A comprehensive theory of induction and abstraction, part I

Cael L. Hasse*

June 27, 2017

Abstract

I present a solution to the epistemological or characterisation problem of induction. In part I, Bayesian Confirmation Theory (BCT) is discussed as a good contender for such a solution but with a fundamental explanatory gap (along with other well discussed problems); useful assigned probabilities like priors require substantive degrees of belief about the world. I assert that one does not have such substantive information about the world. Consequently, an explanation is needed for how one can be licensed to act as if one has substantive information about the world when one does not. I sketch the outlines of a solution in part I, showing how it differs from others, with full details to follow in subsequent parts. The solution is pragmatic in sentiment (though differs in specifics to arguments from, for example, William James); the conceptions we use to guide our actions are and should be at least partly determined by preferences. This is cashed out in a reformulation of decision theory motivated by a non-reductive formulation of hypotheses and logic. A distinction emerges between initial assumptions–that can be non-dogmatic–and effective assumptions that can simultaneously be substantive. An explanation is provided for the plausibility arguments used to explain assigned probabilities in BCT.

In subsequent parts, logic is constructed from principles independent of language and mind. In particular, propositions are defined to not have form. Probabilities are logical and uniquely determined by assumptions. The problems considered fatal to logical probabilities–Goodman's 'grue' problem and the uniqueness of priors problem are dissolved due to the particular formulation of logic used. Other problems such as the zero-prior problem are also solved.

A universal theory of (non-linguistic) meaning is developed. Problems with counterfactual conditionals are solved by developing concepts of *abstractions* and corresponding *pictures* that make up hypotheses. Spaces of hypotheses and the version of Bayes' theorem that utilises them emerge from first principles.

Theoretical virtues for hypotheses emerge from the theory. Explanatory force is explicated. The significance of effective assumptions is partly determined by combinatoric factors relating to the structure of hypotheses. I conjecture that this is the origin of simplicity.

Contents

1 Preface

This series of papers is a labour of love, paid for on borrowed time. I do not guarantee the originality of the ideas, though I haven't seen anything quite like them. Erudite predantry, while good in appropriate doses can suffocate the valuable thought if taken too seriously.

Considering the ambition of the goal, the effort spent on some subjects may be considered inadequate. One reason is time, for which I apologise in advance. Another

*caelthetutor@gmail.com

1

is that the presentation of the theory roughly reflects the theory's own views on evidence and argument. Namely, the pragmatic sentiments that every conception is fallible, and the conceptions we use in the world are at least partly determined by our values; arguments that things must be a certain way[1] are based on premises that, in lieu of strong evidence or theoretical virtue, come from what we *prefer* to be true. Moreover, intuitions do not count as strong evidence. As the theory is based on my own intuitions[2] formed through my training as a physicist, my positive arguments, interpreted in a precise way–just like every logical argument in philosophy–should not be considered very robust[3]. Consider the series of papers as a confession (Nietzsche, 2003); my most coherent account of intuitions built up over the years. Thus instead of detailed arguments from some supposedly necessarily true premises, I focus on providing the conceptual core of the theory, showing its theoretical virtues, features and results. This is in the vain hope that one day robust evidence for or against it can be found[4]. I imagine this could be done in at least two ways. One, if the theory succeeds in helping to solve the original goal of the work of finding a basis for deriving Quantum Theory or a successor from first principles along lines similar to that of Fuchs (2016), and if this successor could lead to new experimental predictions that are seen. Two, if the theory leads to the construction of artificial general intelligence.

In some cases, familiar words are used in perhaps unfamiliar ways. Words like proposition, substance, incommensurability and versimilitude to name a few. This is partly because I lack the imagination to come up with replacements that sound appropriate. But it is also because I want to emphasise the similarity of the intuitions used in defining the concepts.

I genuinely believe there are some good–though not infallible–ideas in the paper. If you take the time to digest them, you might find them worthwhile.

2 Introduction

We have two parts to the problem of induction (Hume, 1910): (1) the justification of general standards of inductive inference and (2) the characterisation of these general standards. Though the series of papers mostly focusses on (2), I'll briefly discuss (1).

Traditionally, an inference to something was considered justified if one is justified in assuming something else and can use ones standards to verify the inference from the latter to the former. The idea is that these standards are efficacious in some way; if one uses deductive inference, one is supposed to be able to avoid error in ones predictions. If one uses inductive inference, error is supposed to be minimised in some way. This idea corresponds to a substantive claim about the world. An example of such a claim is Hume's 'uniformity of nature'; the claim that the past resembles the future (though this is vague). As Hume argued, it is eminently possible that this not be the case. The fear is that if we can't justify the claim in our standards, then we potentially have no basis to know anything substantive about the world. Moreover, we may have no reason why one ought to be rational.

Goodman (1983) considered the traditional version of the question of (1) dissolved and I think his reasons are good, that there is something awry in the formulation of the problem itself. To demand that there is a justification of standards of inference (inductive or deductive) of the form that is traditionally hoped amounts to a demand for efficacy of our predictions that is wishful thinking, a demand of nature that may not be fulfilled and we should not expect it to be.

Goodman's positive account is one where general standards of inference are justified or not by their level of conformity to inferences we are willing to accept as valid. There is a circular process of mutual adjustment between general rules and accepted inferences. The problem of induction reduces to this process of characterisation.

---

[1] As opposed to negative ones such as arguing that a theory is inconsistent, incoherent or has a conceptual gap.

[2] Namely the fallibility of all conceptions, the limitations of analysis of language, the dangers of over formalisation, and a certain distaste for abundance of conceptual ingredients.

[3] I don't state this in a pejorative way; the search for foundations is important, just hard.

[4] See the next section for comments on how the theory provides a limited solution to the is/ought problem.

I argue for certain distinctions that complicate this understanding of the problem. I argue that the traditional version of the question of (1) not only comes from wishful thinking, but that any possible answer will be a form of *nonsense*. Carnap (1991) argued that metaphysical questions such as whether certain things exist have two types of answers, one internal to a linguistic framework and one external. The first type potentially answerable, the second nonsense. Though the theory explicated here will not use linguistic frameworks, a similar distinction will emerge: meaningful answers to questions occur only internal to a hypothesis. For us, hypotheses do not concern only empirical truths, but non-empirical ones also, including ones of validity and justification. The validity of an inference is defined by a hypothesis and we can't externally ask about the validity of inferring said hypothesis by using its internal definition of validity.

Though I doubt if Goodman would think in this way, we can consider his recasting of justification as an example of inventing a definition of 'justification' different to the internal definition by a hypothesis, allowing for meaningful inquiry into the internal justification of the hypothesis.

We must distinguish between standards of inference and the process in which one goes about trying to achieve them. The process may be circular; one may decide to consider justification defined one way in light of what one accepts as instances of justified inference, but then go back and revise the list of accepted inferences in light of the defined notion of justification. However, related to the above comments on meaningfulness, where we require a strict hierarchy of inference, the standards outlined here will not allow circular inference. This affects the dissolution of the problem of justification.

Standards of inference–subsumed into standards of rational action choice–will be considered as comprising of two components; the normative component and a declarative one in the form of a hypothesis.

The normative component is the normative appropriateness of *following* the edicts of the standards. The following of the standards is the alignment of a series of actions and circumstances with the edicts of the standards. The edicts are dependent upon input called *initial assumptions*. In this series of papers, to say X is assumed by an agent is to say that that agent's actions align with the standards in the situation where X is an input into the standards. This notion differs from one where to say an agent assumes something implies something about the state of the agent. An example of how this differs is that in the first definition, it is possible for one to appropriately say that an agent both does and doesn't assume X, if inputing X into the standards has no consequences upon the appropriate choice of actions. Moreover, it is possible to follow the standards and assume nothing.

In this way, the standards may be followed but the declarative component not necessarily assumed; there can be uncertainty about hypotheses of various competing standards[5]. One may in principle gather evidence for or against the hypotheses of different standards. Goodman's circular *process* of justification is to be considered a reflection of the *non-circular standards* of weighing up of various hypotheses for various standards.

Initial assumptions are not and cannot be justified. More will be said on this in following papers. They can however satisfy normative desiderata.

I don't give any positive reasons why the following of the standards described here are normatively appropriate. That is beyond the scope of the project. My goal is to provide what I consider a 'good' characterisation of the standards.

3 Desiderata of characterisation

After considering the justification problem solved, Goodman (1983) understood the problem of induction as one of characterising a 'good' theory of confirmation. Given certain evidence, a confirmation theory determines whether a hypothesis is confirmed or not (and perhaps also to what degree it is or isn't confirmed). An amendment to this is that confirmation also depends on the context

---

[5]This requires our standards to be flexible enough such that it is possible to follow them while assuming others.

of assumptions surrounding the evidence. Bayesian Confirmation Theory is (I consider rightly) one of the most popular of such theories. Our goal is to come up with a theory that has the best features of Bayesian Confirmation Theory but overcomes its problems.

Bayesian Confirmation Theory (BCT) is a theory of probabilities. In some formulations there are two types of probabilities, conditional and non-conditional. In other formulations, there are only conditional ones. Non-conditional probabilities can be written as $p(A)$ and conditional probabilities as $p(A|B)$. We shall consider all probabilities conditional. Probabilities are functions that map to $[0,1]$, and are construed in BCT as degrees of belief. Their inputs A and B are construed in various ways. They can be construed as events, sets or sentences expressing propositions.

If logic is used, it is equipped with the basic conjunction (AB), (inclusive) disjunction ($A+B$), and negation ($\overline{A}$). If one uses sets, set theoretic equivalents are intersection, union and complement respectively.

The probabilities are constrained by various axioms or rules that formalise rationality constraints. Two, which we call the product and sum rules are pretty universal. For any A, B and C,

$$p(AB|C) = p(A|C)p(B|AC) = p(B|C)p(A|BC),$$

and

$$p(A|B) + p(\overline{A}|B) = 1.$$

The principle of total probability is used a lot as a rule. It states that if a set of sentences/sets $\{X_i|i\}$ is exclusive and exhaustive, then our probabilities can be written as weighted sums

$$p(A|B) = \sum_i p(X_i|B)p(A|X_iB).$$

We shall see how, with the right formulation, it can just be understood as an application of the product and sum rules.

Mirroring a conception of science, BCT is often conceived in terms of a number of hypotheses $\{H_i|i\}$ for which some data D is collected in order to provide evidence for or against. There are also contextual assumptions Y which contain things like auxillary hypotheses.

The hypotheses in $\{H_i|i\}$ are assumed to be exclusive and exhaustive. From this and axioms like the principle of total probability we get a couple of equations. We have

$$p(E|DY) = \sum_i p(H_i|DY)p(E|H_iDY), \qquad (1)$$

i.e., predictions of E are equal to a weighted sum of predictions one would make assuming individual hypotheses. Secondly, we get Bayes' theorem with the space of hypotheses

$$p(H_i|DY) = \frac{p(H_i|Y)p(D|H_iY)}{\sum_j p(H_j|Y)p(D|H_jY)}, \qquad (2)$$

where a probability of the form $p(D|H_jY)$ is generally called a likelihood of H_j. It is a measure of how well H_j predicts the data D. A probability of the form $p(H_j|Y)$ is generally called a prior–the degree of belief in H_j prior to collection of the data.

Equation (2) generally represents an ideal situation when one is aware of all possible hypotheses. This is often not the case; noone was aware of General Relativity (GR) until Einstein presented it. To deal with this, science construed through BCT often focusses on comparisons between known hypotheses; we may ask, did data D support H_i compared to H_j? A useful measure of this is the odds

$$\Omega_{ij}(D) \doteqdot p(H_j|DY)/p(H_i|DY) = L_{ij}Q_{ij},$$

where L_{ij} is the ratio of likelihoods and Q_{ij} is the ratio of priors,

$$L_{ij} \doteqdot \frac{p(D|H_jY)}{p(D|H_iY)} \;,\; Q_{ij} \doteqdot \frac{p(H_j|Y)}{p(H_i|Y)}.$$

The change in odds is due to only the ratio of likelihoods. If H_i predicts D better than H_j, then $L_{ij} < 1$, decreasing the odds of H_j over H_i. If H_i predicts D worse than H_j, then $L_{ij} > 1$, increasing the odds. Equation (2) can be rewritten in terms of odds, showing how to connect the concept with the probability of a hypothesis:

$$p(H_i|DY) = (1 + \sum_{j \neq i} \Omega_{ij})^{-1} = (1 + \sum_{j \neq i} L_{ij}Q_{ij})^{-1}.$$

For any H_i, the hypothesis may predict the data better or worse relative the the others, increasing or decreasing the

odds. A net increase in the sum of the odds decreases $p(H_i|DY)$ and vice versa. A low likelihood for H_i makes it easier for alternative hypotheses to predict D better, making it easier for $p(H_i|DY) < p(H_i|Y)$, though doesn't guarantee it.

When determining a probability of say, A, we consider two things. Firstly, we may need assumptions B such that we say the probability we are determining is $p(A|B)$. Using A and B, we may use our rules to constrain this probability as one equal to a function of other probabilities, as done in equations (1) and (2). Secondly, some of these other probabilities are not constrained by rules and hence must be assigned through other means. Generally this involves introspection, perhaps supplemented with certain techniques such as the principle of insufficient reason or the principle of maximum entropy etc.

Priors are the probabilities most in need of assignment though sometimes likelihoods require it too.

The strength of BCT is its simplicity and explanatory force in many nuanced real world situations (Jaynes, 2003). Calculations done with the theory seems to conform very well to nuanced scientific reasoning. That said, different formulations of BCT and different examples of calculations using different techniques to assign probabilities have a larger scope for debate.

Pursuant to a normative characterisation, there is one issue (above others (Earman, 1992)) I want to focus on. The probability $p(A|B)$ is considered the degree of belief someone (perhaps counterfactual) has in A. How, and to what extent, is someone licensed to have such a degree of belief?

I am asking this question in a certain way that differs from others. One view is to consider BCT as a tool for helping real people make precise inductive arguments and decisions. On this view, ought implies can, so any formulation that requires superhuman mental feats is not appropriate. This is not the kind of normativity I am asking about. My concern is what perfect rationality might look like. For example, it is impossible for a person to take into account all possible hypotheses but it is still considered ideal if they could.

We must consider in what way both inputs– assumptions B and the assigned probabilities–are licensed. I have already argued that one is not ultimately justified in ones assumptions but there can be distinctions of justification. In particular, the theory needs to work when a certain ideal is upheld:

(I) One should be as non-dogmatic as possible.

A precise formulation of this ideal is difficult for at least two reasons. Firstly, any formulation can depend upon the original formulation of BCT, whether and what logic is used, how its structure is understood etc. For example, is mathematics to be considered analytic? Is such a choice already dogmatic? Secondly, perhaps the best measure of non-dogmatism–entropy over probability distributions–has limitations such as, for infinite spaces of possibilities, many distributions of probabilities have infinite entropy and thus can't be distinguished. Let's then consider the notion informally, trusting our intuitions to guide us. In standard formulations of BCT, any assumptions preclude possibilities and are hence at least a little dogmatic. Informally, assumptions can be more or less dogmatic, the more substantive the assumptions, the more dogmatic. Perhap a little dogmatism cannot be escaped, but it should be minimal. Assignments of probabilities can also be seen as substantive/dogmatic if some probabilities are seen to be unfairly assigned higher values than certain others.

Taking (I) seriously, we run into the following problem. Every use of BCT in real world scientific situations makes substantive assumptions about the world or use substantive assigned probabilities. Moreover, such substantive assumptions often lead to strong predictions. In order for BCT to satisfy the ideal of non-dogmatism but also be explanatory in the many scientific situations, some way must be found to explain how one can end up with substantive probabilities about the world while starting from very non-dogmatic assumptions, assigned probabilities and realistic evidence.

One possible path to a solution is that substantive assumptions and assigned probabilities are due to purely pragmatic (James, 1897) reasons; that ones degrees of

belief can be due to previous licensed actions. Then, to be a solution, one must consider assigned probabilities coresponding to such degrees of belief as licensed. I will argue that reasons similar in sentiment to the pragmatic ones play a crucial role, though the form of pragmatic reasoning differs fundamentally from common interpretations.

Let's consider the reasons given for choice of assumptions and assigned probabilities in scientific scenarios. Generally, assigned probabilities are given informal plausibility arguments based on assumptions not made in the original determination of $p(A|B)$. For example, if A is a potential final state of some physical system in an experiment, one may assign a probability distribution over initial states, this distribution informally argued as appropriate given ones understanding of the experimental setup–the physics involved, how well calibrated the equipment is etc. Without such plausibility arguments, the choice of assumptions B and assigned probabilities are not considered very licensed. We require an explanation and formalisation of such arguments. The seemingly reasonable answer is that BCT is just such a formalisation of plausibility arguments so we should use the same tools for assigned probabilities as we used for $p(A|B)$; the assigned probability is no longer just assigned but given as a function of other probabilities. The problem with this approach is that we get what Suppe (1989) calls a vicious regress. In trying to replace the assigned probabilities with ones that aren't, we just end up with more.

This regress may not be too much of a problem if it can be shown how, starting from a non-dogmatic position and using only Bayesian reasoning, the addition of evidence leads one to make strong predictions about future observations. Past evidence D affects predictions of future evidence E as given by equation (1). In order for $p(E|DY)$ to be large, the probabilities for hypotheses that strongly predict E must dominate over those that don't. As discussed, $p(H_i|DY)$ can be understood in terms of odds between the possible hypotheses. The odds between two hypotheses changing based on the ratio of their likelihoods; if two hypotheses are observationally distinguishable, their odds change. If the likelihood of H_i is small enough relative to likelihoods of enough other hypotheses, the probability $p(H_i|DY)$ will be lower than $p(H_i|Y)$. If $p(H_i|DY)$ is small enough, H_i can be said to be effectively refuted (for the time being). Thus, given some D that effectively refutes the right hypotheses, strong predictions for future observations may occur.

In this vein, one may try to point to a convergence theorem by Hawthorne (2004, 2014). The theorem makes much weaker (more realistic) assumptions than other convergence theorems like those of Gaifman and Snir (1982). A great virtue of Hawthorne's theorem over Gaifman and Snir's is that it is *not* a theorem of convergence to certainty. I assert that one should not expect certainty, even in the long run; any such result should be seen as a symptom of some dogmatism. If so, such theorems should count against any theory with such a result.

Hawthorne's theorem shows roughly that, given that some hypothesis H_k is true, it is probable in the medium term that one will observe D such that the likelihood ratio L_{kj} is small for hypotheses H_j that are observationally distinguishable to H_k by D. In this way, in the medium term it is probable that one gets evidence that makes it easy to effectively refute some observationally distinguishable hypotheses. There are some caveats. Firstly, one must be in the right situations to get observational distinction; this is why scientists must set up experiments to get the right observations. Just sitting in your living room may not be enough. Secondly, effective refutation of hypotheses becomes easier but not guaranteed. There is however a bigger problem that makes the theorem insufficient to explain why one should be licensed to make strong predictions of future observations.

Consider a Bayesian agent in a medium term position; they have gathered some evidence D about the world and some hypotheses that are observationally distinguishable from the others have been given low probabilities. In order to make strong predictions for any particular future observation E, the probabilities for hypotheses that strongly predict E must be large relative to the others; they must be distinguished in some way. This distinction is not made by the effective refutation of hypotheses whose predictions of past observations differ greatly from what was observed; in principle one can imagine hypotheses that strongly predict D (and are hence not

effectively refuted) but make any predictions of the future one may like. It is completely possible to have effectively refuted many hypotheses and still be unable to make strong predictions.

To make this case stronger, consider a (finite) exclusive and exhaustive set of potential future observations $\mathscr{E} \doteq \{E_i | i\}$. We can measure the strength of the agent's predictions with the entropy over the distribution of probabilities of E_i:

$$S(\mathscr{E}) \doteq -\sum_i p(E_i | DY) \log_2 p(E_i | DY),$$

where base 2 was chosen for the log. The strongest prediction where $p(E_i | DY) = 1$ for some i corresponds to a minimal entropy of 0. There is a useful inequality

$$S(\mathscr{E}) \geq \sum_j p(H_j | DY) S(\mathscr{E}; H_j),$$

where

$$S(\mathscr{E}; H_j) \doteq -\sum_i p(E_i | H_j DY) \log_2 p(E_i | H_j DY),$$

is the entropy of the predictions of future evidence assuming H_j. In order for the agent's predictions to be strong (i.e., for $S(E)$ to be small), one needs the probabilities for hypotheses with weak predictions to be small. But effective refutation of such hypotheses is only determined by the hypotheses' predictions of past data D.

One needs a mechanism to probabilistically distinguish hypotheses that make the same prediction for D but different predictions for E. Current probabilities of hypotheses $p(H_i | DY)$ are dependent upon only the various hypotheses' predictions of past evidence D and their assigned priors $p(H_i | Y)$, i.e., any pair of hypotheses that make the same predictions of D can only have probabilities with different values if their initial assigned priors have different values. BCT alone gives no guidance as to choice of assigned priors and without this guidance cannot solve our problem. Despite the success of BCT, we still don't have a good answer to why evidence of what has happened should tell us anything about what will happen.

If we want our initial assumptions to be non-dogmatic, this problem becomes stark. We need a good reason to assign a different initial prior to one hypothesis that isn't effectively refuted compared to other hypotheses that aren't effectively refuted. It is often argued that some hypotheses have theoretical virtues that distinguish them from others such as simplicity and explanatory force. Indeed, I have made this argument for BCT.

The theoretical virtues are in need of explanation[6]. To do this, one needs a story as to the nature and structure of hypotheses which BCT does not provide. The story that BCT does provide–that hypotheses are exclusive and exhaustive–is simplistic and moreover, is dogmatic! I consider exclusivity and exhaustivity to be substantive assumptions about the world.

Instead of following the vicious regress of BCT, the theory presented in this series of papers will explain the licensing of probabilities approximating Objective Bayesian (Jaynes, 2003) ones in appropriate scientific contexts through a combination of factors and will also explain the plausibility arguments used: firstly, decision theory will be modified due to considerations independent of the above concerns. Secondly, hypotheses will be defined to not be reducible to a single exclusive and exhaustive space of atomic propositions, or similarly, possible worlds. Moreover, exclusivity and exhaustivity will not be assumed but rather emerge. From these factors, the theory will present a way to start from very minimal assumptions but act as if substantive assumptions about the world are true. The structure of hypotheses and the relationships between them will provide an explanation of explanatory force and potentially provides a general explanation of simplicity.

---

[6]For various reasons that I can properly argue only after more is said in following papers, I do not consider Solomonoff induction (Solomonoff, 1964) and the notion of Kolmogorov complexity (Kolmogorov, 1998) to be a good explanation of the theoretical virtue of simplicity. The reasons relate to how one understands the relationship between strings of signs and meaning relevant to inference associated with them.

4 Pragmatism

The conceptions of the world we use to guide our actions are at least partly chosen by what we value. Moreover, there is normative force to such choices. In this I am expressing a certain pragmatic sentiment (James, 1897). The problem of induction can be understood in this light with three consequences: (1) The basis for the reason to accept certain inferences can be, to a certain extent that we prefer them that way; the craving for justification lessens. (2) Preferences are individual. Thus different individuals can in principle appropriately use different standards of inference; intersubjective agreement between individuals is incidental. (3) The solution to the characterisation problem will include a value component. In this way, we shall move from finding standards of inference to finding standards of action choice–a decision theory–where a value component comes into play. This general idea isn't new (Rudner, 1953; Savage, 1972; Jeffrey, 1956) but we shall take it in a new direction.

Consider how decision theory is generally conceived (Savage, 1972; Jeffrey, 1956)–bracketing out specific details of formulation such as the conception of probability used and the conception of what the probabilities are of: we can imagine an agent that has preferences and makes choices of action in some normatively appropriate way. The appropriate action is found by maximising the expected utility

$$e(A_i) \doteq \sum_{j=1}^{n} U(Z_j A_i) p(Z_j | A_i Y), \qquad (3)$$

where U is a utility function determined by the agent's preferences, the set $\{Z_j | j = 1, \ldots, n\}$ is the set of some exclusive and exhaustive outcomes, $\{A_i | j = 1, \ldots, m\}$ is the set of actions to choose from.

Normally in decision theory we pick out only a small subset of outcomes to consider; in deciding whether to buy a toaster, one might consider specifically how much it costs and how much value you expect to get from eating toast. However, in principle buying the toaster may have myriad secondary effects such as effects on your weight and health, whether the toaster looks good in your home etc. Thus, as we aim for a decision theory that aims for universality, we shall consider the theory to take into account all possible outcomes, i.e., the set of outcomes

is also considered *complete*. A precise definition of completeness of a set of outcomes will not be given until we define the logic used later. However, the currently vague notion is still intuitively useful.

Pascal's wager is one of the best known early examples of decision theoretic arguments for choice of action. It and similar ones are used to justify acts that are conjectured to lead to belief (James, 1897). Is it then the case that certain degrees of belief formalised in terms of initial assigned Bayesian probabilities can be justified due to the beliefs being a consequence of justified actions? I shall not use this argument.

Instead, by looking at Pascal's wager, we may come up with a different approach that outlines a solution to the problems considered in the previous section. A decision theoretic formalisation of Pascal's wager asks us to consider the utility of the existence of god. Such a formalisation treats the *hypothesis* that god exists the same as if it were an *outcome* like the result of a roll of the die or the state of the weather. What if there is a functional distinction to be made between these two categories? Consider the following argument schema with premises (A.1)-(A.3):

(A.1) There are various hypotheses about the world.

With this premise I am not talking about the role of non-observational terms in inference. Rather, I am discussing the nature of hypotheses: one can understand hypotheses in terms of a gap between possible epistemic positions and the world. For example, there is epistemic modality and metaphysical modality; the metaphysical kind associated with a hypothesis which distinguishes one or more epistemic positions over others in some way (that will be made clearer later). Hypotheses relate to the context around various outcomes. To say that there are hypotheses in the theory is to say that there is a functional distinction at the level of standards.

(A.2) Outcomes from different hypotheses are different.

Take as an example outcomes as orbits of planets. There are different hypotheses that can predict them such as Newtonian Gravity (NG) and General Relativity (GR). To compare their predictions there is a way to align them; to

8

be able to say that their predictions in some circumstances are very similar or not. But importantly, this premise claims that they are still different, even if in some cases they are perfectly aligned. We shall cash this premise out with

(A.3) Values associated with outcomes are dependent on the hypothesis they come from.

Often in a scientific context, we don't distinguish the value we see in the world by whether we see it in terms of say, NG and GR. In fact this is one important characteristic of science. But this doesn't mean value can't be distinguished this way. The outcomes of NG are distinct from the outcomes of GR and from this I argue there is no in principle reason why the values associated with their outcomes should be aligned in any way. A potential counterargument is that if meaning is defined instrumentally–that if both hypotheses are couched in a neutral observation language–then the aligned outcomes can mean the same thing. We shall not go this route; meaning will not be defined instrumentally.

Taking (A.1)-(A.3) as premises, conclude:

(A.4) In the situation where one is uncertain about hypotheses, one needs a decision theory that is sensitive to the hypothesis dependence of value, distinct from the outcome dependence of value.

Consider that equation (3) sums over utilities of a complete set of possible outcomes. If there is more than one possible hypothesis, then there is a set of outcomes for each hypothesis that should be taken into account. Consider then the following modification to (3) as a schema[7] for a new decision theory:

$$e'(A_i) \doteq \sum_{k=1}^{m} \sum_{j=1}^{n(k)} U(Z_j^k A_i, H_k') p(Z_j^k H_k' | A_i Y)$$
$$= \sum_{k=1}^{m} p(H_k' | A_i Y) \sum_{j=1}^{n(k)} U(Z_j^k A_i, H_k') p(Z_j^k | H_k' A_i Y),$$
(4)

where $\{H_k' | k = 1, \ldots, m\}$ is the set of possible hypotheses.

Compare $e'(A_i)$ with

$$e(A_i) = \sum_{j=1}^{n} U(Z_j A_i) \sum_{r=1}^{s} p(H_r | A_i Y) p(Z_j | H_r A_i Y),$$
(5)

where we have assumed a set of exclusive and exhaustive hypotheses $\{H_r | k = 1, \ldots, s\}$. The salient difference is that the utilities in (4) are explicitly hypothesis dependent while the ones in (3) aren't. Moreover, utilities $U(Z_j A_i)$ could be made to be expressed with compatible hypotheses like $U(Z_j A_i H_r)$ but they will not be under the sum over hypotheses. Such functional differences lead to important consequences.

We can imagine a situation similar to one in Pascal's arguments for wagering for god. I am not going to argue that Pascal's wager is reasonable, merely that there are extreme situations where something like it is. Suppose that H_l' is a hypothesis where god exists and for every other hypothesis this isn't the case. Then suppose that this corresponds to utilities such that

$$\forall k \neq l \ \forall j, i \ \ U(Z_j^k A_i, H_k') = 0,$$

i.e., nothing in the world has value unless god exists. If this is so, then we have

$$e'(A_i) = p(H_l' | A_i Y) \sum_{j=1}^{n(k)} U(Z_j^l A_i, H_l') p(Z_j^l | H_l' A_i Y).$$

If–as is common–A_i is independent to the hypothesis–in this case, H_l'–and the probability for H_l' is non-zero, then $p(H_l' | A_i Y) = p(H_l' | Y)$ such that the model e' becomes functionally equivalent to standard decision theory in the situation where it is assumed that god exists, i.e., one should act as if god exists[8].

The relevant aspect of this situation for us is that preferences can directly force effective assumptions without needing belief; one can be licensed to act as if something were true without believing it to be true. This is different to the argument where one is licensed to act where belief is a consequence and then one uses those beliefs to assign probabilities.

---

[7] I say schema because without an appropriate formulation of the logic used, which shall be presented in sections after this, there are subtle problems one can encounter in this version of the theory.

[8] If H_l' is such that, if it were true then it is strongly preferable to believe in god, then it is preferable to find a way to believe in god, i.e., to wager for it.

The above mechanism is also different from a certain notion of acceptance in practical reasoning due to Bratman (1992). With his notion the focus is on the practicality of (what can be modelled as) certain approximations of standard decision theory as used as a mental tool. Our notion of effective assumptions is not due to an approximation of decision theory but rather a reformulation of it.

Consider (4) generally. The expected utility e' is a weighted sum of the expected utilities one gets upon assumption of a hypothesis–call these *subexpected utilities*. The weights given by the probabilities of each hypothesis, which can be viewed as logical components of utility.

We must distinguish between initial assumptions (in the above case, Y) and *effective assumptions* (in the above case $H_l'Y$). The above case is extreme, where $H_l'Y$ completely dominates. Generally, effective assumptions may be considered more or less significant based on a few factors, including the weightings of the subexpected utilities of such assumptions relative to the others. When using introspection to assign a Bayesian probability, I assert that we are often considering a probability relative to some significant effective assumptions as opposed to relative to ones initial assumptions. There is a disconnect between probabilities arising from ones initial assumptions and ones that are intuitively assigned.

In this way we shall be able to make very weak initial assumptions–where inductive inference alone is impossible–while at the same time make strong effective assumptions. The model (4) is not however sufficient for this task. It shall be built upon with a more nuanced formulation of the structure of hypotheses, motivated from various directions, from a theory of meaning that will be developed, to the difficulties with counterfactual conditionals (Goodman, 1983) which shall be connected to the reductionism that has influenced formulations of probability theory and decision theory since at least Wittgenstein (1994).

Hypotheses in (4) can be understood as providing a context of meaning for outcomes. This notion will naturally extend to another: hypotheses can provide a context of meaning for other hypotheses. As a general example, consider (a version of) Classical Mechanics as providing a context of meaning for specific possible hypotheses of classical physics such as Electrodynamics or Fluid Mechanics. Moreover, Classical Mechanics can be seen as stemming from a conjunction of more abstract hypotheses such as mathematical and metaphysical ones.

We will have a hierarchy of abstraction for hypotheses. We shall be lead to a situation where multiple higher level hypotheses can form a context for a single lower level one, i.e., there are many different paths to forming a context. Hypotheses will be built from parts. Each part a higher level hypothesis which can correspond to a description of some aspect of the lower level hypothesis. There will be many paths to building a hypothesis, each path corresponding to a different collection of parts. For example, Newtonian Mechanics can be formulated in terms of forces or in terms of a minimisation of action, the resultant hypotheses being the same but the paths used to get there different. Other examples include different ways to formulate equivalent geometries, or different ways to formulate equivalent theories about numbers.

From a higher level of abstraction, there will in principle be multiple ways or paths to arrive at a hypothesis. Each path will naturally contribute to the expected utility by increasing the weight of the subexpected utilities for each hypothesis. A hypothesis that has low probability in each path but many paths will be able to dominate in significance to decision choice due to combinatoric reasons. This combinatoric effect will be related to the logical structure of the hypothesis. It is in this way the problems of BCT discussed in the previous section is solved: differences in the logical structures of hypotheses, separate to their predictions of past data, lead to some hypotheses dominating over others.

I conjecture that this combinatoric mechanism is the origin of the notion of simplicity.

Plausibility arguments can be understood with the hierarchical web of hypotheses. In cases where subexpected utilities of some effective assumptions dominate, instead of assigned probabilities, we have effective probabilities due to effective assumptions; their significance determined by the number and significance of more abstract

effective assumptions. In such cases, one can say the theory approximately aligns with an Objective Bayesian one. Supppose one has *X* as an effective assumption. One can give a plausibility argument from a more abstract effective assumption *Y* where *X* is not assumed, i.e., one may argue that if one assumes *Y*, then *X* is probable (and implicitly that *Y* is a significant assumption). A plausibility argument will correspond to a segment of a path from a higher level of abstraction to a lower one.

Probabilities will be argued to be uniquely determined by assumptions. The notion of degrees of belief will be made redundant; there is no single probability an agent assigns to any *X*, only a unique probability of *X* for any assumed *Y* but with many assumptions contributing to action choice. Our notion of probability will be logical rather than Bayesian, though a version of Bayes' theorem and the principle of total probability will drop out from first principles.

There are arguments against theories of logical probabilities that are considered fatal: Goodman's 'grue' argument and the argument against the uniqueness of priors through redescription. These problems will be dissolved due to a reformulation of logic and its relationship to language. Other problems like the zero prior problem will be similarly dealt with.

Logic will be considered non-linguistic: it is subsidiary to a normative decision theory for an agent that need not communicate to anyone or even use symbolic tools for internal reasoning. Propositions are defined to not have form.

References

Bratman, M. E. (1992). Practical reasoning and acceptance in a context. *Mind*, 101:401.

Carnap, R. (1991). Empiricism, semantics, and ontology. In *The philosophy of science*, pages 85–98.

Earman, J. (1992). *Bayes or Bust? A critical examination of Bayesian coniﬅrmation theory*. Cambridge, ma: MIT Press.

Fuchs, C. A. (2016). On participatory realism. *arXiv preprint arXiv:1601.04360*.

Gaifman, H. and Snir, M. (1982). Probabilities over rich languages, testing and randomness. *Journal of Symbolic Logic*, 47(3).

Goodman, N. (1983). *Fact, Fiction and Forecast*. Harvard University Press.

Hawthorne, J. (2004). A better bayesian convergence theorem. In *First Annual Austin-Berkeley Formal Epistemology Workshop, FEW*, pages 21–23.

Hawthorne, J. (2014). Inductive logic.

Hume, D. (1748 [1910]). *An Enquiry concerning Human Understanding*. P. F. Collier and Son.

James, W. (1897). *The Will to Believe and Other Essays*. London: Longmans, Green, and Co.

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

Jeffrey, R. C. (1956). Valuation and acceptance of scientific hypotheses. *Philosophy of Science*, 23(3):237–246.

Kolmogorov, A. (1998). On tables of random numbers. *Theoretical Computer Science*, 207(2):387âĂŞ395.

Nietzsche, F. (2003). *Beyond Good and Evil*. Penguin. pp. 37.

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of science*, 20(1):1–6.

Savage, L. J. (1972). *The foundations of statistics*. Courier Corporation.

Solomonoff, R. (1964). A formal theory of inductive inference parts i/ii. *Information and Control*, 7:1–22 + 224–254.

Suppe, F. (1989). *The semantic conception of theories and scientific realism*. University of Illinois Press. pp. 399.

Wittgenstein, L. (1994). *Tractatus Logico-Philosophicus*.
  Edusp.