target property for the population. For example, suppose a sample of pigeons is found to all have a newly discovered gene that makes pigeons susceptible to a newly discovered disease. Suppose there is no other evidence about the prevalence of this gene or disease among pigeons and birds in general. Surely the sample should provide some evidence as to the relative frequency of the gene among both the population of pigeons and the population of birds more generally. But surely the sample is a greater indication as to the relative frequency among the pigeons than it is among birds more generally.

The problem is exacerbated when many populations are relevant. To give another example, suppose I have sampled 50 Auckland Transport (AT) bus arrivals at my local bus stop and found that all of them were punctual (given some definition of punctuality in terms of, for example, a bus arriving within 4 minutes of its scheduled arrival).[160] Surely this should provide some evidence (of whatever strength) that the population of AT busses that stop at my bus stop are generally punctual. But it should also provide some (weaker) evidence that the population of AT busses stopping at *any stop* in Auckland are generally punctual, that the population of busses operated *by any company* in Auckland are generally punctual or that the

[159] Zabell, "Predicting the Unpredictable," and Jaakko Hintikka and Ilkka Niiniluoto, "An Axiomatic Foundation of Inductive Generalization," in *Formal Methods in the Methodology of Empirical Sciences*, eds. Marian Przelecki, K. Szaniawski and Ryszard Wójcicki (Dordrecht: Reidel, 1976).

[160] To illustrate the problem of similarity, I am adapting an example that Brendon Brewer uses when teaching Bayesian statistics (albeit not in the context of similarity). Brendon J. Brewer, "Introduction to Bayesian Statistics" (Course notes, STATS 331, The University of Auckland, 2014). (Accessible at https://www.stat.auckland.ac.nz/~brewer/stats331.pdf.)

population of busses operated by *any company throughout New Zealand* are generally punctual.

This illustrates that samples can be from many populations and the extent to which a sample is reflective of a population's relative frequency depends largely on the known similarity between the sample and the relevant population.

In these cases, the kind of evidential implication concerns the extent to which the *similarity between various individuals or groups* suggest that they share in common *one* property. The samples' similarity to the different populations has probabilistic implications about the extent to which the various populations likewise have the property of the gene or the property of punctuality.

The second kind of implication of similarity concerns the extent to which one group or individual having one property suggests that another group or individual will have *a distinct but similar* property. For example, suppose that a sample of adult pigeons are all found to weigh between 250 to 350 grams. Not only should this raise the probability that other members of the population of adult pigeons would weigh between 250 to 350 grams, but it should also raise the probability that other members in the population possess similar but distinct properties, such as weighing 355 grams.

Here we will focus on the problems raised by the first kind of implication from similarity.

In the literature on analogical reasoning, there are analogues of the problems that arise in accounting for these evidential implications in the context of inverse inference. Hence, we might understand these problems for inverse inference analogously.

We might say that one set of problems are *classificational* – they concern how we classify similarity or dissimilarity. In this respect, we face what is called the *counting problem* in the literature on analogical reasoning.[161] How do we count the number of similarities and dissimilarities between a sample and a population? Does, for instance, the fact that both pigeons and other birds have five-digit feet mean that that there is one similarity between them, or does that mean that there are five similarities corresponding to each digit?

Sometimes there may be an objective or natural metric by which similarity is to be measured. For example, one might think that a theory of genetics would provide such a metric for determining the similarity between pigeons and other birds. Whether there is such a theory is a question for geneticists, although I suspect it may be difficult, if not impossible, to objectively measure the similarity of pigeons to a population as genetically diverse as all

[161] Paul Bartha, "Analogy and Analogical Reasoning," *The Stanford Encyclopedia of Philosophy*, accessed May 5, 2016, http://plato.stanford.edu/archives/fall2013/entries/reasoning-analogy/.

birds in general. This difficulty is exacerbated since presumably one would need to make an inverse inference about the genetic makeup of birds in general to estimate the similarity between pigeons and birds in general on genetic grounds. Yet this inverse inference may itself require judgments of similarity on non-genetic grounds.

Of course, even there is such an objective metric, there is still the question of why one chooses to use that metric rather than another metric (one which may be equally natural or objective). Carnap alludes to this problem when noting that similarity among colours can be measured by one's perceptual sensations of colour or by frequencies of light waves.[162] He mentions that, given a set of axioms, the two corresponding ways of representing the space of possible colours can lead to certain probabilistic assignments that are incompatible. Perhaps one might regard the frequency of light waves as providing a natural metric of similarity, but there is still the question of why this is so or why it should be the preferred metric in place of possible perceptual sensations of colour. So sometimes there are multiple relevant and incommensurable metrics along which similarity can be measured.

In any case, sometimes it is not obvious that there is even at least one objective metric for measuring similarity. Suppose we had no other information about the genetics of pigeons and birds more generally. Surely the observable but non-genetically detectable similarity between the sample of pigeons and the different populations should still affect our credences regarding the relevant relative frequencies of the gene. But these observable similarities do not necessarily provide an objective and clear cut metric by which similarity is to be measured, such as a metric specifying that one foot with five digits counts as exactly one similarity rather than five or more. There may also be no such metric in many other cases, such as those concerning the similarity between my local AT buses and the many other populations of buses.

Yet another classificational problem is what we might call the *weighting problem*. Some similarities are accepted as being important, such as any similarities between objects that have a *causal relation* to the potentially shared property of interest. But, as Paul Bartha states, "[s]ome similarities and differences are known to be (or accepted as being) utterly irrelevant and should have no influence whatsoever on our probability judgments."[163] Both the sample of pigeons and the population of birds more generally are mentioned in this sentence, yet this similarity between the groups is irrelevant to the probability that the birds generally share the

[162] Ultimately, however, Carnap constructed his Basic System for an observational language, one which would not measure similarity in this context by the frequency of light. Carnap, "A Basic System of Inductive Logic, Part I," 51-52.
[163] Bartha, "Analogy."

gene in common with pigeons. So some similarities or dissimilarities should be weighted more or less heavily than others, or even completely disregarded, but producing general guidelines for such weighting is a complicated affair.[164]

Having counted and weighted similarities and dissimilarities, however, a related problem arises which we may say is *implicational*, that is, it concerns how to formalise the evidential implications of the similarities and dissimilarities. In this sense, an ideal theory of inverse inference would specify precisely how the similarity between a sample and population affects the probability of proportion formulas.

The aforementioned measures do not offer any obvious solution to this problem. For example, given that the case of the pigeons concerns degrees of similarity between birds generally and pigeons, a measure might offer a solution to the problem by somehow incorporating a specified degree of similarity between the predicates Bx which stands for "x is a bird" and Px which stands for "x is a pigeon". Furthermore, Bx and Px should have a higher degree of similarity to each other than either of them do to certain other predicates that may appear in the language (such as, say, Wx which stands for "x is wombat"). But the measures do not account for such similarity relations. They are simply measures on sample spaces constructed from predicates and constants denoting individuals – all such predicates and individuals are treated the same in the sense that there is no consideration given to the degree of similarity between them. Given that the sample of pigeons should provide some evidence about the prevalence of the gene in both the population of pigeons and population of birds in general, it is doubtful that the measures could discriminate between the populations so as to account for the evidential implications of the varying degrees of similarity.[165]

One might think of other ways that the measures may potentially address these problems, yet none of the ones that I am aware of are promising.

In any case, it is not controversial that these measures cannot adequately account for the problems posed by similarity since it is precisely for this reason that Carnap and Hintikka later abandoned the measures for more complicated systems of inductive inference. Carnap's last two-part work, "A Basic System for Inductive Logic", presents an axiomatic system which generates predictive probabilities (i.e. the probabilities that are the outputs of predictive inference; see Section 2.2.). This system features an "analogy parameter" η, that is, a parameter which specifies to what extent the observation of $Q_l a_1$ should influence the

[164] Bartha, "Analogy."
[165] Here, I am presupposing an understanding of a "population" in terms of the definition given in Section 2.2., not Carnap's definition in Section 5.1.

expectation that the next observed individual will satisfy $Q_j x$ when the agent has no other relevant evidence (where Q_l and Q_j are Q-predicates such that $l \neq j$).[166] This η-parameter in turn gives rise to Carnap's λ-parameter, often interpreted as an indicator of the weight given to empirical and non-empirical factors when estimating predictive probabilities.[167] Hintikka and Ilkka Niiniluoto provided a generalisation of Carnap's system, one which was alluded to earlier and which allows the predictive probabilities to also depend on another parameter representing the number of different properties in a sample.[168] Yet these systems are problematic since it is doubtful that the parameters always have objectively correct values which determine a uniquely correct probability function. Carnap was unable to constrain η to generally generate one uniquely correct probability function and Hintikka later, in his own words, "lost interest" in inductive logic in part because the optimal choice of the λ-parameter and others cannot be determined on a priori or "purely logical basis".[169] In any case, these systems do not specifically concern inverse inference, so they are of limited relevance here.

Similarity also poses a problem for the theories of inverse inference proposed by Pollock, Kyburg and Teng.[170] Their theories for inverse inference only consider the proportion of objects satisfying a target formula in a sample of size *n*. To assign probabilities to population relative frequencies, then, one simply counts the objects in a sample with the target property and the ones without it. There is no consideration given to the degree of similarity between the sample and the population.[171]

---

[166] According to Niiniluoto, Carnap characterises η as an "analogy parameter" in Rudolph Carnap and W. Stegmüller, *Inductive Logik und Wahrscheinlichkeit* (Vienna: Springer-Verlag, 1959). Niiniluoto, "The Development," 341. Marta Sznajder also characterises it in this way in Marta Sznajder, "What Conceptual Spaces can do for Carnap's Late Inductive Logic," *Studies in History and Philosophy of Science* 56 (2016):66.

[167] See, for example, Patrick Maher, "Explication of Inductive Probability," *Journal of Philosophical Logic* 39, no. 6 (2010): 601-2.

[168] Hintikka and Niiniluoto, "An Axiomatic Foundation of Inductive Generalization."

[169] Jaakko Hintikka, "Reply to Isaac Levi," in *Philosophy Jaakko Hintikka*, eds. Randall E. Auxier and Lewis Edwin Hahn (Chicago: Open Court, c2006), 777-778, Niiniluoto, "The Development," 349 and Risto Hilpinen, "Carnap's New System of Inductive Logic," *Synthese* 25, no. 3 (1973): 313. Nevertheless, Carnap hoped to eventually find further constraints on probabilities and he did spend 35 pages discussing different values of the λ-parameter before concluding that he was "inclined" to set it to 1 for many cases. Carnap, "A Basic System of Inductive Logic, Part II," 119.

[170] See a summary of Pollock's take on inverse inference in Pollock, "The Theory of Nomic Probability," 285-296. His extended discussion of inverse inference can be found in *Nomic Probability and the Foundations of Induction*. Kyburg and Teng's account of inverse inference can be found in Kyburg and Teng, *Uncertain Inference,* ch. 11.

[171] While they have some requirements for inverse inference from samples, these do not suffice to address the problem of similarity. Pollock merely requires that the sample be "innocent until proven guilty" while Kyburg and Teng require there to be "no grounds for impugning the fairness of the sample". See Pollock, *Nomic Probability,* 114 and Kyburg and Teng, *Uncertain Inference,* 270. One might think that the sample of pigeons is fair and innocent when considering its implications for the prevalence of the gene in both the populations of pigeons and birds more generally. One might think this on the grounds that there is no reason to deny that the prevalence of the gene in the sample diverges from the prevalence in either of the populations. Based on their

So the problem of similarity is one which casts doubt on the utility of the aforementioned measures and theories, at least in many contexts.

Yet it may also cast doubt on the possibility of finding what we might call a *formal and objective method of inverse inference*. A formal and objective method of inverse inference is a means for inverse inference which takes several quantifiable and well-defined characteristics of a sample and returns uniquely correct probabilities for proportion formulas, perhaps relative to a language and population size. Carnap's $m*$, Hintikka's measures, Pollock's theory of statistical induction and Kyburg and Teng's theory of evidential probability all offer such methods. For example, all one needs to know to use these methods of inverse inference is the proportion of objects in a sample of size n satisfying a formula and, for Carnap's $m*$ and Hintikka's measures, the size of the population of individuals and features of a language. These methods, then, purport to take objective quantitative inputs (perhaps given a language) and return objective quantitative outputs.

While these methods may be useful in some contexts of inverse inference, they are not useful in all contexts given the above problems.

The problem of similarity, then, is a formidable, if not insurmountable, obstacle in the path to finding "the one true formal method of inverse inference" so to speak. For one, we might have qualms about the possibility of there being such a method since specifying the relevant degree of similarity involves subjective or intuitive judgements and, hence, any method which takes these judgments as inputs may not return *the* objectively correct output. For example, it is doubtful that there is some objective fact that, say, each digit shared by pigeons and other birds counts exactly as one similarity of weight f and that the degree of similarity between pigeons and other birds is precisely 93% rather than 85% (whatever that means). Furthermore, even if there is some specifiable degree of similarity, it is not obvious that there is a uniquely correct way of formalising its probabilistic implications. We may doubt, for example, the possibility that there will be a formal method which correctly specifies that each digit shared by both pigeons and other birds raises the probability of some proportion formula of interest by exactly 0.0134 rather than 0.05.

(lack of) restrictions on permissible samples to make inferences from, I think that Kyburg, Teng and Pollock would probably have regarded the sample of pigeons as fair for estimating both populations; yet this would counter-intuitively generate the same probabilities regarding possible relative frequencies of the gene in both populations. But even if they were to regard the sample of pigeons as partly guilty or unfair when estimating the prevalence of the gene among the population of pigeons or birds, it should still provide *some evidence* about the prevalence, evidence whose import for the different populations varies. Yet the theories of Pollock, Kyburg and Teng are silent on how this is so.

A pessimism toward formalising all inverse inferences as such is consonant with some of the literature on analogical reasoning. There are various accounts aiming to shed light on the evaluation of analogical argument, accounts which are explicitly normative or from which one may attempt to draw norms.[172] Most recent accounts of analogical inference do not propose formal rules of inference for obtaining probabilities on the basis of analogical considerations.[173] Rather, they only provide general criteria and procedures for analogical inference. Bartha, in fact, is of the opinion that "[d]espite the confidence with which particular analogical arguments are advanced, nobody has ever formulated an acceptable rule, or set of rules, for valid analogical inferences."[174] John Norton provides a similar perspective on analogical inference. He disparages attempts to formally distinguish good analogical inferences in general from bad ones, asserting that the legitimacy of each inference depends on local facts and intuitive judgments specific to the particular analogy at hand.[175] He notes that, in the past, formal analyses of analogical inference have always confronted some problem case and that "there always seems to be some part of the analysis that must be handled intuitively without guidance from strict formal rules."[176] This, in fact, is an expression of his *material theory of induction*, the cornerstone of which is that "there are no universal rules of inductive inference" (perhaps aside from any ones in this sentence) but rather that "all induction is local."[177]

We might see inverse inferences from non-probability samples as being similarly resistant to objective and formal analysis. In fact, arguably such inverse inferences are just large-scale analogical inferences inferring the probability that the relative frequency in one set of objects is similar to that of another set of objects based on the similarity between the sets. If this is the case, then perhaps the rationality of such inverse inferences must be assessed like analogical inferences, taking into account facts local to the situation at hand and the agent's intuitive judgments of the similarity and its implications. This is a take on inverse inference which accords more with common inferential practices. Real-world agents often (if not always) cannot articulate formal principles specifying the degree of similarity between, say, my local buses and the various populations that they are members of. But they can make intuitive judgments about such similarity and its implications. Furthermore, these judgments

---

[172] Bartha, "Analogy."
[173] Ibid.
[174] Ibid.
[175] John Norton, "Analogy," *University of Pittsburgh*, accessed May 7, 2016, http://www.pitt.edu/~jdnorton/papers/Analogy.pdf.
[176] Ibid.
[177] John D. Norton, "There are No Universal Rules for Induction," *Philosophy of Science* 77, no. 5 (2010): 765.

are useful for making inferences. I would think that agents often successfully infer that most of their local buses arrive within a certain period of their scheduled time and that the next sampled bus is probably punctual. I would also think that the typical agent would have many other success stories of intuitive inferences too.

In any case, let us summarise the implications from the problem posed by similarity and the other topics in the next section.

6.4. Summary: A General and Formal Method for Inverse Inference?

This chapter examined one response to the problem of inverse inference that confronted the random-worlds method. The response was to appeal to alternative but similar measures which could allow useful inferences from samples. As we saw, however, various other problems confronted these measures. These were:

- the putative illegitimacy of their vindication of inferences involving non-probability samples
- the problem of being frequently inapplicable by virtue of the complicating background evidence which varies from context to context
- the dependence of certain probabilities on the problematic choice of language
- the difficulties in encapsulating the agent's uncertainty about the possible categories of object in the population, particularly given variation of categories in a sample and
- the difficulty in accounting for the problem (or problems) posed by similarity.

Some of these problems seem more severe to me than others. Indeed, some seem potentially unsolvable, at least without appeal to potentially intuitive and informal judgments (such as the problem of similarity).

What, then, are the implications of these problems?

In my opinion, we probably lack a generally applicable, formal and objective method for inverse inference, at least for now. By this, I mean that there is no one correct method that instructs the agent exactly how the probability of various proportion formulas should be affected by her sample of pigeons sharing a particular gene, her sample of days she has lived since her birth, her sample of black toy cars from a factory, her sample of coloured balls in an urn, her sample of objects in the universe obeying theorised laws (to some degree of approximation), etc. There is currently no formal and objective method of inverse inference which provides a general and uniquely correct way of accommodating the evidence that varies from context to context, including the similarity between the sample and the population, the variety of categories of object in a sample and the evidence about other relevant samples and populations.

Furthermore, I am pessimistic that such a method will be found. This is especially since inverse inference from non-probability samples is, in my opinion, basically a large-scale analogical inference inevitably involving intuitive judgments about the similarity between a sample and one or another of the populations that it is from. Naturally then, John Norton and subjective Bayesians would probably share my pessimism too. (However, I doubt that Pollock, Kyburg and Teng would given the formal methods which they espouse.)

Note that my opinion and pessimism is construed probabilistically. This is because one cannot review every piece of extant literature that is possibly relevant and maybe there is or will be such an objective, formal and universally applicable method after all.[178] Regardless, given what I have seen of the main philosophical and statistical methods of inverse inference (including classical confidence intervals and others), they are not such generally applicable and formal methods. While they seem useful and insightful, they are unable to fully accommodate the issues posed by similarity and other problems. Furthermore, just by reflecting on the issues and on how their nature intuitively seems to preclude one universal, objective and formal analysis, I suspect that there is no such analysis.

This is not to say, however, that there are currently no formal and objective methods which account for a large category of inverse inferences *in certain circumstances*. For example, it seems quite reasonable that some measures may be generally applicable for certain types of inferences from probability samples, such as random samples.

Unfortunately, though, the overwhelming majority of contexts in which humans engage in inverse and predictive inference involve non-probability sampling. Surely we would like a theory of inference for these contexts, one partly for the layman who wishes to make inferences yet lacks the money, time and expertise to do so on the basis of sophisticated probability sampling strategies.

How, then, should we constrain our credences on the basis of evidence about frequencies in non-probability samples? What are the right inferences from such samples?

In the next chapter, I provide a response to these questions.

[178] Another relevant body of literature concerns machine learning and Kolmogorov complexity. Yet, from what I have seen, this literature does not discuss similarity in the context of inverse inference (as opposed to just universal or predictive inference), even though it may contain some relevant insights. See Ming Li and Paul Vitányi, *An Introduction to Kolmogorov Complexity and its Applications* (New York: Springer, c2008).

7. Intuition, Evidence and Inference

"It has been recognized by all, beginning with Laplace, that the purpose of a statistical analysis is to aid our common sense by giving a quantitative measure to what we feel intuitively."

E.T. Jaynes

"[S]ome subjective element exists in all scientific appraisal…"

Colin Howson and Peter Urbach

"[T]he true test of our semantics requires seeing whether it yields intuitively plausible results in simple cases, to begin with, and reasonable results in more complex cases in which our intuitions are not so strong."

Henry Kyburg and Choh Man Teng

The above authors all reflect very different strands of thought about statistical reasoning. Jaynes is a pioneering objective Bayesian. Howson and Urbach are champions of subjective Bayesianism. Kyburg and Teng are more accurately placed in the classical frequentist camp, if anywhere.[179]

Regardless, they are all united by at least two things. First, they share a common aim of finding the right theory for inductive inference. Second, they allow, or even emphasise, a role for intuition in pursuing this aim. Jaynes rails against classical statistical methods and favours Bayesian inference. He does this by adducing cases in which classical methods putatively deliver results that are inconsistent with the intuition enshrined in common sense.[180] Howson and Urbach place no constraints on the permissible prior probabilities in Bayesian inductive inferences, thus allowing space for intuitive priors to make an appearance.[181] Kyburg and Teng test their theory of evidential probability against intuitive desiderata for a theory of

[179] Kyburg and Teng, *Uncertain Inference*, 195. This is also apparent from their advocacy of (quasi-)classical methods for inferences from samples and a non-Bayesian theory of acceptance. See Kyburg and Teng, *Uncertain Inference*, 261 and Chapter 11.

[180] E.T. Jaynes and Oscar Kempthorne, "Confidence Intervals vs. Bayesian Intervals," in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, vol. II., eds. William L. Harper and C.A. Hooker (Dordrecht: Springer Netherlands, 1976), 175-257.

[181] Howson and Urbach, *Bayesian Reasoning*, 265.

inductive inference, such as allowing basic direct inference, a preference for statistics for more specific reference classes and the like.[182] Bacchus et al. similarly evaluate their theory against intuitive desiderata of this sort.[183]

In this sense, there is a precedent for evaluating theories of inference based on pre-theoretic intuitions specifying what credences or probabilities are epistemically rational in certain cases. Recall, for example, the intuition underlying basic direct inference: the (perhaps subjective) probability that Aaron was fatally injured should be 0.8 if the agent knows that Aaron was on a train where 80% of the passengers were fatally injured (and that is the only relevant knowledge regarding Aaron possibly being injured). Bacchus et al., Kyburg and Thorn all share this intuition, but they do not all agree on a theory that explains, formalises or justifies it.[184]

An analogous situation is present in other areas of inquiry. Many philosophical views are motivated by intuition about particular thought experiments, including the view that knowledge is not merely justified true belief (as per the Gettier cases), that computers lack semantic understanding (as per the Chinese room) and that mental states are at least partly non-physical (as per Mary the colour scientist).

In Chapter 6, I attempted to motivate the claim that there is currently no general, objective and formal method for inverse inference which succeeds to account for our intuitions about various problems.[185] There, I said that such methods purport to take several objective, quantifiable and well-defined characteristics of a sample (perhaps given a language) and return objective quantitative outputs in the form of uniquely correct probabilities.[186] I will henceforth call such methods merely "formal methods" for short. This is not to say, however, that there are no contexts in which extant formal methods do deliver the right answers, but only that there are some contexts - particularly with non-probability samples - where formal methods do not and perhaps never will.

[182] Kyburg and Teng, *Uncertain Inference*, 244.
[183] Bacchus et al., "From Statistical Knowledge Bases," 79.
[184] For instance, Bacchus et al. explain this intuition via an indifferent probability measure over the models that are consistent with a knowledge base; in contrast, Kyburg, Teng and Thorn do not. See Kyburg and Teng, *Uncertain Inference*, 226 and 262, Thorn, "Two Problems of Direct Inference," 307 and Thorn, "Three Problems of Direct Inference," 70.
[185] As mentioned in Section 6.1., even if the reader is not sympathetic with the claim, they can perhaps appreciate this chapter and the next as being a conditional claim of the sort "If my pessimism is justified, then here is what one can do."
[186] It should be clear, then, that I do not take subjective Bayesian methods of data analysis to be such methods since they explicitly involve (relatively) unconstrained, subjective distributions as inputs, not just things such as the objective number of objects in a sample satisfying some target formula.

Despite this situation, we need to make inverse inferences in both scientific and everyday contexts. Surely my having consistently observed punctual busses at my bus stop should cause me to adjust my credences in the proportion of buses that are punctual, making me more confident that at least most buses at that stop are punctual. Likewise, having sampled several guests at a party who all know Ben should adjust my credences about the proportion of people at the party who all know Ben. Similarly, a sample of 1,000 emeralds, all of which are green, should adjust my credences regarding the proportion of emeralds that are green. I could produce example after example in which samples, including non-probability samples, should affect my credences for population relative frequencies.

What, then, are the right inverse inferences in the absence of a general formal method?

One suggestion is to follow the aforementioned tradition and seek answers from intuitions similar to the basic direct inference one, although perhaps via an unexpected route.

This chapter will outline a theory which emphasises a role for intuition in inverse inference, particularly in the context of inferences from non-probability samples. The theory involves two dimensions along which the epistemic rightness or rationality of an inverse inference is assessable. The chapter then critically discusses the claim that intuition should be trusted when making inverse inferences, concluding that certain intuitions are trustworthy. Following this, I defend the theory as one which has content for the individual when applied to their life. I lastly summarise the theory and delineate problems and avenues for further research. In Chapter 8, I will then explore how the theory can supplement the random-worlds method.

However, as a preliminary, it would be useful to consider the nature of the relevant intuitions given their importance in this chapter. I will understand these intuitions as beliefs that prescribe inferences to credences, perhaps on certain conditions. For example, consider again the aforementioned intuition about basic direct inference. This can be understood as the belief that one's credence that Aaron was fatally injured should be 0.8 if one knows that Aaron was on a train where 80% of the passengers were fatally injured (and this is the only relevant knowledge regarding Aaron possibly being injured). Many have this intuition about what their credence should be given certain conditions regarding a knowledge base. Furthermore, they might have this intuition, as I do, without have actually having the credence under discussion because the relevant conditions do not hold for them. Hence, I understand intuitions about what credences are appropriate as being distinct from those credences. I also understand these intuitions as not having any articulable justification, at least initially.[187] For example, one may have the basic direct inference intuition without being able to explain why that intuition is trustworthy. In this sense, we can say that the intuitions under discussion are (initially) unjustified beliefs about epistemically rational inferences (perhaps of a conditional sort stating that "if one has evidence *e*, then they should have a credence of strength *f* for proposition *p*"). I will also suppose in this discussion that the intuitions of any agent are coherent and I will call the agent's inferences that accord with such intuitions "intuitive" inferences.

I am assuming this understanding of intuition to be adequate, even though one might challenge it one various grounds. For one, an agent's intuitions might not be coherent; one might simultaneously have intuitions that they should and should not believe proposition *p* to degree *f* given their circumstances. For another, the line between credences and intuitions is blurry. When I am in a situation like the direct inference one above where I do have such knowledge about Aaron, I may not have both a credence of 0.8 *and* an intuition about that credence – rather I might just have the credence which is in some sense intuitive. Furthermore, intuitions may themselves be graded. I might only be 90% confident that one's credence that Aaron was fatally injured should be 0.8 if one knows that Aaron was on a train where 80% of the passengers were fatally injured. But perhaps I am also 10% agnostic, so to speak, about what that same credence should be in those circumstances (perhaps because of Levi's objection to direct inferences with non-random samples). Here, then, my intuitions are characterised by a degree of confidence and are therefore perhaps best understood as credences.

On this note, we can see how complicated the nature of credences and intuitions may be. To avoid writing an unduly long thesis extending into empirical psychology, I will just suppose that this (perhaps overly) simplistic understanding of intuition is correct. However, I explicitly acknowledge that there are legitimate questions about this understanding which are relevant and worth pursuing, but which cannot be pursued here. I also acknowledge that the answers to these questions may have implications for what I say in the following. But to have space to at least say something, let us bracket these questions and move on.

[187] I use the term "articulable justification" to denote a justification for the intuition which the agent can speak of. This is in contrast to an externalist notion of justification whereby an agent may have a justification for an intuition because it is, say, produced by a reliable process, even if the agent cannot articulate this latter justification and may not even know that they have it.

7.1. Objectively Rational Inverse Inferences

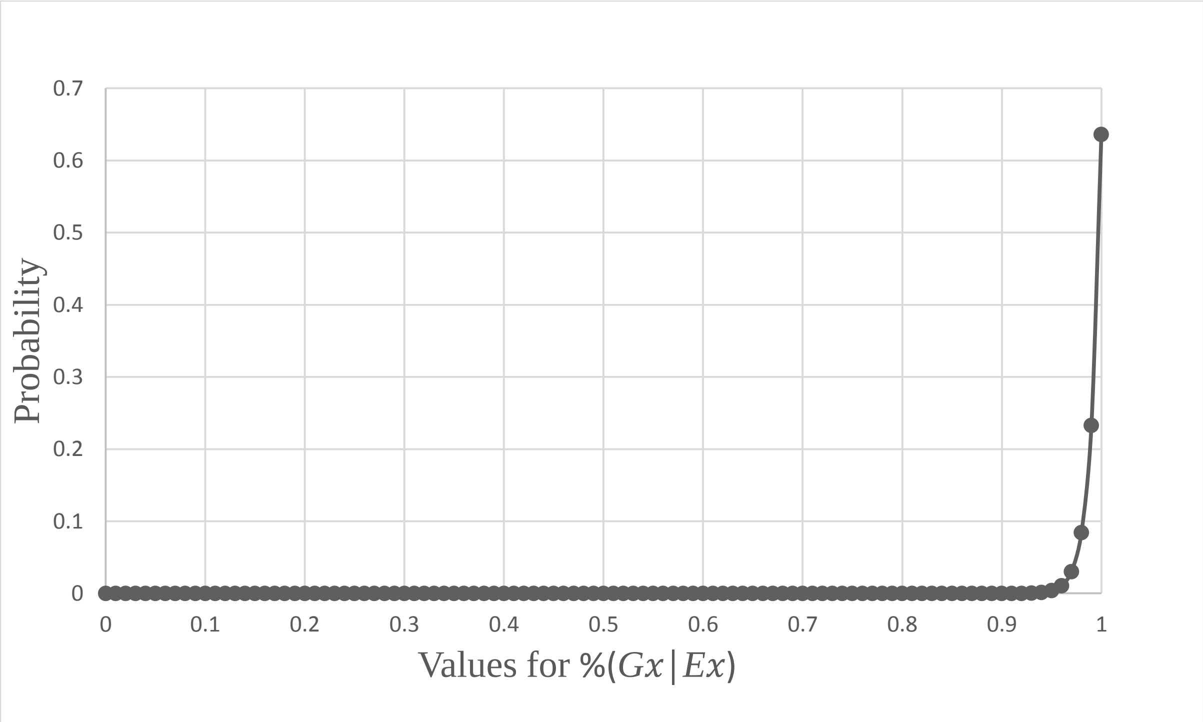
In making inferences, some intuitions are less trustworthy than others. We now turn to a thought experiment that capitalises on this idea in the context of inverse inference and that will be useful as a point of reference later.

Suppose that two agents, sensible Sarah and radical Mitchell, are interested in estimating the proportion of emeralds that are green in a particular area. Suppose that they have exactly the same evidence (or lack thereof) bearing on the colour of emeralds and the only things that differ are their intuitions about the probabilistic implications of that evidence (or lack thereof). (This supposition may be unrealistic, but this does not mean that it is no less useful or conceptually possible than many other bizarre philosophical thought experiments involving teletransporters, time travel, Chinese rooms and the like.) We can suppose they have no knowledge about emeralds that bears on the probable proportion(s) of emeralds (in the area or in general) that are of any particular colour.[188] Here we represent the space of possible proportions with formulas of the form $\%(Gx|Ex) = f$ where where Ex stands for "x is an emerald in the area," Gx stands for "x is green" and f is some real number in $[0,1]$.[189] Suppose that they both sample 10,000 emeralds by haphazardly inspecting rocks around the area together. All of the emeralds are green and so they have the following intuitive inverse inferences based on the sample:

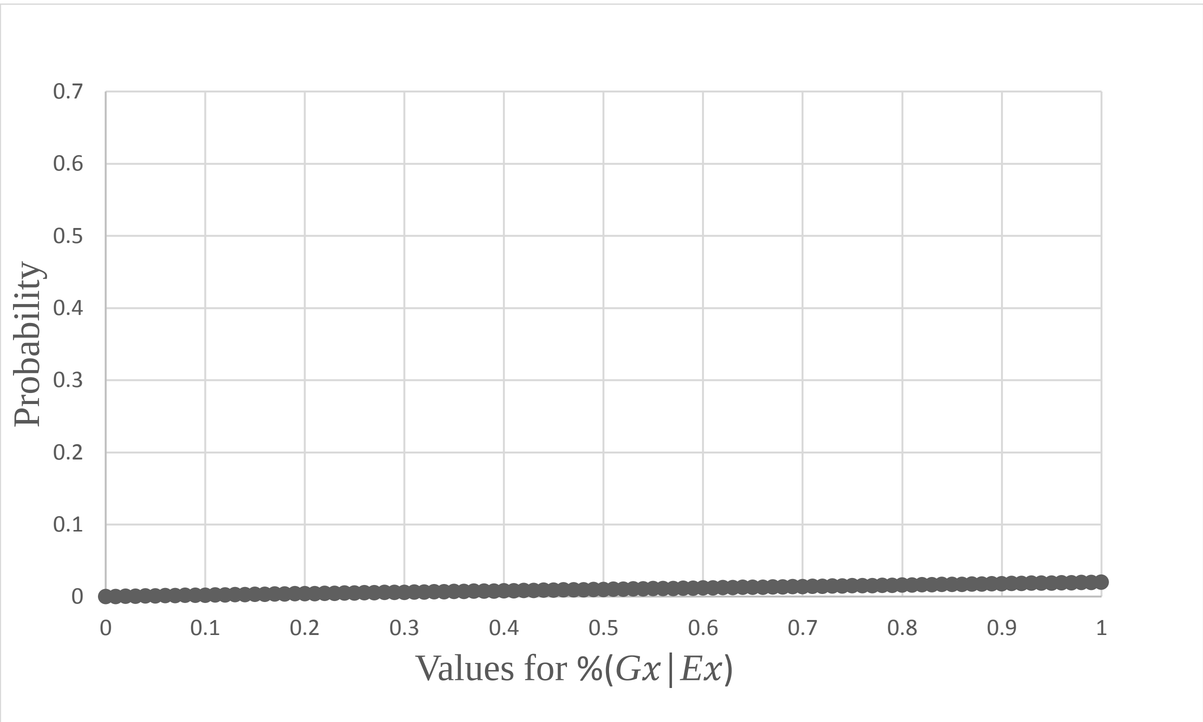
[188] If need be, we can also suppose that they (somehow) know that the population of emeralds in the area is 1,000,000 or some other number.

[189] In doing so, we will use an idealisation of the space of possible proportions whereby the space consists of 101 values in [0,1] that are separated by equal distances. The use of this idealisation should not be too objectionable since it is a commonplace and (frequently) harmless practice in statistics to represent the "parameter space" of possible proportions with idealised continuous or discrete spaces. See, for example, the use of idealised continuous spaces alluded to in Howson and Urbach, *Scientific Reasoning,* 32.

Sarah’s distribution after sampling 10,000 green emeralds



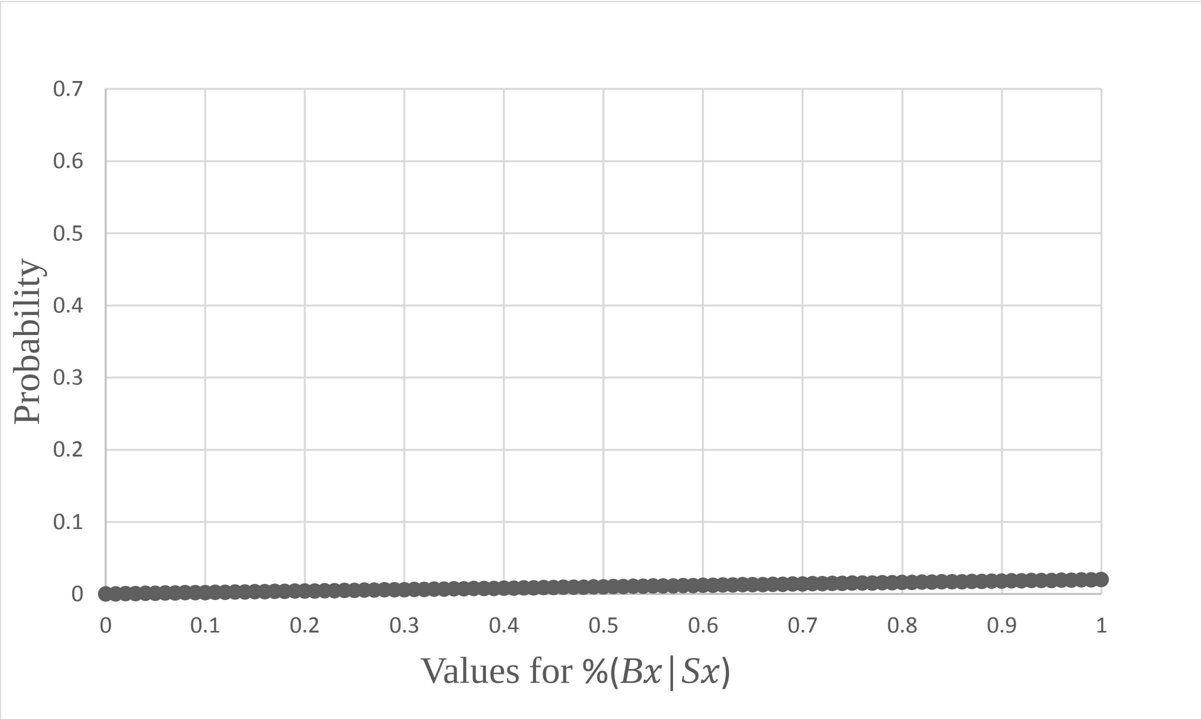
Mitchell’s distribution after sampling 10,000 green emeralds



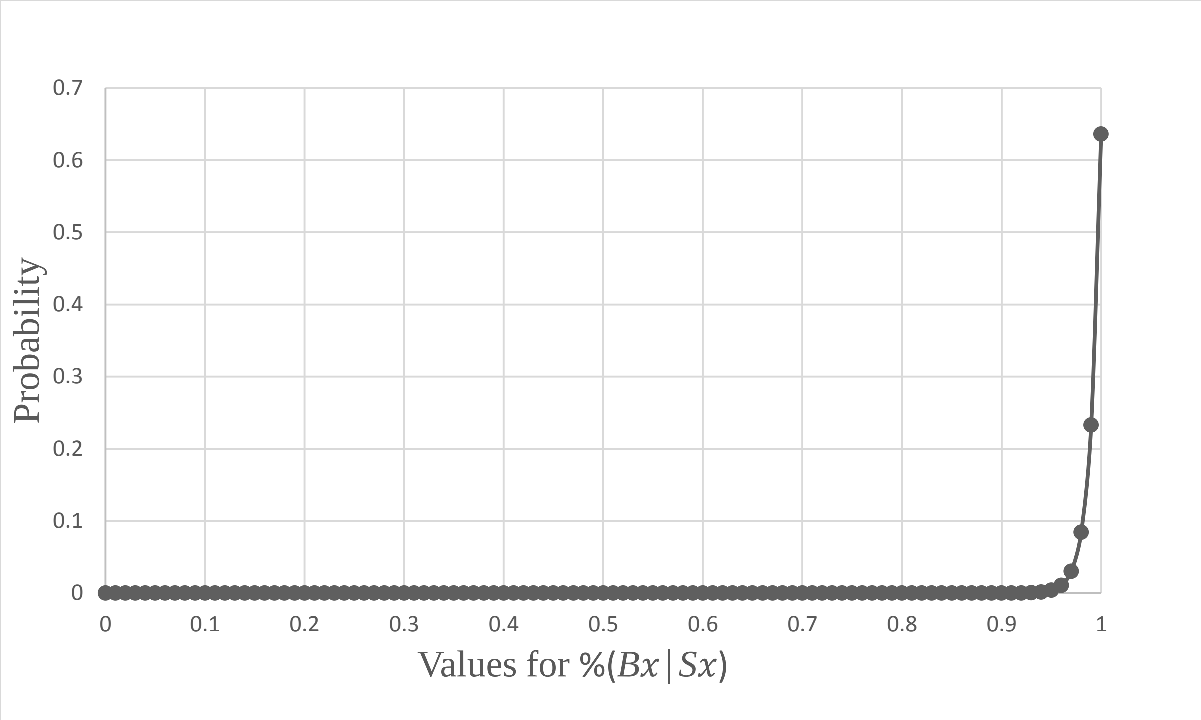
Sarah's distribution is fairly sensible. The sample of emeralds makes her highly confident that at least almost all emeralds are green even though she still attaches a positive but negligible probability to the other proportion formulas that are consistent with the evidence. Mitchell's distribution is less sensible. The thousands of emeralds that are consistently observed to be green has given him an almost non-committal attitude about the proportion of emeralds that are green. While Sarah's distribution seems sensible or close to it, we might say that Mitchell's is radically under-confident that at least most emeralds are green.

But suppose Mitchell also has other bizarre proclivities. Suppose that the thought experiment is same as the above except that the agents sample only two sapphires, both of which are blue, to estimate the proportion of sapphires in the area that are blue. They then have the following distributions where $\%(Bx|Sx)$ is the proportion of sapphires in the area that are blue, Sx stands for "x is a sapphire in the area" and Bx stands for "x is blue".

<u>Sarah's distribution after sampling two blue sapphires</u>



Mitchell’s distribution after sampling two blue sapphires



Sarah’s distribution is largely non-committal and at least somewhat sensible. The slight amount of evidence causes her intuitively to slightly favour the formulas where (approximately) over two thirds of sapphires are blue. In contrast, Mitchell’s distribution is radically over-confident that almost all sapphires are blue.

How do intuitions play a role in these distributions?

Well, perhaps Sarah and Mitchell update in a Bayesian style and intuition features in the determination of priors for the proportions. Maybe Sarah has an objective Bayesian intuition that she should use a distribution that is uniform over the above proportion formulas. Maybe Mitchell is also an objective Bayesian who, in the case of emeralds, intuitively thinks he should use a complicated language featuring a dazzling amount of possible colour predicates. This gives rise to a very complicated space of probability vectors for proportion expressions for each of the colours. (For example, consider a probability vector of proportion formulas corresponding to the statement “1% of emeralds are slightly pink with purple dots, 2% are blue with red stripes, 0% are magenta,…, and the other 97% are a yellowy-green”.) The enormity of colour predicates and consequent enormity of outcomes leads him to a uniform prior distribution that assigns minuscule probabilities to the formulas where at least most emeralds are green. They then update with the different intuitive priors to obtain drastically different posteriors. So that's one story. But perhaps both agents are instead just subjective Bayesians where the prior probabilities reflect intuitions that do not necessarily conform to some formal principle.

Or perhaps both agents just reflect on the sample evidence, then they directly form the distributions in a way which does not reflect Bayesian updating and they cannot specify why they think that those distributions are correct aside from just appealing to "intuition". This case may be more realistic insofar as agents seldom have a precise probability distribution which they then systematically update; they may not reflect on their credences or even have relevant credences until after the relevant evidence is presented to them. They may both also have putatively relevant background evidence, perhaps concerning the constancy of colour for other types of gemstone, but they may have different intuitions about the relevance of this evidence.

In any case, the details do not matter so much. What does matter is that intuition somehow influences the distributions and that no matter how this is done, one final distribution is more sensible than another.

In any case, I will assume that the reader shares the intuitions of myself and certain others that Sarah's distributions are at least close to being sensible whereas Mitchell's distributions are far from it.[190]

So Mitchell's intuitions are bizarre. We can suppose that, unknowingly for Mitchell, they are induced by some rare brain disorder triggered by observing certain colours. We can also suppose that they are nevertheless consistent with his evidence and other (perhaps equally bizarre) intuitions. From this perspective, Mitchell is blameless for his radical intuitions insofar as he knows no better than to make such inferences.

Regardless, we probably think that there is a sense in which Mitchell's intuitions and inverse inferences are incorrect.

Furthermore, these intuitions are, I would say, objectively incorrect in the sense that they are wrong irrespective of anyone's perspective, just as the claims that $2 + 2 = 5$ or that the earth is under 10,000 years old are objectively incorrect, irrespective of anyone's

[190] Even if the reader is unsympathetic with either agent's distributions, I suspect they would agree (along with the others that I have spoken to) that evidence of this kind should affect one's credences for proportion expressions. After all, how could the observation of 10,000 green emeralds fail to affect one's confidence that all or almost all emeralds are green? Presumably, then, the reader would think that an agent in these circumstances should have some probability distribution and that some distributions are more or less sensible than others. The reader can then substitute a more and a less sensible distribution of their choosing into the above thought experiment.

perspective.[191] In this sense, I take the thought experiment to suggest that there is an objective standard of correct intuitions just as there is an objective standard of correct claims about the physical world or mathematical truths.

The critic might object that claims of objective correctness about intuitions have unacceptably spooky truth-makers since it is unclear what makes it true that such intuitions are objectively correct.

However, one does not need to know what the truth-makers of a claim are in order to know that it is true. Indeed, I do not know what makes it true that $2 + 2 = 4$ (is it a platonic fact or some claim about the real-world?), but this is an objectively true claim irrespective of my ignorance (or indeed suspicion) of any truth-makers.[192] Presumably advocates of basic direct inference or the principle of indifference often agree that the intuitions underlying these norms are objectively correct while failing to agree about what makes them correct. Likewise, I may not know what makes it true that Mitchell's intuitions are objectively incorrect, but this is still a true claim irrespective of my comparable ignorance or suspicion.

Nevertheless, there are some candidate truth-makers for objective correctness. For example, perhaps, as a matter of objective fact, a certain kind of intuition relative to a certain kind of body of evidence would hinder the agent from the truth most of the time. Consider again Mitchell's under-confident or over-confident intuitions whereby enormous amounts of evidence make him only slightly confident that (approximately) all emeralds are green and slight amounts of evidence make him enormously confident that (approximately) all sapphires are blue. Statistically speaking, people may be hindered from getting close to the truth about relative frequencies most of the time if they had similar intuitions about similar evidence.[193]

I think that something like this kind of truth-maker is plausible. There arguably is something about the world that would make these kinds of intuitions deter the agent from the

[191] This notion of objective correctness may or may not sit well with what the reader's concept of 'objectivity' or 'correctness' is. In a sense, though, I am not necessarily trying to adhere to those concepts. I am just appealing to the intuition that, independent of anyone's perspective, there is something wrong about Mitchell's inferences and intuitions and then I am labelling this thing 'objective incorrectness' even if the reader objects to the label.

[192] As with almost any position in philosophy, there are opponents of the notion that mathematical claims are true, namely, fictionalists about mathematics. Nevertheless, I am conforming to the widely held thought that some mathematical claims are objectively correct. In any case, presumably the fictionalist has some story about why we should take seriously, or act on the assumption of, claims like $2 + 2 = 4$ instead of $2 + 2 = 5$. Perhaps something like this story may be articulated in defence of taking certain inductive intuitions seriously or acting on them *as if they were objectively correct*. The theory I outline may then be articulable with such a story.

[193] This kind of thinking resembles certain accounts of justification, including certain kinds of reliabilism and Ernest Sosa's account of *aptness* in his virtue epistemology. See Ernest Sosa, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge, U.K.; New York: Cambridge University Press, 1991), ch. 9 and Lemos, *An Introduction,* ch. 5.

truth in some sense. Regardless, defending this thought would require more space for definition and defence than I can afford here, particularly when closeness to the truth may require a technical account like so-called *gradational accuracy* and this thought also encounters an analogue of the reference-class problem known as *the generality problem*.[194]

So I settle for the claim that some intuitions are objectively correct or incorrect, even if we do not what know the truth-makers are for correctness claims.

But even if we do not know what the truth-makers are, we do know something about what they are not. Mitchell's intuitions are not objectively incorrect just because they conflict with our perspective of what is correct or some community's shared perspective of what is correct. If Mitchell believed that that the earth is currently less than 10,000 years old or that it is not the case that $2 + 2 = 4$, these beliefs are not incorrect because they conflict with the perspectives that we or our communities have. Supposing that we or our communities did not exist for there to even be such a conflict, Mitchell's inferences would still be objectively incorrect in a sense similar to these beliefs. For example, if Mitchell was the only person in existence and still had under-confident intuitions given enormous amounts of evidence and over-confident intuitions given scant amount of evidence, these intuitions would still be incorrect, perhaps because they would statistically lead him to be overly confident in false propositions and underly confident, so to speak, in true propositions a high percentage of the time. Intuitively, then, intuitions can be incorrect in a sense which transcends the perspective of any agent just as the falsity of these other beliefs do.

So I suggest we say that an inverse inference is *objectively rational* to the extent that it accords with relevant and objectively correct intuitions about the agent's evidence.[195] Note that this definition refers to *relevant* intuitions. This is because, in theory, some intuitions are relevant to the credence(s) of interest whereas others are not. The agent may have objectively correct intuitions underlying her inverse inferences, but objectively incorrect intuitions underlying her direct inferences. Yet the latter are (we suppose) irrelevant to her inverse inferences and so should not count against them.

This definition accommodates the possibility that one can have an objectively rational inverse inference on the basis of intuition despite lacking a formal method to tell them as

[194] For the generality problem, see Lemos, *An Introduction*, 93-94. Richard Pettigrew informally characterises gradational accuracy as a measure of "how good an agent's credal state is by its proximity to the omniscient state" which assigns credences of only 1 or 0 to true and false propositions respectively. See Richard Pettigrew, "Accuracy, Risk, and the Principle of Indifference," *Philosophy and Phenomenological Research* 92, no. 1 (2016): 36.

[195] I would follow tradition by instead speaking of objectively rationally *permissible* inverse inferences, but this term is more cumbersome. Hence, I just speak of objectively rational inverse inferences instead.

such. This should not be objectionable. It is analogous to one having an objectively rational basic direct inference because they have the putatively correct intuition which Kyburg, Teng, Bacchus et al. and others share even though one lacks a formal method to tell them as such.

However, a few disclaimers are warranted. The existence of an objective standard of correct intuition does not entail that we necessarily know what it is or that we agree about it. Like intuitions, certain mathematical theorems or hypothesised laws of physics may be (approximately) objectively correct even if no one knows or agrees as such. Furthermore, the correctness of an intuition is not necessarily solely a binary matter whereby the intuition is either correct or not. Instead, correctness may be graded concept whereby an intuition's *degree of* correctness is assessed by its proximity to some correct intuition.[196] Additionally, there might not necessarily be only one correct intuition in a given case. Perhaps there may be a set of objectively correct intuitions and a set of incorrect ones. These are topics that I wish to remain neutral on.

Also note that this definition of objectively rational inverse inferences does not negate the utility and normative power of formal methods. Perhaps the set of objectively correct intuitions for certain cases are those which align with the prescriptions of a formal method, such as those in statistics.

7.2. Subjectively Rational Inverse Inferences

Yet there is a sense in which Mitchell's inferences are not wrong.

Let me spell this out with an analogy. Suppose that a man named Daniel sees a hungry dog and, out of his compassion, feeds the dog a block of chocolate. Unbeknownst to Daniel, however, the chocolate is poisonous to the dog. Hence, it kills the dog gradually and painfully within a couple of hours. A question arises: did Daniel do something wrong? There is a sense in which he did do something wrong insofar as his action had a bad outcome relative to the objective fact that he painfully killed the innocent dog. But there is a sense in which he did not do something wrong insofar as his action was entirely well-motivated from his perspective and he could not have known better. After all, it was not as if he did something wrong in the sense that he intended to knowingly harm the dog to no benefit.

[196] For example, if there is exactly one correct intuition specifying that a particular credence should have value f given evidence e, then the proximity and degree of correctness for some other intuition prescribing value g regarding the same credence and evidence could be measured by the numerical distance between f and g.

Inverse inferences likewise are assessable relative to two dimensions of rightness or wrongness, for lack of better words. There is the objective dimension according to which Mitchell's inferences were objectively irrational. However, there is a sense in which Mitchell's inference were not wrong insofar as, from his perspective, he was blameless. His evidence did not obviously entail that his intuition was wrong. For example, it is not as if knowledge that 10,000 emeralds are green is logically equivalent to knowledge that 10,000 emeralds are green *and* the probability that all emeralds are green should be 87%. His intuition also did not conflict with any of his other intuitions. He just unknowingly had a misleading intuition brought about by a brain disorder. Given his innocence from his perspective, then, we might say that there is a sense in which his inference was *subjectively rational,* but *not objectively rational*. The subjective rationality of an inverse inference, then, is another dimension along which the rightness or wrongness of an inverse inference may be assessed.

On this note, we may say that an agent's inverse inference is *subjectively rational* to the extent that it accords with the agent's relevant intuitions and evidence.[197]

One might object that this definition of subjective rationality entails that agents like Mitchell have subjectively rational inferences when we think that they are *irrational*. Yet we can accommodate such thinking by saying that agents like Mitchell have *objectively irrational inferences.*

It is important to distinguish these two dimensions for appraising inductive inferences. Otherwise, like the case with Daniel, there is a potentially confusing tension between the desires to evaluate the rightness of an inference relative to factors internal to the subject's perspective or relative to factors that somehow transcend or are external to it. Indeed, in traditional epistemology, an analogous tension manifests itself in debates between internalists and externalists about justification, with some such as Sosa resolving the tension by appealing to two dimensions of the epistemic merit of a belief.[198] Somewhat following Sosa, I likewise propose two distinct dimensions – one regarding subjectively rational inferences and the other regarding objectively rational inferences.[199]

Nevertheless, the two dimensions are related. All objectively rational inferences are the subjectively rational inferences of some agent, if not only a hypothetical one. I would also

[197] This does not rule out the possibility of subjectively irrational inferences. For example, one could have intuitions that their evidence shows that something bad is very probably about to happen but they may choose to irrationally ignore to the evidence and wishfully think that it will not happen.
[198] See a useful summary of the debate in Lemos, *An Introduction to the Theory of Knowledge,* ch. 6.
[199] See Sosa, *Knowledge in Perspective*, ch. 9.

say that subjectively rational inferences (tacitly) aim to be objectively rational and one estimates what the objectively rational inference would be via what their subjectively rational inference would be. Just as Daniel aimed to have an action that had objectively good consequences according to his subjective viewpoint, so too do epistemically rational agents aim to have objectively rational inverse inferences through their subjectively rational inferences.

However, subjectively rational inverse inferences can fail to achieve their goal. Indeed, although Mitchell was subjectively rational, his inference failed to be objectively rational.

7.3. Arguments for and against Trusting in Intuition

So we have categorised epistemically rational inverse inferences based on intuition into subjectively and objectively rational inferences, but it is another question to ask whether people should sometimes rely on their intuitions when making inverse inferences even though they lack a formal method to guide them.

Let us consider arguments for and objections to an affirmative answer to this question.[200]

One objection is that people's intuitions sometimes can lead them astray (à la Mitchell) and so they should not rely on intuition, but they should instead rely on or try to find formal methods of inductive inference. This is a particularly seductive objection given the findings from the so-called *heuristics and biases research programme*. The findings show that humans often have misleading intuitions underpinning their probability judgments.[201] For example, an occasionally pernicious heuristic is known as the *availability heuristic* whereby the probability of a proposition is judged based on how easily instances of that proposition (or similar propositions) being true come to mind. This is problematic when, say, the agent is afraid that there is a decent chance of being attacked by a shark at a beach merely because she recently watched the film *Jaws,* even though the chance such an attack is negligible. The objection may then take these findings as evidence that people can have untrustworthy intuitions.

[200] For many of the following objections, I thank a commentator who shall remain anonymous.

[201] See a concise and insightful discussion of the programme in Anthony O'Hagan, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley and Tim Rakow, *Uncertain Judgements: Eliciting Experts' Probabilities* (London; New Jersey: Wiley, c2006), 31-52.

However, that some intuitions are untrustworthy does not entail that all others are similarly untrustworthy. Someone's proclivity to fall prey to the availability heuristic does not mean that I or numerous others should discard our intuitions supporting, say, the rationality of basic direct inference or the principle of indifference in certain circumstances.

Furthermore, formal methods, just like intuitions, sometimes also lead us astray for certain inferences - arguably methods like classical confidence interval estimation.[202] Presumably this shows that we need not use theories like these for the inferences that we have mistakenly applied them to. Yet we similarly need not use misleading intuitions for guiding the inferences that we have mistakenly applied them to.

Surely the lesson is that if a formal method or intuition is untrustworthy, it is because of factors specific to the formal method or intuition and not just because other methods or intuitions were untrustworthy.

In any case, it is consistent with the preceding discussion that one should sometimes reason in accordance with a formal method. Arguably a formal method is just codified intuition dressed in formal clothing inasmuch as, say, the random-worlds method formally encodes intuitions about basic direct inference, independence, indifference and others. Perhaps, in fact, the set of objectively correct intuitions for inductive inference are ultimately expressible in some statistical theory to be discovered (although I doubt it). Plausibly, many statistical theories do reflect objectively correct intuition, including certain forms of Bayesian and classical data analysis for random-samples. So trusting in intuition accommodates their guidance. *But* it can also guide agents in the innumerable everyday contexts of interest where the agent needs to make inferences but does not have enormous amounts of time, expertise and money to design and implement a sophisticated probability sampling strategy.

Another objection is that trusting in intuition as Sarah does and as the theory allows has no precedent in statistical practice and so it should be treated with suspicion. But it explicitly does have precedent in subjective Bayesian data analysis where the priors (and, by implication, the posteriors in part) are determined with the help of intuition. Even then, as mentioned, other formal methods for statistical analysis are arguably just codified intuitions in disguise.

A further objection is that something better than intuition may come along, so we should not trust intuition. However, something better than any formal method may come along too,

[202] See Jaynes and Kempthorne, "Confidence Intervals vs. Bayesian Intervals."

so this cannot count as a good objection to trusting intuition any more than it is a good objection to trusting formal methods.

One might also object that some do not have intuitions to guide inverse inferences in the first place, so they cannot be relied on. But at most, this objection only shows that some people lack intuitions, not that those who have them should not rely on them.

Another objection is that people disagree about the right intuitions for inverse inferences and so this should cause us to doubt their trustworthiness. However, people have such disagreement about all sorts of things including basic direct inference (think of Kyburg and Levi), the principle of indifference (think of White and James Joyce), Chinese rooms and the like.[203] Yet I have never seen someone doubt the trustworthiness of their intuitions merely because others do not have them, and rightly so.[204] I would not then think that any disagreement about intuitions regarding inverse inferences should make trusting in these intuitions particularly controversial or objectionable. Furthermore, the disagreement about appropriate intuitions for inverse inferences should not be exaggerated so as to obscure any agreement about them. For instance, the six people I have asked all agree that Sarah's inferences and intuitions are at least better than Mitchell's; while this is not a large sample to draw rigorous inferences from, it does support (to whatever the extent) the possibility that there is intersubjective agreement about epistemically rational inverse inferences. So I do not find this objection compelling both because we rationally trust in intuitions despite disagreement and because it is not clear that there is that much significant disagreement about intuitively rational inverse inferences in any case.

A further objection is that we have unclear intuitions about appropriate inverse inferences and so we should not trust the intuitions. We might only be able to intuitively specify, for instance, that the reasonable inference lies in the general ballpark of a range of possible inferences, but intuition does not make it clear exactly where it is. Regardless, one might not sympathise with this objection since they might have clear intuitions about what their distribution should be. But even if intuitions are unclear, this does not mean that there should be no inferences based on them; rather, the inferences, or one's opinion about them, should just reflect the unclarity in the intuitions. For example, one's inferred probability distribution may be understood as a *rough approximation* to their (unclear) intuition. Or perhaps their

[203] See White's defence of the principle in White, "Evidential Symmetry." See also Joyce's critique of it in James M. Joyce, "A Defense of Imprecise Credences in Inference and Decision Making," *Philosophical Perspectives* 24, no. 1 (2010): 281-323.

[204] Unless I do not have those intuitions or similar ones, in which case that someone should doubt those intuitions.

inference should somehow involve *indeterminate* or *imprecise probabilities* which specify the range of (possibly) permissible inferences. Or perhaps their distribution should be a *weighted average* of the distributions which could possibly reflect their unclear intuitions. Precedents for these kinds of probabilities and inferences are found either in practice or in theory, and it is doubtful that something like them would not be useful here.[205] In any case, even if we do have some unclear intuitions which do not to favour specific inferences, this does mean that we should not trust them, but just that it is unclear what trusting in them would mean.

However, if one can rely on their intuitions for assigning probabilities, this raises the potentially troubling question of why one might want to gather more evidence to further inform their probabilities, including perhaps random-samples analysed with a formal statistical theory. Yet this question arises for many forms of inference, including inferences with the principle of indifference or direct inferences where one might want to gather more specific statistics or evidence to inform probabilities. Surely it should not then be a particularly concerning problem for intuitive inverse inferences.

One might also object that we will descend a slippery slope if we abandon the use and pursuit of formal methods when making inverse inferences from non-probability samples. In particular, one could argue that if we do abandon such methods because they are frequently unable to accommodate intuitions about similarity, background evidence and other issues, then we may also abandon the pursuit and use of other useful theories because they also occasionally conflict with intuition. However, this is not necessarily true. For example, one might (as I do) accept the random-worlds method as an insightful theory of direct inference to validate and guide our intuitions and they may even seek to extend its applications. This is perfectly consistent with abandoning formal methods of inverse inference from non-probability samples. Recall also that the motivation for abandoning formal methods of inverse inference from non-probability samples is simply that the relevant factors guiding such inferences are too subjective, vague or complicated to be stuffed into a general formal method, either now or in the future. This is not necessarily the case for many other subjects which theories speak to, so there is no reason to think that we will descend such a slippery slope when the motivation for doing so is lacking.

[205] Brendon Brewer informs me that statisticians occasionally see their distributions as approximations to the appropriate distribution or treat their distribution as a weighted average of the set of possibly appropriate distributions. We have also seen that the concept of indeterminate probabilities is advocated by, or sympathised with, a range of scholars, including Levi, Hájek, Smithson and others.

Against these objections, I would like to contend that (at least some) agents should sometimes trust in their inverse inferences and the intuitions underpinning them, even if the intuitions might (unknown to the agents) be objectively incorrect.

Defending my contention, however, is problematic.

Let us see why.

Three potentially appealing arguments may be given for this contention. One argument relies on the premise that humans make inverse inferences that are or appear to be reliable most of the time and hence most other inverse inferences are likely to be reliable. An example of such a reliable inference is when the agent inversely infers that most of the people at a party know Ben based on a sample of the party-goers and the agent subsequently finds that most of the party-goers do in fact know Ben. One could probably give many other personal success stories like this, success stories which arise so frequently that their banal success goes unnoticed. A second argument is that nature has fashioned humans through evolution so as to have (at least approximately) correct intuitions as these would conduce to survival and reproduction; hence, we should trust in our intuition. A third argument is that we make progress in determining what intuitions are correct by considering different situations (like thought experiments) and potential responses to them; so we should be confident that these intuitions which we have progressed to are trustworthy. Examples of such progress have arguably been the intuitions spoken to in chapter 5 and 6 and which I list in the next section (Section 6.4.).

While I am sympathetic to these arguments and think they should (somehow) make us feel more comfortable about relying on intuition, they are not entirely compelling. The first tacitly presupposes the conclusion that intuitions can be trusted; it relies on an intuitive inverse inference about the reliability of a population of inverse inferences based on the reliability of a sample of such inferences. The second argument merely as it is presented above is not compelling since it is possible (albeit not necessarily probable) that nature fashioned us to have unreliable intuitions. So we need a reason to discredit this possibility if the argument is to work. But we cannot have this reason without the first argument or an analogously circular and intuitive inference that evolution would be likely to bestow adaptive and correct intuitions on humans given a sample of other beneficial adaptations that it has bestowed.[206]

[206] This "analogous intuitive inference" may take the form of a singular predictive inference rather than an inverse inference. The inference might be, for example, that a sample of human faculties (such as sight, hearing and others) proved to be well-adapted to the environment, so the human faculty of intuitive probability estimation is probably likewise well-adapted and reliable. But this predictive inference likewise presupposes some trust in intuition by intuitively assigning the probability on the basis of this sample evidence.

The circularity here is reminiscent of the problem confronting structuralism which was mentioned in Section 4.1. The third argument is also weak since we cannot assume that we have made progress toward trustworthy intuitions unless we already question-beggingly trust the intuitions that we supposedly progressed to, at least more so than prior intuitions.

Here, then, is how I support the contention.

First, I do not contend that *anyone* should trust *any* of their intuition and inferences, but only that *certain* intuitions may be trusted. More specifically, these are the intuitions that I think are objectively correct, that I and others somewhat like me have and that Mitchell and others like him lack. Of course, specifying what these intuitions are is beyond the scope of this thesis just as much as the task is of specifying what one's evidence is or should be generally speaking; the appropriate intuitions vary from context to context just as the evidence does.

Then I take a foundationalist approach insofar as I rely on intuitions that I cannot articulate non-circular justifications for: I intuitively think that we need to make inductive inferences in everyday life, that I should trust these intuitions even if they are possibly incorrect and that there is no better alternative to such trust. Naturally, I cannot defend such thinking without (controversially) presupposing the trustworthiness of one or another set of intuitions.[207] Regardless, I take this thinking as foundational in the sense that I can trust it despite lacking articulable reasons for doing so.[208]

The critic might ask, "Why do you think that your thinking is right, especially when someone like Mitchell could think similarly regarding the trustworthiness of his bizarre intuitions and inferences?"

I have no answer to give, but the critic should not be dissatisfied. If Hume's problem of induction has taught us anything, I would say it is that induction inevitably involves foundational assumptions about rational inferences which we regard as (objectively) correct and which cannot be given articulable and non-circular justifications.[209] Despite this, everyone relies on induction and thinks that this is (often) the right thing to do. So why not

[207] I do concede that one might take the premises of the first two aforementioned arguments about past success and evolution as foundational and use these premises to justify the trustworthiness of intuitions about appropriate inferences. I do not find this approach remarkably objectionable. Regardless, I intuitively take the intuitions about rational inferences themselves to be the natural stopping point in the chain of justification, and I do so for no articulable reason. In any case, the substance of the theory of inverse inference that I endorse does not need to rest on treating this as the natural stopping point.

[208] See a useful introduction to foundationalism in Lemos, *An Introduction*, Chapter 3.

[209] See the problem of induction discussed in Section 4.1 and Vickers, "The Problem of Induction." I think this lesson from the problem of induction would be accepted as uncontroversial, particularly since Howson, Urbach and others appear to think that the problem of justifying induction non-circularly is insoluble. See Howson and Urbach, *Scientific Reasoning*, 2.

just extend this foundational reliance to intuitions about appropriate inverse inferences where formal methods are lacking? In fact, I think that we often can do no better, that this is often exactly what we do and that this is rightly so.

7.4. A Theory without Content?

The ideas that I endorse in the preceding discussion, then, constitute a theory about inverse inference. The theory specifies two dimensions by which to assess the epistemic rationality of inverse inferences, permits one to make such inferences in the absence of a formal method and recommends that certain intuitions be trusted in doing so.

Yet the critic might object that this theory lacks normative content; it does not constrain or tell us what specific credences we should have.

This is dubious since the theory can have normative content for the individual who accepts it. For example, unlike any other theory that I know of in the literature, it allows me to claim that some of my necessary inverse inferences in every day contexts are objectively rational (or close to it) while Mitchell's are not, even though I lack a formal method telling me as such. Not only does no formal method tell me as such, but the problems in Chapter 6, I believe, raise doubts that a formal method will ever be able to claim as such by objectively and quantitatively accommodating all of my background evidence and judgments of similarity between the sample(s) and population(s) of interest. The formal methods of objective Bayesians and proponents of (quasi-)classical statistical theory either do not speak specifically to it or do speak to it but are unable to accommodate cases with varying background evidence and judgments of similarity. Subjective Bayesianism only validates part of this claim on certain conditions. So long as one's credences conform to the probability calculus and updating by conditionalisation, they are rational and in this sense I can form probability distributions based on intuitions. However, subjective Bayesianism does not permit me to object to Mitchell's inferences as being objectively irrational. This is because there are (updated) probability distributions which accord with Mitchell's distributions *and* are permissible in the subjective Bayesian's eyes because they accord with the calculus and conditionalisation. In a sense, the theory presented in this chapter harmonises the subjective Bayesian thought that rational inferences have intuitive and non-formal elements with the objective Bayesian thought that not all such elements are objectively correct. What I propose, then, is a mixture of the two thoughts, one that is perhaps best called *sobjective Bayesianism*.[210]

In any case, the theory I have articulated has no less content than subjective Bayesianism and surely subjective Bayesianism is a theory with enough content to deserve some attention.[211]

Of course, the theory does not answer all specific questions about what one's distribution should be. For example, just as Bacchus et al.'s method does not specify what one's KB should be, so too does the theory not specify exactly what all of one's intuitions should be (although I list some suggestions in what follows). But like every theory of inductive inference, its content has to stop somewhere. The theory articulated here, I believe, is at least worth some attention insofar as it started to make some progress. However, specifying exactly where current formal method signs off and brute intuition begins to guide inference is a task beyond the scope of this thesis. Nevertheless, inverse inferences from non-probability samples are beyond the boundaries of formal methods, at least in many cases.

Regardless, I think that several substantive points for guiding our inferences already have emerged from this thesis.

We make progress toward obtaining inferences that are objectively rational by a process of acquiring, refining and discarding intuitions (and perhaps by other things).[212] This is done through considering various situations (whether hypothetical thought experiments or real-world scenarios) and the various inferences that can be made from them. This thesis has appealed to many thought experiments to uncover intuitions for assessing particular inferences. This was the case for thought experiments regarding the pigeons (see Sub-Section 6.3.5.), the toy car factory (see Sub-Section 5.1.) and others.

Consequently, via these experiments and other considerations, this thesis has unearthed various more specific intuitions that should guide intuitive inverse inferences. These are some of them:

[210] The fact that at least one scholarly commentator on the theory has objected to it also inclines me to think that it has some significant content.
[211] Perhaps an analogous point can be made for John Norton's material theory of induction.
[212] Note that acquiring better intuitions about evidence is distinct from acquiring *better evidence* to inform inferences.

- *Non-Zero Probability of Universal Generalisations:* Universal generalisations in infinite domains should not receive a probability of 0 when counter-evidence to the generalisation is lacking.[213] (See Section 5.1.)
- *Sensitivity to the Population Size:* The probabilities of proportion formulas about populations should (often) be sensitive to the (perhaps estimated) size of the population so that the probabilities would be affected were the population size to vary. (See Section 5.1.)
- *Method of Sampling:* Inferences from samples should consider whether the sample has been obtained in a way which is biased or should affect the inferences from the sample evidence. (See Sub-Section 6.3.1.)
- *Evidence about Other Populations and Samples:* Where relevant, evidence about other populations or samples should be taken into account. Perhaps this could be done via the direct route of assigning probabilities directly to proportion formulas on the basis of all the relevant evidence or it may be done indirectly by affecting the prior probabilities of proportion formulas which are then updated on the basis of sample evidence (as per subjective Bayesian data analysis). (See Sub-Section 6.3.2.)
- *Language Invariance:* Probabilities for propositions should depend on the evidence bearing on the propositions themselves and not on the choice of language by which propositions are expressed. (See Sub-Section 6.3.3.)
- *Accommodating Uncertainty about the Categories in a Population:* Probability distributions should take into account evidence or intuitions about the categories of object that could possibly be in the population, particularly given observed variation in categories in the sample. In particular, if it is possible that an unobserved category could appear in the population, then the proportion formulas should have probabilities which do not preclude this possibility. For example, a sample consisting of only green emeralds (that does not include the whole population of emeralds) should not itself cause one to assign a probability of 1 to the proportion formula stating that all emeralds are green since this precludes this possibility.[214] (See Sub-Section 6.2.4.)
- *Evidential Implications of Similarity:* Loosely speaking, the more similar a non-probability sample is to the population that it is a subset of, the greater the probability

[213] It should go without saying that I am assuming that such counter-evidence does not include the assumptions that the domain size is infinite and that this requires the use of a limit for the probability which tends to zero as the domain size goes to infinity.
[214] Of course, to give rigour to this guideline, one would need to give an account of what is "possible" in terms of logical possibility, physical possibility or some other concept. I do not currently have such an account, but I still think that this guideline provides useful, albeit rough, guidance for inverse inferences.

that the sample relative frequency of interest reflects the population relative frequency. However, the agent should consider the relevance of certain similarities or dissimilarities as some may be more significant than others (such as similarities that are causally relevant to any target property of interest). (See Sub-Section 6.3.5.)

- *Heuristics and Biases:* The agent should guard against the undue influence of biases and heuristics such as the availability heuristic.

This list of guidelines may not prescribe specific inferences from non-probability samples, but I think that it, or indeed any theory, can do no better given the complexity of the issues involved. Note also that these are not the only guidelines that one might take into account and indeed other guidelines may be relevant in the agent's specific context. In particular, the theory I have outlined is consistent with a Bayesian guideline that many would endorse (including myself) but that I cannot argue for or elaborate on here: the agent should conform to axioms of probability and where she has subjectively rational prior credences and likelihoods and (virtually) certain new evidence, she should update via conditionalisation.

7.5. Summary and Further Research

So we can now summarise what we can jestingly call a sobjective Bayesian theory of inverse inference.[215] There is a standard of objectively correct intuitions which an agent's intuitions may or may not satisfy, even if no one knows what the correct intuitions are or what makes them correct. An inverse inference is *objectively rational* to the extent that it accords with relevant and objectively correct intuitions about the agent's evidence. An agent's inverse inference is *subjectively rational* to the extent that it accords with the agent's relevant intuitions and evidence. When evaluating the epistemic "rightness" or "wrongness" of an agent's inverse inference, we do so with respect to at least one of these dimensions of epistemic rationality. Through our subjectively rational inverse inferences, we aim for objectively rational inverse inferences and we appraise inferences as (perhaps more or less) objectively rational. One progresses toward objectively correct intuitions via (perhaps among other things) considering various responses to situations like thought experiments or real-world problems. From this thesis, a list of more specific intuitions have emerged to guide

[215] I have narrowly avoided the temptation to self-servingly call this the *intuitive theory of inverse inference.*

inverse inferences. Formal methods provide useful norms for inverse inferences in some contexts (such as those involving random samples), but they do not and may never provide useful norms in other contexts. In these latter contexts, the agent might nevertheless make objectively rational inferences using intuition while lacking such a method. Agents sometimes should (in a foundationalist manner) trust in their intuitions and subjectively rational inverse inferences, even if they are potentially not objectively rational. So that is the sobjective Bayesian theory of inverse inference.

However, this theory, like other philosophical theories, has problems and areas for further research. These include:

- explicating the relationship of intuition to credence
- accommodating the possibility of incoherent intuitions
- formulating the theory in a way which accommodates unclear intuitions
- articulating a specific account what the truth-makers are for claims about objectively correct intuitions and
- ascertaining more precisely where formal methods of inverse inference are not helpful and where brute intuitions are.

Another problem and area for further research concerns the question of how to represent the space of possible proportions or relative frequencies in a population, particularly when the population size is unknown. Some insight on this question may be gained by turning to statistical practice whereby the space of proportions is idealised in a continuous or discrete manner, seemingly with no harm.[216]

And there may be other problems for the theory which I have not considered, especially given philosophical tradition whereby someone proposes a theory and another person completely disagrees it and has an objection to show why.

[216] That this is statistical practice is affirmed by Brendon Brewer and also Howson and Urbach, *Scientific Reasoning*, 32.

8. Random-Worlds and Total Probability

The preceding chapter presented a theory about epistemically rational inferences to probability values for proportions formulas.

This theory can supplement the random-worlds method. Both the random-worlds method and the aforementioned theory are concerned with answering the following question: given information about frequencies, what credences are the right or rational inferences? The random-worlds method is a theory about direct inference from proportion formulas; the theory in the preceding chapter is a theory about inverse inferences to probabilities for proportion formulas. How, then, can the resulting probability distribution over the proportion formulas be integrated with the random-worlds method to constrain credences?

In response to this question, this chapter outlines a proposal for integrating the resulting probability distributions with the insights from the random-worlds method, one which invokes the spirit of the theorem of total probability.[217]

8.1. The Core Idea

It is useful to outline this proposal using a thought experiment. Suppose the agent picks up a newspaper at a café which someone has spilled coffee onto. The headline reads "…0% fatally injured on *Midnight Express* train yesterday" where part of the first digit is covered in coffee.[218] Nevertheless, the first digit visibly has an "o" shape at the bottom of it, meaning that it is either "6" or "8". Unfortunately, the rest of the newspaper does not indicate what the percentage of fatally injured passengers was. The agent is concerned because she knows that her friend Aaron was on that train yesterday when it crashed because she saw him off when the train departed.

[217] An alternative to this proposal may be to somehow introduce probability statements for proportion formulas into the language of the random-worlds method and thereby into knowledge bases. Bacchus et al. briefly hint at the possibility of this alternative in Bacchus et al., "From Statistical Knowledge Bases," 129. Assessing the merit of this alternative is beyond the scope of this thesis.

[218] I thank Jeremy Seligman for proposing this more realistic and cunning version of a previous hypothetical situation of mine.

Given the visible information on the newspaper, the agent is indifferent between the proportion formulas that are consistent with the headline. So she has the credences $P(\%(Ax|Bx) = 0.6) = 0.5$ and $P(\%(Ax|Bx) = 0.8) = 0.5$ where Ax stands for "x was fatally injured", and Bx stands for "x was a passenger on the *Midnight Express*". The agent is wondering what her credence $P(Aa)$ should be where a stands for Aaron and Aa expresses the proposition that Aaron was fatally injured. Clearly this headline should affect the agent's credence that Aaron had been fatally injured, but the agent is not sufficiently confident in either proportion formula in order for one of them to be accepted into her knowledge base.

What should be the value of $P(Aa)$, then?

As a first step, we can model the agent's evidence as consisting of two parts, a (virtually) certain part and an uncertain part. The first part is a set of accepted statements that we can symbolise with KB'. Here, this includes statements such as Ba (Aaron was a passenger on the train). The second part is a set of propositions that the agent is uncertain about. In this case, we can suppose that $\%(Ax|Bx) = 0.6$ and $\%(Ax|Bx) = 0.8$ constitute the uncertain part of the agent's evidence. We can suppose here that if the agent were to accept either $\%(Ax|Bx) = 0.6$ or $\%(Ax|Bx) = 0.8$ into her knowledge base, then she would have a knowledge base KB which satisfies the conditions of Bacchus et al.'s theorem for basic direct inference (see Sub-Section 4.2.1.). That is, if $KB \equiv (\%(Ax|Bx) = f \,\&\, KB')$ where f is either 0.8 or 0.6, then a does appear in Ax, Bx or KB' and so $P_{Random-Worlds}(Aa|KB) = f$ (supposing that the limit exists).

As a second step, we can turn to the spirit of the total probability theorem and a related norm known as *Jeffrey conditionalisation*. The theorem of total probability is as follows:

$$P(e) = P(e|h_1)P(h_1) + \cdots + P(e|h_k)P(h_k)$$

where e denotes some proposition and $\{h_1, \dots, h_k\}$ is a partition of hypotheses.

The structure of the theorem of total probability finds expression in Jeffrey conditionalisation, also known as *probability kinematics.*[219] Jeffrey sought to provide an extension of standard conditionalisation via Bayes's theorem where some hypothesis h is updated in the light of

[219] This rule can be found in Richard C. Jeffrey, *The Logic of Decision,* 2nd ed. (Chicago: University of Chicago Press, 1983), ch. 11. The rule is also implied by Maximum Entropy methods. See Ariel Caticha,"Lectures on Probability, Entropy, and Statistical Physics," *arXiv: 0808.0012* (2008): 139.

some evidentiary statement e where one's credence for e is some value aside from 0 or 1.[220] Jeffrey conditionalisation is the prescription to update in accordance with the following formula:

$$P'(h) = P(h|e)P'(e) + P(h|{\sim}e)P'({\sim}e)$$

where $P(.)$ and $P'(.)$ are respectively the agent's prior and posterior credences relative to the receipt of the evidence.

We can then reflect the agent's two-part evidence in an equation similar to the two above, an equation that integrates the agent's uncertainty about the proportion formulas and random-worlds credences. However, the random-worlds credence will be used in an unconventional way. They will involve formulas whose probabilities are conditional, so to speak, on proportion formulas such as $P_{Random-Worlds}(\alpha(c)|\%(\alpha(x)|\beta(x)) = f \,\&\, KB')$ where $\alpha(c)$ is some target formula mentioning c, $\%(\alpha(x)|\beta(x)) = f$ is a proportion formula and $\beta(c) \in KB'$. The discussion concerns proportion formulas which the agent is uncertain about and so it is misleading to represent them in part with a KB symbol which denotes a knowledge base of accepted propositions. Hence, we will use cases in which the proportion expressions feature in the antecedent of the random-worlds credences without being included in, or denoted by, a KB.

The proposal is that:

$$P(Aa) = P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.8 \;\&\, KB')P(\%(Ax|Bx) = 0.8 \;\&\, KB') \\ + P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.6 \;\&\, KB')P(\%(Ax|Bx) = 0.6 \;\&\, KB')$$

We can now calculate $P(Aa)$. Recall that $P(\%(Ax|Bx) = 0.6) = 0.5$ and $P(\%(Ax|Bx) = 0.8) = 0.5$. By the probability calculus and the supposition that KB' is certain, $P(\%(Ax|Bx) = 0.6 \,\&\, KB') = 0.5$ and $(\%(Ax|Bx) = 0.8 \,\&\, KB') = 0.5$. Recall also that if $KB \equiv (\%(Ax|Bx) = f \,\&\, KB')$ where f is either 0.8 or 0.6, then a does appear in Ax, Bx or KB' and so $P_{Random-Worlds}(Aa|KB) = f$. Hence, $P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.8 \;\&\, KB') = 0.8$ and $P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.6 \;\&\, KB') = 0.6$.

[220] Here, I understand an evidentiary statement to be a statement expressing some proposition which is uncertain but which should have implications for the agent's other credences.

Therefore, we have all the assignments necessary to assign a value to $P(Aa)$:

$$\begin{aligned} P(Aa) &= P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.8 \ \& \ KB')P(\%(Ax|Bx) = 0.8 \ \& \ KB') \\ &+ P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.6 \ \& \ KB')P(\%(Ax|Bx) = 0.6 \ \& \ KB') \\ &= 0.8 \times 0.5 + 0.6 \times 0.5 = 0.7 \end{aligned}$$

0.7 is an intuitively reasonable value for $P(Aa)$ to take in the agent's case, irrespective of the probabilistic machinery by which one obtains it.

Here, then, the proposal preserves the insights from the random-worlds method by utilising random-worlds credences and it also accommodates the clearly relevant information about the uncertain proportion formulas $\%(Ax|Bx) = 0.8$ and $\%(Ax|Bx) = 0.6$.

From this, it should also be clear how the proposal can be implemented with other kinds of uncertain propositions, such as those of the form *Ba*. Nevertheless, I will not dwell on this topic here as it is not the focus of the thesis.

8.2. The Reference Class Problem and Combinations of Uncertain Evidence

The proposal can be extended to cases where the agent is uncertain about multiple partitions of evidentiary statements, particularly involving ones about alternative reference classes

Here we will take the problem of "incomparable reference classes" as our point of illustration (recall the discussion of such classes in Sub-Section 4.2.7.). Let us make the proposal more abstract and suppose that *A*, *B* and *C* are some arbitrary one-place predicates and *a* some arbitrary constant. Suppose *a* is known to belong to the two incomparable reference classes, ***B*** and ***C***. Suppose further that the agent has the credences $P(\%(Ax|Bx) = 0.9 \,\&\, KB') = 0.5$, $P(\%(Ax|Bx) = 0.8 \,\&\, KB') = 0.5$, $P(\%(Ax|Cx) = 0.1 \,\&\, KB') = 0.5$ and $P(\%(Ax|Cx) = 0.4 \,\&\, KB') = 0.5$. Similarly to the preceding case, suppose that if $KB \equiv (\%(Ax|Bx) = f \,\&\, \%(Ax|Cx) = g \,\&\, KB')$ where f is either 0.9 or 0.8 and g is either 0.1 or 0.4, then KB meets Bacchus et al.'s theorem regarding the treatment of incomparable reference classes. In this case, neither *B* or *C* mention Ax or a, the intersection of the classes ***B*** and ***C*** is known to consist of exactly one member denoted by a, KB contains no other information and so $P_{Random-Worlds}(Aa|KB) = \frac{fg}{fg+(1-f)(1-g)}$. Suppose also that these proportion formulas are independent; so $P(\%(Ax|Bx) = 0.9\ \&\%(Ax|Cx) = 0.1\ \&\ KB') = P(\%(Ax|Bx) = 0.9\ \& KB')P(\%(Ax|Cx) = 0.1\ \&\ KB')$ and similarly for the probabilities for the other pairs of proportion formulas. In this sense, the agent is uncertain about two partitions of evidentiary statements, the partition $\{\%(Ax|Bx) = 0.8, \%(Ax|Bx) = 0.9\}$ and the partition $\{\%(Ax|Cx) = 0.1, \%(Ax|Cx) = 0.4\}$. This gives rise to a probability distribution over what I will call a partition of *all of the possible evidentiary outcomes*. In this case, this is the partition:

$$\{\%(Ax|Bx) = 0.9\ \&\ \%(Ax|Cx) = 0.1, \%(Ax|Bx) = 0.9\ \&\ \%(Ax|Cx) = 0.4,$$
$$\%(Ax|Bx) = 0.8\ \&\ \%(Ax|Cx) = 0.1, \%(Ax|Bx) = 0.8\ \&\ \%(Ax|Cx) = 0.4\}$$

By the probability calculus, the aforementioned assignments of 0.5 to each proportion formula and the assumption of independence, $P(\%(Ax|Bx) = f\ \&\ \%(Ax|Cx) = g\ \&\ KB') = 0.25$ where f is either 0.9 or 0.8 and g is either 0.1 or 0.4. Hence, the probability of each of the possible evidentiary outcomes has a value of 0.25.

The proposal is that we can use a total-probability-style formula to determine $P(Aa)$ in this case with the four possible outcomes:

$$\begin{gathered} P(Aa) = \\ P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.9\ \&\ \%(Ax|Cx) = 0.1\ \&\ KB') \\ \times P(\%(Ax|Bx) = 0.9\ \&\%(Ax|Cx) = 0.1\ \&\ KB') \\ + P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.9\ \&\ \%(Ax|Cx) = 0.4\ \&\ KB') \\ \times P(\%(Ax|Bx) = 0.9\ \&\ \%(Ax|Cx) = 0.4\ \&\ KB') \\ + P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.8\ \&\ \%(Ax|Cx) = 0.1\ \&\ KB') \\ \times P(\%(Ax|Bx) = 0.8\ \&\ \%(Ax|Cx) = 0.1\ \&\ KB') \\ + P_{Random-Worlds}(Aa|\%(Ax|Bx) = 0.8\ \&\ \%(Ax|Cx) = 0.4\ \&\ KB') \\ \times P(\%(Ax|Bx) = 0.8\ \&\ \%(Ax|Cx) = 0.4\ \&\ KB') \end{gathered}$$

The values of the random-worlds credences are provided by the random-worlds method (see Sub-Section 4.2.7.). As mentioned, the other probabilities in the right-hand expression are each 0.25.

We then have the following result:

$$P(Aa) = 0.6 = 0.5 \times 0.25 + 0.86 \times 0.25 + 0.31 \times 0.25 + 0.73 \times 0.25$$

This illustrates how the proposal may be extended to handle cases involving relevant statistics for multiple reference classes, and indeed cases with multiple relevant but uncertain statements in general.

8.3. Summary and Further Research

So this is the core idea behind the proposal in this chapter. Suppose $\alpha(c)$ is some formula which mentions the constant *c*.[221] Suppose also that the agent's evidence consists of 1) some uncertain and relevant proportion formula(s) of the form $\%(\alpha(x)|\beta(x))$ that give rise to a partition of all of the possible evidentiary outcomes and 2) a KB' that, if conjoined with one of the outcomes, satisfies the conditions for a KB in some relevant random-worlds theorem. Then, credences for $\alpha(c)$ should be calculated via an equation patterned after the theorem of total probability. The left-hand expression in the equation is $P(\alpha(c))$. The right-hand expression in the equation is the sum of the products of pairs of terms. One of the terms in such a pair is a random-worlds credence for $\alpha(c)$ given the proportion formula(s) and the KB'. The other term is the agent's credence for the proportion formula(s) conjoined with KB'. These credences for proportion expressions collectively form a (complete) probability distribution over the partition of all of the possible evidentiary outcomes. (Of course, if no relevant random-worlds credence exists, perhaps because the relevant limits do not exist, then the proposal is silent on how to constrain one's credences.)

We might say, then, that the proposal is one for *direct inference from uncertain relative frequencies*. As far as I know, a proposal of this kind is fairly novel since the extant theories on direct inference assume that the relevant proportion formulas for the relative frequencies are accepted.

The values of credences for proportion formulas might be obtained in different ways. One way is via intuitive inverse inferences from sample evidence. Another way is via probability

[221] The proposal can also be extended to handle direct inferences about multiple objects in a sample, not just one, but this is more complex.

assignments to proportion formulas on the basis of testimonies that are relevant to the formulas (such as the case of the newspaper).

The proposal bears structural similarities to the theorem of total probability and it resembles Jeffrey conditionalisation insofar as one's uncertainty for a formula equals a sum of products for two functions. However, it differs from Jeffrey conditionalisation in that Jeffrey conditionalisation is a dynamic constraint for updating: it specifies that if one has certain *prior* credences and *then* they receive some uncertain evidence at *a later time*, then their *posterior* credence for a formula should be a function of their other credences *at that time* as well as their *prior* credences. The proposal above, however, is a static constraint regulating one's credences at one time: it specifies that if one's evidence can be modelled in a certain way *at a given time*, then some other credence should be a function of the random-worlds credences as well as some of their other credences *at that same time*.

How the proposal relates to updating is a topic that is beyond the topic of this thesis, particularly when it may be relevant to such complicated topics as belief revision and commutativity.[222] This, then, is an area for further research.

8.4. Problems

So that is the proposal.

As is always the case, however, there are problems. In this section, I will survey the ones that appear salient to me.

[222] For example, there is a complicated debate about whether the non-commutativity of Jeffrey conditionalisation is a defect. See Marc Lange, "Is Jeffrey Conditionalization Defective by Virtue of being Non-Commutative? Remarks on the Sameness of Sensory Experiences,"*Synthese* 123, no. 3 (2000): 393-403, Zoltan Domotor, "Probability Kinematics and Representation of Belief Change," *Philosophy of Science* 47, no. 3 (1980): 395, Brian Skyrms, *Choice and Chance: an Introduction to Inductive Logic*, 2nd ed. (Encino, Calif.: Bickenson Pub. Co., 1975), 197 and Frank Döring, "Why Bayesian Psychology is Incomplete," *Philosophy of Science* 66, no. 3 (1999): 382-4. Another relevant paper is one by some physicists which defends the non-commutativity of Maximum Entropy methods which imply Jeffrey conditionalization. See Adom Giffin and Ariel Caticha, "Updating Probabilities with Data and Moments," *AIP Conference Proceedings* 954, no. 1 (2007): 74-84.

8.4.1. Direct Inference and Intuitive Inverse Inference vs. Intuitive Predictive Inference

This chapter and the preceding one outlined ideas by which one may make inferences from a sample to credences about a population and then inferences from those credences to credences about another sample, perhaps consisting of one object.

Yet if one can intuitively make an inverse inference from a sample and then direct inferences from possible proportions to a credence about an object of interest as per the proposal, why not just make an intuitive singular predictive inference straight from the sample to the object?

I cannot see anything objectionable about such predictive inferences. However, any such inference would ideally be coherent with the agent's other inferences.

Does this mean that this chapter and the preceding have no significance?

Well, my answer is partly "no" and partly "maybe no".

The "maybe no" part is that the most rational way of making inferences about the probable features of currently unobserved objects *might* (often) consist of inverse inferences from a sample and direct inferences. Yet that is a topic that I am agnostic about.

Regardless, the "no" part of my answer is that the proposals of this chapter and the preceding are useful in their own right. Agents may have uncertain information about proportion formulas that derives from sources other than sampling (as per the newspaper case above). Hence, the proposal in this chapter may be useful even in the absence of relevant inverse inferences. Additionally, the theory in the preceding chapter may provide guidance for inverse inferences that are of interest independently of direct or predictive inferences. For example, a bus company manager may be considering whether to give a bus driver a salary raise based on how punctual a driver he is. So she takes a non-probability sample of trips from the bus driver to estimate how punctual his trips generally are. Here, the manager's intention is to make an inverse inference about the proportion of bus trips that are punctual to estimate the punctuality of the driver. The intention is not to make a direct or predictive inference about the next sampled trip. Here, then, the theory about intuitive inverse inferences is relevant. Hence, the proposals in this chapter and the preceding one are arguably significant in their own right.

In any case, this thesis has responded to a problem posed by Bacchus et al., that is, the problem of how to account for sampling and for uncertainty about proportion formulas on the random-worlds account. Insofar as this problem is significant, so too is my response to it.

Regardless, this thesis has followed in the footsteps of Kyburg, Teng and Pollock in considering constraints on probabilities arising from both inverse and direct inferences rather than predictive inferences. So if they do not consider these kinds of constraints to be objectionable or useless, perhaps others would not too.

8.4.2. Computational Complexity

The reader may have noticed that the preceding cases in which I outline the proposal are often unrealistic. They involve credences for only a few proportion formulas, but the agent is often best understood as having non-zero credences for many proportion formulas, or rather a continuous space of proportion formulas. This makes the calculation of probabilities computationally complex.

To me, this is not a theoretical problem. The agent could, in principle, utilise the proposal with credences for finitely many proportion expressions, irrespective of their number. I also think that the proposal could be extended to continuous spaces.

Regardless, utilising the proposal with many credences is computationally complex. This, in turn, may make the proposal difficult to implement in practical circumstances.

In a sense, I agree that this a problem, but it is not my problem. The aim of this thesis has been to propose an epistemically rational approach to constraining credences on the basis of information about relative frequencies. There is no requirement that an epistemically rational approach be one that is practically easy to adopt. This is particularly the case for Bayesians who advocate conformity to the standard axioms of probability, a putatively rational standard that is impossible for real agents to fully satisfy. Therefore, I do not see how the difficulties in implementing this proposal would constitute a serious objection, at least any more so than for many widely-accepted norms of rationality.

8.4.3. The Problems Confronting the Random-Worlds Credences

The proposal to use the random-worlds method as per the aforementioned proposal risks inheriting various problems from the method. In particular, the proposal may be thought to be language dependent (see Sub-Section 6.3.3.).

My response to this worry is to suggest that the random-worlds credences only be used when the agent has their own probability distribution over the relevant proportion formulas or when such formulas are accepted. This entails that the proposal will only utilise probabilities that, as mentioned in Sub-Section 6.3.3., are not dependent on language by virtue of their dependence on proportion formulas.

However, this response risks accusations of *ad hockery* as it appears to say that the method merely should not be used to address some questions which it clearly delivers counter-intuitive verdicts on and which clearly count against its general applicability.

I sympathise with this accusation and I will respond to it in the next sub-section.

8.4.4. The Problem of a Better Measure, Ad Hockery and Other Problems

While I have defended the proposal in this chapter, I also candidly acknowledge some reasons to doubt its finality or legitimacy as an answer to the question of how to constrain credences on the basis of frequency information.

For one, as mentioned, the proposal might look ad hoc.

Furthermore, if there is a better measure which prescribes useful inverse inferences, it very probably does not look like the random-worlds method's uniform measure. So perhaps this better measure will ultimately invalidate inferences sanctioned by the random-worlds method. Perhaps a case of this would include the method's prescribed inferences that involve incomparable reference classes.

In response to these concerns, I confess to ad hockery and acknowledge the possibility of there being a better measure or theory which would rightly supplant using this proposal.

Nevertheless, I still think that the proposal is rational. This proposal, while not problem free, at least validates intuitions about reasonable inductive practices (such as the newspaper case above) and it allows for uncertainty regarding proportion formulas to be incorporated into inductive inferences. The random-worlds method faces problems, but only in circumstances where we need not apply it. Additionally, in other circumstances where it does not appear to face problems (such as the problem of incomparable reference classes), there is no reason to think that some other particular inference is better than the one which it prescribes.

For these reasons, I offer the proposal regardless of these problems, suggesting that we use the method where it appears to work well and ignore it where it does not.[223]

[223] This suggestion is analogous to how scientists, particularly physicists, use idealised models. They will sometimes use idealised models for making certain predictions that are adequate for their purposes and even if the models deliver false predictions about a subject matter that is not of immediate interest. For example, the physicist's model of an ideal pendulum swinging in a room without air resistance may be adequate for making approximate predictions about the movements of certain swinging pendulums but inadequate for making specific predictions about the air resistance acting on those pendulums. Similarly, the random-worlds method may be adequate as an idealised model of rational direct inferences, but inadequate as, say, a model of rational inverse inferences. For more on the widespread use of idealisations in science, see Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Oxford University Press, 1983) and Paul Teller, "Fictions, Fictionalization, and Truth in Science," in *Fictions in Science: Philosophical Essays on Modelling and Idealisation,* ed. Mauricio Suárez (New York: Routledge, 2009).

9. Conclusion

This thesis considered the question of how information about frequencies should constrain credences. The random-worlds method was presented as a potential answer to the question that has favourable features. Nevertheless, the method has its limitations, particularly when it comes to prescribing rational inverse inferences and encapsulating the uncertainty that agents occasionally have about proportion formulas. I explored some alternative measures which hoped to overcome these limitations, ultimately concluding that they face a variety of problems which make the measures often inapplicable to actual cases of interest.

The thesis then took a positive route in articulating some new ideas. I outlined and defended a theory of inverse inference, one which is consistent with agents potentially making epistemically rational and intuitive inferences from samples with or without formal methods. Lastly, I presented a proposal for how to constrain credences via integrating the resulting uncertainty from these inferences with the random-worlds method. These ideas, then, constitute a suggestion for how evidence about relative frequencies should influence Bayesian probabilities.

Appendix: Logical Glossary

What follows is an informal glossary of logical notation which attempts to provide non-specialist audiences with a rough understanding of the notation in this thesis.

Symbol	Informal Translation	Example
Fx	"x has the property F."	"x has the property of being a frog."
$\boldsymbol{F}$	"The set of all objects satisfying Fx."	"The set of all frogs."
$\sim\alpha$	"It is not the case that α."	"It is not the case that Adam is happy."
$\alpha \supset \beta$	"If α, then β." "α implies β."[224]	"If Adam is happy, then Billy is sad."
$\alpha \ \&\ \beta$	"α and β are true."	"Adam is happy and Billy is sad."
$\alpha \vee \beta$	"α is true, β is true or both α and β are true."	"Adam is happy, Billy is sad or Adam is happy *and* Billy is sad."
$\alpha \equiv \beta$	"α is true if and only if β is true." "α implies β and β implies α."	"If Adam is happy, then Billy is sad and if Billy is sad, then Adam is happy."

[224] The relevant concept of implication here is material implication, not the concept of metalinguistic implication which features at the end of the glossary.

$\forall x(Fx)$	"For all *x*, *Fx* is true."	"Everything is a frog."
$\forall x(Fx \supset Gx)$	"For all *x*, such that *Fx* is true, *Gx* is true." "Every *F* is a *G*."	"All frogs are green."
$\exists x(Fx)$	"There exist some *x* such that *Fx*."	"There exists some thing such that it is a frog." "A frog exists."
$\exists x(Fx \,\&\, Gx)$	"There exists an *x* such that *Fx* and *Gx*."	"There exists a thing such that it is a frog and it is green." "A green frog exists."
$\%(Fx) = f$	"The proportion of things for which Fx is true is *f*."	"The proportion of things that are frogs is 0.000000000001."
$\%(Gx\|Fx) = f$	"The proportion of things satisfying $\beta(x)$ that also satisfy $\alpha(x)$ is *f*."	"The proportion of frogs that are green is 0.8."
$\alpha \vDash \beta$	"α implies β." "In every model in which α is true, β is also true."	"In every model in which it is true that there is a square object, it is also true that there is a rectangular object."

Bibliography

Ayer, A.J. *Probability and Evidence*. London: Macmillan, 1972.

Bacchus, Fahiem, Adam Grove, Joseph Halpern and Daphne Koller, D. "From Statistical Knowledge Bases to Degrees of Belief." *Artificial Intelligence* 87, no. 1 (1996): 75-143.

Bacchus, Fahiem, Adam Grove, Daphne Koller and Joseph Y. Halpern. "From Statistics to Beliefs." In *AAAI-92: Proceedings*, 602-608. Menlo Park, Calif.; Cambridge: MIT Press, c1992.

Bartha, Paul. "Analogy and Analogical Reasoning." *The Stanford Encyclopedia of Philosophy.* Accessed May 5, 2016. http://plato.stanford.edu/archives/fall2013/entries/reasoning-analogy/.

Bradley, Seamus. "Imprecise Probabilities." *The Stanford Encyclopedia of Philosophy*. Accessed June 21, 2016. http://plato.stanford.edu/archives/sum2015/entries/imprecise-probabilities/.

Brewer, Brendon J. "Introduction to Bayesian Statistics." Course notes, STATS 331, The University of Auckland, 2014. (Accessible at https://www.stat.auckland.ac.nz/~brewer/stats331.pdf.)

Butler, Joseph. *The Analogy of Religion, Natural and Revealed to the Constitution and Course of Nature*. 20th ed. New York: Mark H. Newman & Co., 1851.

Carnap, Rudolf. "The Aim of Inductive Logic." In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress,* edited by Ernest Nagel, Patrick Suppes and Alfred Tarski, 303-318. Amsterdam: North-Holland, 1966.

———. "On Inductive Logic." *Philosophy of Science* 12, no. 2 (1945): 72-97.

———. "A Basic System of Inductive Logic, Part I." In *Studies in Inductive Logic and Probability,* edited by Rudolf Carnap and Richard Jeffrey, Vol. 1, 33–165. Berkeley: University of California Press, 1971.

———. "A Basic System of Inductive Logic, Part II." In *Studies in Inductive Logic and Probability,* edited by Richard Jeffrey, Vol. 2, 7-155. Berkeley: University of California Press, 1981.

———. *Logical Foundations of Probability.* London: Routledge and K. Paul, 1951.

———. *The Continuum of Inductive Methods*. Chicago: University of Chicago Press, 1952.

Cartwright, Nancy. *How the Laws of Physics Lie*. Oxford: Oxford University Press, 1983.

Caticha, Ariel. “Lectures on Probability, Entropy, and Statistical Physics.” *arXiv: 0808.0012* (2008): 1-162.

Colyvan, Mark, Helen M. Regan and Scott Ferson. “Is it a Crime to belong to a Reference Class.” *Journal of Political Philosophy* 9, no. 2 (2001): 168-181.

Crupi, Vincenzo. “Confirmation.” *The Stanford Encyclopedia of Philosophy*. Accessed May 5, 2016. http://plato.stanford.edu/archives/fall2015/entries/confirmation/.

Dietz, Thomas and Linda Kalof. *Introduction to Social Statistics: The Logic of Statistical Reasoning*. Chichester, West Sussex; Malden, MA.: Wiley-Blackwell, 2009.

Domotor, Zoltan. “Probability Kinematics and Representation of Belief Change.” *Philosophy of Science* 47, no. 3 (1980): 384–403.

Döring, Frank. “Why Bayesian Psychology is Incomplete.” *Philosophy of Science* 66, no. 3 (1999): 379-389.

Elga, Adam. “Subjective Probabilities should be Sharp.” *Philosophers’ Imprint* 10, no. 5 (2010): 1-11.

Goodman, Nelson. *Fact, Fiction, and Forecast.* Cambridge, MA: Harvard University Press, 1955.

Giffin, Adom and Ariel Caticha. “Updating Probabilities with Data and Moments.” *AIP Conference Proceedings* 954, no. 1 (2007): 74-84.

Hájek, Alan. “Interpretations of Probability.” *The Stanford Encyclopedia of Philosophy*. Accessed May 8, 2015. http://plato.stanford.edu/archives/win2012/entries/probability-interpret/.

———. “The Reference Class Problem is your Problem too.” *Synthese* 156, no. 3 (2007): 563-585.

———. “Symmetry is the Very Guide of Life.” *The Pennsylvania State University*. Accessed May 3, 2016. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.318.2402&rep=rep1&type=pdf.

———. “What Conditional Probability Could Not Be.” *Synthese* 137, no. 3 (2003): 273-323.

Hájek, Alan and Michael Smithson. “Rationality and Indeterminate Probabilities.” *Synthese* 187, no. 1 (2012): 33-48.

Hilpinen, Risto. “Carnap’s New System of Inductive Logic.” *Synthese* 25, no. 3 (1973): 307-333.

Hintikka, Jaakko. “On a Combined System of Inductive Logic.” In *Studia Logico-Mathematica et Philosophica in Honorem Rolf Nevanlinna die Natali eius Septuagesimo,* 21-30. Helsinki: Societas Philosophica Fennica, 1965.

———. “Reply to Isaac Levi.” In *Philosophy Jaakko Hintikka*, edited by Randall E. Auxier and Lewis Edwin Hahn, 777-782. Chicago: Open Court, c2006.

———. “Towards a Theory of Inductive Generalization.” In *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*, 2nd ed., edited by Yehoshua Bar-Hillel, 274–288. Amsterdam: North-Holland Pub Co., 1965.

Hintikka, Jaakko and Niiniluoto, Ilkka. “An Axiomatic Foundation of Inductive Generalization.” In *Formal Methods in the Methodology of Empirical Sciences*, edited by Marian Przelecki, K. Szaniawski and Ryszard Wójcicki, 57–81. Dordrecht: Reidel, 1976.

Howson, Colin and Peter Urbach. *Scientific Reasoning: The Bayesian Approach,* 3rd ed. Chicago: Open Court, c2006.

Huber, Franz. “Formal Representations of Belief.” *The Stanford Encyclopedia of Philosophy*. Accessed January 31, 2016. http://plato.stanford.edu/entries/formal-belief/.

Jaynes, E.T. *Probability Theory: The Logic of Science*. New York: Cambridge University Press, 2003.

Jaynes, E.T., and Oscar Kempthorne. “Confidence Intervals vs. Bayesian Intervals.” In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. II., edited by William L. Harper and C.A. Hooker, 175-257. Dordrecht: Springer Netherlands, 1976.

Jeffrey, Richard C. *The Logic of Decision,* 2nd ed. Chicago: University of Chicago Press, 1983.

Joyce, James M. “A Defense of Imprecise Credences in Inference and Decision Making.” *Philosophical Perspectives* 24, no. 1 (2010): 281-323.

Koons, Robert. “Defeasible Reasoning.” *The Stanford Encyclopedia of Philosophy*. Accessed June 21, 2016. http://plato.stanford.edu/archives/spr2014/entries/reasoning-defeasible/.

Kyburg, Henry and Choh Man Teng. *Uncertain Inference*. New York: Cambridge University Press, 2001.

Lange, Marc. “Is Jeffrey Conditionalization Defective by Virtue of being Non-Commutative? Remarks on the Sameness of Sensory Experiences.”*Synthese* 123, no. 3 (2000): 393-403

Levi, Isaac. “Direct Inference.” *The Journal of Philosophy* 74, no. 1 (1977): 5-29.

———. “Imprecision and Indeterminacy in Probability Judgment.” *Philosophy of Science* 52, no. 3 (1985): 390-409.

Li, Ming and Paul Vitányi. *An Introduction to Kolmogorov Complexity and its Applications.* New York: Springer, c2008.

Lucas, Samuel R. “Beyond the Existence Proof: Ontological Conditions, Epistemological Implications, and In-Depth Interview Research.” *Quality & Quantity* 48, no. 1 (2014): 387-408.

Maher, Patrick. “Explication of Inductive Probability.” *Journal of Philosophical Logic* 39, no. 6 (2010): 593-616.

Meacham, Christopher. “Impermissive Bayesianism.” *Erkenntnis* 79, no. 6 (2014): 1185–1217.

Momen, Moojan, ed. *The Bábí and Bahá'í Religions 1844-1944: Some Contemporary Western Accounts.* Oxford: George Ronald, 1981.

Niiniluoto, Ilkka. “The Development of the Hintikka Program.” In *Handbook of the History of Logic*, vol. 10, edited by Dov M. Gabbay, and John Woods, 311-356. Amsterdam; Boston: Elsevier, 2010.

Norton, John D. “Analogy.” *University of Pittsburgh.* Accessed May 7, 2016. http://www.pitt.edu/~jdnorton/papers/Analogy.pdf.

———. “There are No Universal Rules for Induction.” *Philosophy of Science* 77, no. 5 (2010): 765-777.

O'Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley and Tim Rakow. *Uncertain Judgements: Eliciting Experts' Probabilities.* London; New Jersey: Wiley, c2006.

Pettigrew, Richard. “Accuracy, Risk, and the Principle of Indifference.” *Philosophy and Phenomenological Research* 92, no. 1 (2016): 35-59.

Pollock, John. “Direct Inference and Probable Probabilities.” Accessed December 6, 2015. http://johnpollock.us/ftp/PAPERS/Direct%20Inference.pdf.

———. *Nomic Probability and the Foundations of Induction*. New York: Oxford University Press, 1990.

———. “Probable Probabilities (with Proofs).” Accessed December 6, 2015. http://johnpollock.us/ftp/PAPERS/Probable%20Probabilities%20with%20proofs.pdf.

———. “The Theory of Nomic Probability.” *Synthese* 90, no. 2 (1992): 263-299.

Reichenbach, Hans. *The Theory of Probability*. 2nd ed. Berkeley: University of California Press, 1949.

Shapiro, Stewart. “Classical Logic.” *The Stanford Encyclopedia of Philosophy*. Accessed April 20, 2016. http://plato.stanford.edu/archives/win2013/entries/logic-classical/.

Sosa, Ernest. *Knowledge in Perspective: Selected Essays in Epistemology.* Cambridge, U.K.; New York: Cambridge University Press, 1991.

Skyrms, Brian. *Choice and Chance: an Introduction to Inductive Logic,* 2nd ed. Encino, Calif.: Bickenson Pub. Co., 1975.

Strevens, Michael. “Bayes, Bayes’ Theorem, Bayesian Approach to Philosophy of Science.” Vol. 1 of *Encyclopedia of Philosophy,* edited by Donald M. Borchert, 495-502. 2nd ed. Detroit: Macmillan Reference USA, 2006.

———. *Tychomancy*. Cambridge, Mass.: Harvard University Press, 2013.

Suárez, Mauricio, ed. *Fictions in Science: Philosophical Essays on Modelling and Idealisation.* New York: Routledge, 2009.

Sznajder, Marta. “What Conceptual Spaces can do for Carnap’s Late Inductive Logic.” *Studies in History and Philosophy of Science* 56 (2016): 62-71.

Talbott, William. “Bayesian Epistemology.” *The Stanford Encyclopedia of Philosophy*. Accessed July 22, 2014. http://plato.stanford.edu/archives/sum2011/entries/epistemology-bayesian/.

Thompson, Steven K. *Sampling,* 3rd ed. Hoboken, NJ: John Wiley & Sons, 2012.

Thorn, Paul. “Three Problems of Direct Inference.” Doctoral thesis, The University of Arizona, 2007.

———. “Two Problems of Direct Inference.” *Erkenntnis* 76, no. 3 (2012): 299-318.

Titelbaum, Mike. “Fundamentals of Bayesian Epistemology: Chapter 5.” *Fitelson.* Accessed May 8, 2015. http://fitelson.org/bayes/titelbaum_ch5.pdf.

Vickers, John. “The Problem of Induction.” *The Stanford Encyclopedia of Philosophy*. Accessed May 6, 2016. http://plato.stanford.edu/archives/spr2016/entries/induction-problem/.

Vineberg, Susan. “Dutch Book Arguments.” *The Stanford Encyclopedia of Philosophy*. Accessed May 6, 2016. http://plato.stanford.edu/archives/spr2016/entries/dutch-book/.

Weisberg, Jonathan. "Varieties of Bayesianism." In *Handbook of the History of Logic*, vol. 10, edited by Dov M. Gabbay, and John Woods, 477-551. Amsterdam; Boston: Elsevier, 2010.

White, Roger. "Evidential Symmetry and Mushy Credence." In *Oxford Studies in Epistemology*, vol. 3, edited by Tamar S. Gendler and James Hawthorne, 161-186. Oxford: Oxford University Press, 2010.

Williamson, Jon. *In Defence of Objective Bayesianism*. Oxford; New York: Oxford University Press, 2010.

Zabell, Sandy L. "Carnap and the Logic of Inductive Inference." In *Handbook of the History of Logic*, vol. 10, edited by Dov M. Gabbay, and John Woods, 265-309. Amsterdam; Boston: Elsevier, 2010.

———. "Confirming Universal Generalizations." *Erkenntnis* 45, no. 2-3 (1996): 267-283.

———. "Predicting the Unpredictable." *Synthese* 90, no. 2 (1992): 205-232.