Formal Methods
for *Cambridge Handbook of Analytic Philosophy*

Richard Pettigrew
Richard.Pettigrew@bris.ac.uk

October 27, 2020

It is not uncommon to open an analytic philosophy paper written during the past century and find logical symbols and mathematical notation among the prose. Sometimes these are used rather lightly, perhaps in setting out the logical form of an argument; sometimes they play a more central role, perhaps stating a mathematical theorem that features as a key premise in the argument. Why do analytic philosophers do this? They study the nature of knowledge and beauty and truth, the analysis of language, the most just way to organise our society, the best way to live our lives, the existence of God, the fundamental nature of reality, among many other things. How do such formal, mathematical methods help them to investigate these questions? In this chapter, I'd like to persuade you that they help in many different ways. They allow us to analyse and scrutinise philosophical theses and arguments, perhaps disambiguating a premise or conclusion that is unclear in its natural language formulation, perhaps rendering the argument in a particular logic and thereby either allowing us to identify missing premises, assumptions that have been smuggled in, or even fallacies that the argument commits, or else assuring us that the conclusion truly follows only from the assumptions explicitly stated and does not require anything further (§1-4). They sometimes allow us to establish that, while we would like our account of some topic of interest to boast a range different features, it is in fact mathematically impossible for an account to have all of those features at once, and so we must settle for second-best; and indeed they sometimes help us investigate the options for second-best (§5). They sometimes allow us to discover new arguments for our conclusions, and in particular better arguments that appeal to weaker, more plausible assumptions in their premises; sometimes this is because we can appeal to a surprising and powerful mathematical theorem to boost those seemingly weak premises to establish a powerful conclusion (§8), and sometimes it is because we can use computer software to search the space of possible arguments for our conclusions to find ones that assume less than

we've been assuming so far (§1). They sometimes allow us to run computer simulations of idealised mathematical models of the situations that our arguments treat to see exactly what sort of consequences we should expect to follow from the claims we make about those situations (§7). Sometimes they help us to better understand a phenomenon, perhaps by showing that it shares a mathematical structure with another phenomenon, sometimes by giving an idealised model of it (§6). And sometimes they are necessary even to represent the phenomenon with which our arguments are concerned (§2-3).

In short: the uses of logical and mathematical methods in philosophy are multifarious. And so too are the areas in philosophy that deploy these methods. In this chapter alone, my examples draw from philosophical discussions of democratic theory, the existence of God, the foundations of mathematics, the oppression of minority groups, and the epistemology of uncertain belief, and the semantics of natural language. Let's get started.

1 Deductive arguments in philosophy

Two of the most familiar uses of formal methods in philosophy are the use of formal logic and probability theory to study reasoning. Let's take them in turn.

Early in our philosophical education, we learn to wrangle arguments into a list of premises followed by the conclusion that those premises seek to establish. We also learn how to symbolise those premises and the conclusion in propositional logic and first-order logic.[1] Later, we might learn how to translate into logics that have more expressive power, such as temporal, modal, epistemic, or deontic logic, as well as the first- and higher-order versions of these. Why do we do this?

There are two reasons to formalise an argument. First, rendering propositions in their logical form goes some way to making them precise enough for us to scrutinise them; second, considering the logical relationships between the premises and between the premises and the conclusion goes some way to revealing the inferential power of the argument.

An elegant example of both of these comes from St Thomas Aquinas' argument for the existence of God in the third of his Five Ways. Here is a crucial step of the argument:

> [I]f everything is possible not to be, then at one time there was nothing in reality. (As translated in (Anders, 2012).)

It is natural to render the argument of this sentence in first-order logical form as follows, where Cx symbolises *x is contingent* and Ext symbolises x

[1] Introductions to formal logic include (Priest, 2001; Hodges, 2001; Restall, 2004; Magnus et al., 2019).

exists at t.

(TW1) $\forall x(Cx \rightarrow \exists t \neg Ext)$

Therefore,

(TWC) $\exists t \forall x(Cx \rightarrow \neg Ext)$

But this argument form is invalid. There are countermodels in which the premise is true and the conclusion false. We might, for instance, let the domain of quantification be the set of natural numbers (that is, 0, 1, 2, 3, ...), let the extension of C be the whole domain, and let E be the relation that one number bears to another if the first is greater than the second. Then (TW1) is true: for all natural numbers, there is a natural number that is greater than it. But the conclusion is false: there is no natural number that is greater than all others.

By formalising the argument in this way, we learn something of its inferential force. But it's important not to overstate what we learn. When we discover that a given argument is not logically valid when formalised in a particular logic, we do not thereby learn that the premises of the original argument do not support its conclusion. We simply learn that they do not guarantee the conclusion in virtue of their logical form as represented in that logic. It's possible that they do guarantee the conclusion in virtue of their logical form as represented in another logic—for instance, an argument whose translation into propositional logic is invalid might have a translation into first-order logic that is valid. And it's possible that they do not guarantee the conclusion in virtue of their logical form in any logic, but nonetheless do support the conclusion in some weaker way—for instance, they might make it likely to be true, for instance.

Nonetheless, formalisation is a powerful tool. We might use it to discover what premises we'd have to add to an argument in order to make it logically valid in that logic, and then ask whether they are plausible. Or we might consider countermodels of the logical form of the argument when it is translated into its most natural logic to understand how the world would have to be for the premises to be true and the conclusion false. So, for instance, we might be inspired by the natural numbers countermodel to the formalization of Aquinas's argument from above to describe a universe in which Aquinas' original premise is true but his conclusion is false.

Formalisation requires us to render the premises and conclusion of our argument in logical form. In the previous example, that was straightforward and a means to the end of analysing the argument's inferential strength. But sometimes the formalization itself is illuminating. Take, for instance, a different argument for the existence of God. This time, St Anselm's ontological argument from his *Proslogion*. The American philosopher David

Lewis (1970) analyses the argument into four premises and a conclusion. Here's the third premise:

(OA3) Something exists in the understanding than which nothing greater is possible.

He formalizes this premise along with the other three. They all involve quantification over things, but also modal notions, such as possibility. So it might seem most natural to formalize it in modal logic. But Lewis thinks it is more perspicuous to formalise the argument using quantification over possible worlds, which in any case is a standard way of giving the meaning of modal notions. However, while he manages this easily for the other three premises, when he tries to do it for (OA3), he realises that this premise is ambiguous. That is, we can make it precise using quantification over possible worlds in two different ways.

Here's the first:

(OA3_1^*) There is an understandable thing x such that there is no world w and thing y such that y is greater at w than x is at the actual world.

If we render (OA3) in this way, then, when combined with his formalizations of the other premises, the argument is valid. But the premise understood in this way is implausible. Why should we think that there is any understandable thing that is at its very best in the actual world?

Here's the second reading of (OA3):

(OA3_2^*) There is an understandable thing x and a possible world v such that there is no world w and thing y such that y is greater at w than x is at v.

If we disambiguate (OA3) in this way, the premise is much more plausible. It is plausible that there is an understandable thing that, in the possible world in which it is at its best, is better than anything else at any other world. But, when combined with the formalizations of the other premises, the argument is invalid.

By excavating the logical form of St Anselm's argument using quantification over possible worlds, David Lewis unearthed an ambiguity in one of the premises. Until we resolve that ambiguity, we've no way to assess whether the argument is valid or invalid, sound or unsound. And indeed when Lewis does the work of disambiguating the premise, he discovers that the argument must fail in one way or the other: with (OA3_1^*), it is valid but unsound; with (OA3_2^*), it is invalid, but sound.

So, formalising an argument can bring greater precision to the claims that constitute it, help identify the commitments of the argument, identify missing premises, and evaluate the inferential power of the argument.

It can also put your argument into the form where we can investigate it using a theorem prover, which is a piece of computer software that assists with some aspect of logical reasoning.[2] For instance, a theorem prover might search for a deduction from your premises to your conclusion, or even from some weaker set of premises to your conclusion, thereby strengthening your argument. Or, it might search for countermodels of the argument to show it's invalid, or for models of your premises to ensure that they're consistent. Different theorem provers specialise in finding proofs or models in different logics. They have been used in philosophy to help understand the Austrian logician Kurt Gödel's formulation of the ontological argument for the existence of God. Gödel didn't publish his proof himself, but showed it to the American logician Dana Scott, who fixed an inconsistency in Gödel's original premises (Gödel, 2004; Scott, 2004). Of course, if we're working in a logic that is complex enough that the great Gödel didn't spot an inconsistency, we might be worried that Scott's alternative premises are also inconsistent. Using a theorem prover called `Nitpick`, Benzmüller & Paleo (2014) discovered a model for those premises, showing that they are consistent. And then they used two further theorem provers, `LEO-II` and `Satallax`, to show that the theorem does not require the strong background modal logic S5, which Gödel used, but only the much weaker system KB.

Branden Fitelson (2013) has also used the theorem provers `paradox`, `vampire`, `prover9`, and `otter` to strengthen a well-known argument by showing that much weaker premises are required. The argument is Allan Gibbard's Collapse Theorem (Gibbard, 1981). Gibbard showed that, from certain assumptions about the indicative conditional (i.e., the natural language connective in sentences of the form *If A is true, then B is true*), we can show that it must be logically equivalent to the material conditional (i.e., the logical connective symbolised as $A \rightarrow B$). But the assumptions are quite strong and it seems we can escape the undesirable conclusion by rejecting some of them. However, Fitelson shows that a much weaker set of assumptions ensures the same collapse result. There's no easy escape!

2 Probabilistic arguments in philosophy

Not all arguments seek to establish their conclusion definitively on the basis of their premises. Some seek only to make the conclusion likely. Formal logic is usually not the best way to study these arguments. It is often better to turn to the theory of probability. Staying in the philosophy of religion, let's consider an example.[3] Given the physical laws that govern our universe, there is a rather narrow range of values within which the fundamen-

[2] For an introduction to theorem provers and automated reasoning, see (Portoraro, 2019). For the original manifesto arguing for their use in philosophy, see (Fitelson & Zalta, 2007).

[3] For an introductory overview of this argument, see (Friederich, 2018).

tal constants must lie if the universe is to be hospitable to life. Let's symbolise that fact as F, for *fine-tuning*. But of course there is life. Let's symbolise that fact as L. Some philosophers wish to appeal to the conjunction LF to argue that God exists, which we'll symbolise as G. The argument proceeds via two probabilistic claims. The first is that, given F, L is very unlikely given $\neg G$. That's because, given F, there is a much larger range of values for the fundamental constants on which the universe doesn't support life than on which it does, and if God doesn't exist, any value is as likely as any other. The second claim is that, given F, L is very likely given G. That's because God presumably wants to create a universe that supports life, and is able to do what They want and so able to set the fundamental constants within the range that makes life possible.

I'll formalise this argument using the Bayesian approach.[4] The core Bayesian assumption is that we can measure the strength of someone's belief in each proposition they consider using a real number from 0 to 1—we call this strength of belief their *credence* in A. We record a person's credences in their *credence function*. If P is their credence function and A is a proposition they consider, $P(A)$ is their credence in A. If $P(A) = 1$, then the person has maximal credence in A; if $P(A) = 0$, they have minimal credence in A. What's more, $P(A|B)$ gives their conditional credence in A given B. The Bayesian assumes that a rational person's credence function P is a probability function. That is, P has the following properties:

(PA1a) If $\top$ is a tautology, $P(\top) = 1$;

(PA1b) If $\bot$ is a contradiction, $P(\bot) = 0$;

(PA2) If A and B are propositions, $P(A \vee B) = P(A) + P(B) - P(AB)$.

Various results follow from this—I'll mention just a couple:

(i) If A entails B, then $P(A) \leq P(B)$. For instance, $P(A) \leq P(A \vee B)$ and $P(AB) \leq P(A)$.

(ii) If A and B are logically equivalent, then $P(A) = P(B)$.

(iii) If A and B are logically inconsistent, then $P(A \vee B) = P(A) + P(B)$. For instance, $P(A) + P(\neg A) = 1$, so $P(A) = 0.5$ if you think A is as likely as not.

What's more, Bayesianism demands that a rational person's conditional credences are related to their unconditional credences by the so-called *ratio formula*:

[4]Introductions to Bayesianism and Bayesian epistemology include (Talbott, 2008; Titelbaum, ???; Easwaran, ???). The approach is named in honour of Reverend Thomas Bayes, the eighteenth century English Presbyterian minister who first stated its central result, Bayes' Theorem, which we will meet below.

(RF) $P(A|B) = \frac{P(AB)}{P(B)}$, provided $P(B) > 0$.

The most famous consequence of this is *Bayes' Theorem*, which underpins nearly all applications of Bayesian epistemology:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\neg A)P(B|\neg A)}$$

And, the final piece in the Bayesian jigsaw: the rule for updating your credences when new evidence comes in, which is sometimes called *Bayes' Rule* or *Bayesian Conditionalization*:

(BC) If P is your credence function at one time, P' is your credence function at a later time, and E is the strongest evidence you obtain between those times, then, for all propositions A, $P'(A) = P(A|E)$.

That is, your later unconditional credence in A should be your earlier conditional credence in A given E.

Before we move on to apply this framework, note that this is a case in which formal methods are required just to represent the phenomenon of interest, namely, a person's degrees of belief, let alone to investigate them. Each credence is represented by assigning a real number, itself a mathematical object, to a proposition. It is important to note that, while we are familiar with real numbers from high school, and while we use them to represent distances, weights, and so on throughout our everyday life, a lot of work is required in the background to ensure that such a representation is legitimate. This is the subject matter of *measurement theory*, which provides the foundations for representing quantities numerically (Tal, 2020, Section 3).

With all of this in hand, we can now formalise the probabilistic part of the fine-tuning argument from above:

(FT1) $P(L|\neg GF)$ is low

(FT2) $P(L|GF)$ is high

Therefore, by (FT2) and (FT3),

(FTC) $P(G|LF)$ is high.

We can then note that LF is true, and conclude that your credence in G should be high. Now, we might resist the argument by denying (FT1) or (FT2), but there is another objection to it: it is probabilistically invalid. The problem is that it is possible to have a probability function with the properties demanded by (FT1-2) that does not have the property demanded by the conclusion (FTC); so (FTC) does not follow from (FT1-2).

Essentially, this argument commits a version of what is sometimes called the *base rate fallacy* (Kahneman & Tversky, 1973; Bar-Hillel, 1980). The classic example of this fallacy occurs when you test positive for a rather unusual disease using a slightly imperfect test, and you become very confident that you have the disease on this basis. When you do this, you ignore the fact that the disease is very rare, which means that you are much more likely to be one of the healthy people who get a false positive test than you are to be one of the sick people who get a true positive test. The fine-tuning argument fails in the same way. Suppose that, given F, you just think that it's very unlikely that God exists. So, $P(G|F)$ is very low. Then $P(L|\neg GF)$ might be low and $P(L|GF)$ might be high, but $P(G|LF)$ is still low. We can see this by using Bayes' Theorem, which tells us that

$$P(G|LF) = \frac{P(G|F)P(L|GF)}{P(G|F)P(L|GF) + P(\neg G|F)P(L|\neg GF)}$$

Suppose, for instance, that

- $P(G|F) = 0.1\%$,
- $P(L|GF) = 99\%$, and
- $P(L|\neg GF) = 1\%$.

Then $P(G|LF) \approx 10\%$.

So, the Bayesian method of rendering arguments using subjective probabilities or credences allows us to assess their inferential power, and identify fallacies in reasoning.

3 Probabilistic arguments in science

Of course, philosophers do not only scrutinise their own arguments. In the philosophical study of academic disciplines like science, mathematics, psychology, history, etc. we are interested in evaluating the inferences that the practitioners of those disciplines use. And Bayesian methods can help here as well. Let's take a famous example from the philosophy of science.

A central question that arises in the philosophy of any science is what evidence confirms which hypotheses. The French philosopher Jean Nicod (1924) proposed the following principle, which identifies one type of evidence that confirms one sort of hypothesis:

(NC) Observing something that has property R and property B confirms the hypothesis that all Rs are Bs.

In symbols, $RaBa$ confirms $\forall x(Rx \rightarrow Bx)$.

However, the German philosopher Carl Hempel (1945) identified an unwelcome consequence of (NC) when it is combined with the following plausible claim:

(EC) If E confirms H, and H is logically equivalent to H', then E confirms H'.

First, note that $\forall x(Rx \rightarrow Bx)$ is logically equivalent to $\forall x(\neg Bx \rightarrow \neg Rx)$. Then note that, by (NC), $\neg Ra \neg Ba$ confirms $\forall x(\neg Bx \rightarrow \neg Rx)$. And finally combine these and note that, by the (EC), $\neg Ra \neg Ba$ confirms $\forall x(Rx \rightarrow Bx)$. And this, Hempel thought, must be wrong.

To illustrate the problem, he gave a famous example, which gives the problem its standard name, *the paradox of the ravens*. Let Rx symbolise *x is a raven* and let Bx symbolise *x is black*. In this case, (NC) tells us that seeing a black raven confirms that all ravens are black, which seems right. It also tells us that seeing a non-black non-raven confirms that all non-black things are non-ravens. And that also seems right. But, because the latter hypothesis is logically equivalent to the former, the (EC) tells us that seeing a non-black non-raven confirms that all ravens are black. Thus, when I see my red shoes, I confirm that all ravens are black; similarly when I see a yellow banana, a green leaf, and a scarlet ibis. And that seems wrong.

To state this paradox, it was helpful to use a little formal logic. But to solve it, we turn to the probability theory that we introduced in the previous section. The Bayesian solution to Hempel's paradox was given originally by the Polish logician Janina Hosiasson-Lindenbaum (1940). Using probability theory to explicate the notion of confirmation, and also to measure its strength, Hosiasson-Lindenbaum showed that, on reasonable assumptions, seeing a non-black non-raven does indeed confirm that all ravens are black, but only to a very small degree; while seeing a black raven confirms that all ravens are black to a much greater degree. Her solution thereby retains (NC) and (EC), but takes the sting out of their apparently paradoxical consequence. As we'll see, however, Hosiasson-Lindenbaum's approach also entails that seeing a black nonraven disconfirms the hypothesis that all ravens are black, and that creates a new paradox. We'll then meet Susanna Rinard's (2014) version of the Bayesian approach, which builds on Hosiasson-Lindenbaum's to solve this.

Probabilities were used early in the philosophy of science to understand the notion of confirmation. The Bayesian account is simple: evidence E confirms hypothesis H just in case $P(H|E) > P(H)$, where again $P(H)$ measures your confidence in H and $P(H|E)$ measures your confidence in H given E. That is, evidence confirms a hypothesis when the evidence raises the probability of the hypothesis. What's more, we can measure how strongly it confirms the hypothesis by looking at how much it raises the

probability. Now, recall Bayes' Theorem:

$$P(H|E) = P(H)\frac{P(E|H)}{P(E)}$$

Together with the Bayesian account of confirmation just given, it follows that E confirms H just in case $P(E|H) > P(E)$, and how strongly E confirms H is determined by how much greater $P(E|H)$ is than $P(E)$.

To see how Hosiasson-Lindenbaum's solution works, let's set out the relevant probabilities before and after we condition on H.

X	$P(X)$	$P(X\|H)$
$RaBa$	a	p
$Ra\neg Ba$	b	q
$\neg RaBa$	c	r
$\neg Ra\neg Ba$	d	s

What we want: first, $RaBa$ confirms H reasonably strongly; second, $\neg Ra\neg Ba$ confirms H only very weakly. As we've just seen, we'll get this just in case

$$\frac{P(RaBa|H)}{P(RaBa)} \gg \frac{P(\neg Ra\neg Ba|H)}{P(\neg Ra\neg Ba)} > 1$$

That is,

$$\frac{p}{a} \gg \frac{s}{d} > 1.$$

The usual Bayesian solution that follows Hosiasson-Lindenbaum's lead, arranges this by making four assumptions:

(a) Learning H lowers your probability for observing a non-black raven to zero. That is, $P(Ra\neg Ba|H) = 0$. That is, $q = 0$.

(b) Learning that all ravens are black wouldn't change how likely it is you'll observe a raven. That is, $P(Ra|H) = P(Ra)$. That is, $p + q = a + b$.

(c) Learning that all ravens are black wouldn't change how likely it is you'll observe something non-black. That is, $P(\neg Ba|H) = P(\neg Ba)$. That is, $q + s = b + d$.

(d) You're much less likely to pick a raven than something that isn't black. That is, $P(Ra) \ll P(\neg Ba)$. That is, $a + b \ll b + d$.

So, by (a-c), $p = a + b$ and $s = b + d$. So

$$\frac{P(RaBa|H)}{P(RaBa)} = \frac{p}{a} = \frac{a+b}{a} \quad \text{and} \quad \frac{P(\neg Ra\neg Ba|H)}{P(\neg Ra\neg Ba)} = \frac{s}{d} = \frac{b+d}{d}$$

And by (d), $a \ll d$, so

$$\frac{P(RaBa|H)}{P(RaBa)} = \frac{a+b}{a} \gg \frac{b+d}{d} = \frac{P(\neg Ra\neg Ba|H)}{P(\neg Ra\neg Ba)} > 1$$

And this is exactly what we wanted. However, (a-c) also entail something less welcome: since $p+q+r+s = 1 = a+b+c+d$, (a-c) give us, $(a+b)+r+(b+d) = a+b+c+d$, and so $r = c-b < c$. That is, $P(\neg RaBa|H) < P(\neg RaBa)$. So $P(H|\neg RaBa) < P(H)$. So observing something black that isn't a raven disconfirms the hypothesis that all ravens are black. Oh dear!

Fortunately, the American philosopher Susanna Rinard (2014) provides a way out. She retains (a), (b), and (d) from above, but she strengthens (c) as follows:

(c1) Learning that all ravens are black wouldn't change how likely it is you'll observe something black that isn't a raven. That is, $P(\neg RaBa|H) = P(\neg RaBa)$. That is, $r = c$.

(c2) Learning that all ravens are black wouldn't change how likely it is you'll observe something non-black that isn't a raven. That is, $P(\neg Ra\neg Ba|H) = P(\neg Ra\neg Ba)$. That is, $s = d$.

This then gives us:

$$\frac{P(RaBa|H)}{P(RaBa)} = \frac{p}{a} = \frac{a+b}{a} \quad \text{and} \quad \frac{P(\neg Ra\neg Ba|H)}{P(\neg Ra\neg Ba)} = \frac{s}{d} = \frac{d}{d} = 1$$

So $RaBa$ confirms H, $\neg Ra\neg Ba$ neither confirms nor disconfirms H, and so $RaBa$ confirms H more strongly than $\neg Ra\neg Ba$ does. And, what's more, $P(\neg RaBa|H) = P(\neg RaBa)$. Notice, of course, that this solution violates (NC), since $\neg Ra\neg Ba$ does not confirm H and therefore does not confirm $\forall x(\neg Bx \rightarrow \neg Rx)$. But perhaps this is an advantage. After all, it accords with our intuitive judgment.

So, rendering arguments using probabilities helps us to understand and justify reasoning that scientists routinely use, just as it helps us analyse and critique philosophical reasoning. In this way, it contributes to the epistemology of science.[5]

4 Deductive arguments in mathematics

Nowadays, philosophers use formal logic most often to analyse their own arguments. But the logic that we know as first-order logic was introduced by the German mathematician and philosopher Gottlob Frege in 1879 to

[5]For more on the Bayesian reconstruction of scientific reasoning, as well as some of the problems it faces, see (Earman, 1992; Sprenger & Hartmann, 2019).

investigate reasoning in mathematics (Frege, 1879). Like many mathematicians in the eighteenth and nineteenth centuries, Frege wanted to set his subject on a firmer foundation by expunging geometrical intuitions from its reasoning. To do this, Frege created a formal language in which to write mathematics, as well as a set of simple, mechanically-applicable rules for deriving one sentence in this language from others. Frege thought that these rules of inference should be used one after the other, starting with the assumptions of a theorem and ending with the theorem itself, to create an unimpeachable chain of reasoning that would guarantee for us that no geometrical intuition had slipped into the proof. Frege hoped that, by formalising mathematical arguments in this new language, we would become more certain that their conclusions genuinely follow from their premises.

However, around the turn of the century, a new threat to the foundations of mathematics emerged. How do we know our mathematical assumptions are consistent? Who is to say that, when we start with those assumptions and follow these unimpeachable chains of mathematical reasoning, we might not end up with a contradiction? These concerns led the great German mathematician David Hilbert to seek a new way to improve the epistemology of mathematics by establishing the consistency of its assumptions (Zach, 2019). And Frege's approach of mechanising mathematical reasoning was the perfect tool. The idea was that, since formal deductive systems of logic are specified so precisely, they are in fact mathematical objects in their own right. Hilbert therefore proposed to prove mathematically that there could be no chain of steps in a formal system of proof like Frege's that starts with the assumptions of some part of mathematics, such as geometry or arithmetic or real analysis, and ends in a contradiction.

It turned out, however, that Hilbert's project was doomed by one of the most powerful pieces of mathematics that has been used as a formal tool to answer a philosophical question. This is the pair of incompleteness theorems due to the Austrian logician Kurt Gödel (1931). The second of these shows that any mathematical theory with enough expressive power to formulate arithmetic cannot prove its own consistency. The upshot is that, in order to prove the consistency of any significant mathematical theory, you need to assume something stronger than the theory itself. And this undermines the epistemic benefits that Hilbert wanted to gain from such a proof.[6]

With Gödel's theorems, we meet our first no-go theorem. Such results form an important species of formal method, and we'll meet another in the next section. They show that we can't always get what we want. Specifically, they take a list of desiderata, and they show mathematically that they are not jointly satisfiable. Gödel's second incompleteness theorem showed

[6]For an overview of Gödel's results and their effect on Hilbert's program, see (Zach, 2019; Raatikainen, 2020).

that Hilbert could not have what he wanted. He could not have a consistency proof for a significant part of mathematics that assumes less than what was already assumed in that part of mathematics. So he could not establish that the part of mathematics was risk-free without already assuming it was risk-free.

5 The Condorcet paradox and Arrow's theorem

One of the early pioneers of mathematical methods in philosophy was the French Enlightenment philosopher, the Marquis de Condorcet. We will meet his greater achievement in the next section, but first let's consider what is sometimes called *the Condorcet paradox* (de Condorcet, 1785).

The Condorcet paradox belongs to the theory of voting. In its most general form that theory concerns situations in which we have a group of people, each with their individual preferences over a range of options, and we want to identify the group's preferences over those options. The group might comprise all the voting-age citizens in a country, or all the members of a hiring committee, or just a group of friends deciding what to do together on a Saturday afternoon. The options might be policy proposals on a referendum ballot, or candidates for a job, or a range of board games. A voting method takes the preferences of individuals, each of which we represent by an ordering over the options, and spits out the preferences of the group, which we also represent by an ordering over the options. A popular method is *majority voting*. This takes the group to prefer one option to another just in case the majority of individuals in the group prefer the first to the second.

Condorcet's paradox shows that majority voting leads to problems. Suppose there are three job candidates—Amira, Benedict, and Cyra—and three people on the appointment panel—Prija, Quentin, and Roko. These are the panellists' preferences:

	First	Second	Third
Prija	A	B	C
Quentin	C	A	B
Roko	B	C	A

Then, according to majority voting, the panel prefers A to B because Priya and Quentin do, it prefers B to C because Prija and Roko do, and it prefers C to A because Quentin and Roko do. That is, the group's preferences are cyclical. There is no first, no second, and no third. Each candidate is preferred to the other two. And indeed the group preference order is intransitive: A is preferred to B, B is preferred to C, and C is preferred to A, but A is not preferred to itself. Something has gone badly wrong. In the

terminology of voting theory, majority rule violates the conditions of *Unrestricted Domain* and *Social Order*. Together, they say that, if the individuals' orderings are weak orderings, then the voting method should spit out a weak ordering.[7]

Now, majority rule is a very common method for aggregating the different opinions of individuals to generate the group's opinions. But it isn't the only one. Perhaps there is another method that avoids this problem. And of course there are many. But it turns out that every alternative faces its own problem. This is the upshot of Kenneth Arrow's startling impossibility theorem (Arrow, 1951). Like Gödel's incompleteness theorem, it is a no-go result. Where Gödel's result showed that Hilbert's project of proving the consistency of a part of mathematics using proof theory cannot provide the epistemic benefits Hilbert hoped for, Arrow's theorem shows that there can be no voting method that has all of the features we would like it to have.

What are these features? Unrestricted Domain and Social Order furnish the first two, and there are three more. First, *No Dictator* says that there should be no individual in the group such that, whatever preferences the other members of the group have, the group ordering will always match that individual's ordering. Second, *Weak Pareto* says that, if all members of the group prefer one option to another, then the group does. And third, *Independence of Irrelevant Alternatives* says that the order in which the group places to options should be determined entirely by the orders in which each individual places those two options, and should not be affected by the order in which the individuals place any other options. Arrow's theorem states that there is no voting method that has all five properties.

This is a remarkable result. Of course, you might not agree that each of the properties is desirable, in which case it will have little bite. But if you do think they're desirable, Arrow's theorem tells you that your democratic system must be flawed. Presumably some conditions are less crucial to a democratic system, and violations of those might not be fatal to a democracy, but it will nonetheless not be perfect.

Another remarkable feature of Arrow's theorem is its generality. While I introduced it by talking of individual people and their preferences over different options, the mathematical result itself simply concerns the properties of functions that take a sequence of orderings over a range of things and return a single ordering over those things. As a result, it has been applied well outside the boundaries of voting theory. For instance, Samir Okasha (2011) has used Arrow's theorem to investigate Thomas Kuhn's claim that there is no algorithm by which to choose a scientific theory on the basis of theoretical virtues like simplicity, explanatory power, and uni-

[7] A binary relation $\preceq$ is a *weak ordering* if it is reflexive (for all a, $a \preceq a$) and transitive (for all a, b, c, if $a \preceq b$ and $b \preceq c$, then $a \preceq c$).

fication (Kuhn, 1969). In this application of Arrow's theorem, the options are the scientific theories in question, while the voters are the theoretical virtues. Each virtue orders the theories by how well they exhibit that virtue. Arrow's theorem then shows that there is no algorithm that takes these orderings and returns an aggregate ordering of the theories that satisfies the relevant versions of the five properties we described above. Kuhn's claim is therefore supported if those five properties are indeed features we would want to see in an algorithm for theory choice.

Arrow's theorem therefore reveals one of the most powerful features of mathematical methods; a feature that has long been appreciated by mathematicians themselves. A general mathematical result will often have many different interesting applications, some of them well beyond the original area for which the result was conceived. In mathematics, group theory, which has its origins in number theory and geometry, is applied now in physics and in musical theory. Here, a result devised in the analysis of democratic systems illuminates our understanding of scientific theory choice.

6 The Ship of State and the Condorcet Jury Theorem

In the *Republic* Book VI (488), Socrates conjures a fictional scene aboard a ship.

> Picture a shipmaster in height and strength surpassing all others on the ship [...] Conceive the sailors to be wrangling with one another for control of the helm, each claiming that it is his right to steer though he has never learned the art and cannot point out his teacher or any time when he studied it.

So begins the famous 'ship of state' analogy, which compares navigating a ship with running a state. Socrates uses it to persuade his interlocutor that ships—and, by analogy, states—will always be run better by experts like the shipmaster than by the inexpert sailors, who have none of the knowledge and experience required to do the job effectively. In his more famous contribution to democratic theorising—*the Condorcet jury theorem*, introduced in the same remarkable essay in which he formulated the Condorcet paradox—Condorcet rebuts Plato's strong universal claim (de Condorcet, 1785).

To introduce this result, imagine a magician with a trick coin. Unlike a normal fair coin, this coin is more likely to land heads than to land tails on any given toss—for the sake of definiteness, let's say it's 53% likely to land heads on any given toss. Like a normal fair coin, the outcome of each toss is independent of the outcome of all the others. Our magician now starts tossing the coin. Intuitively, the longer they toss the coin, the more likely it is that the majority of the tosses will land heads. And indeed we can prove

this using probability theory. Indeed, after not very many tosses at all, it's extremely likely that the majority of tosses will have landed heads. After 100 tosses, it's about 70% likely, 300 tosses it's about 80% likely, and by 500 tosses, it's almost 90% likely. What's more, it will approach 100% as the number of tosses grows and grows.

Condorcet's jury theorem uses exactly this mathematics, but applies it not to tosses of a biased coin, but to the judgments of slightly reliable members of a population. Suppose we are interested in the answer to a question. The question is yes/no, and the matter is factual, so there's a right and a wrong answer. Suppose we will try to answer this question by asking each person in our population what they think, and taking whatever is the majority opinion. And suppose that each of these people is more likely to be right than wrong about the question. Indeed, let's suppose that each has exactly the same chance of getting the answer right—for the sake of definiteness, let's say they're 53% likely to get it right. And suppose that a given individual getting the answer right is independent of every other individual getting it right. Then, Condorcet proved, just as the chance of the majority of tosses landing heads increases as the coin is tossed more and more times, similarly, the chance of the majority opinion of this group of individuals being right increases as more and more individuals are added to the group. And just as the probability of the coin landing heads on a majority of tosses approaches certainty as the number of tosses increases, so the probability of the majority opinion being correct approaches certainty as the size of the population increases.

How does Condorcet's jury theorem tell against Plato's ship of state argument? Well, Plato assumes that a small number of experts—in the analogy, the shipmaster—will always be more likely to make decisions beneficial to the state than a large group of non-experts—in the analogy, the motley group of sailors. Condorcet's theorem shows this is not necessarily the case. The result is, I think, best seen as a 'how possibly' result. Plato rejects or ignores the possibility that the multitude on the ship of state who do not have any high level of maritime expertise could, collectively, handle the ship better than the single expert shipmaster. Condorcet shows how this is in fact possible. He doesn't, of course, claim that it is actually the case. After all, his theorem has no empirical assumptions about how reliable members of a society are, nor about the independence of their opinions. But he shows a way that the world might be in which a moderately large number of non-experts can, collectively, be more reliable than a small number of experts.

One attractive feature of formal methods in philosophy is that they readily invite constructive extensions, developments, and strengthenings of earlier arguments, rather than encouraging the hunt for the devastating objection. And the history of Condorcet's result since it was rediscovered in the twentieth century has been an inspiring example of that. Condorcet's

own assumptions are very strong, and it's unlikely any society is exactly as he describes: there will surely be some variation in the reliability of individuals; their opinions are surely not completely independent; and so on. So the question arises: under what more realistic conditions do Condorcet's conclusions, or conclusions similarly optimistic about collective wisdom, still hold? And, thanks to a huge literature that has grown up around those results, we now know many many such conditions and their relationship to these conclusions (Estlund, 1994; Kanazawa, 1998; List & Goodin, 2001; Dietrich & Spiekermann, 2013).

7 Modelling intersectionality

So we use a 'how possibly' result to rebut an argument that relies on denying that something is possible. Here's another example—I introduce it because it uses a different formal method, namely, computer simulation. A prominent thesis in social theory says that there are specific disadvantages suffered by individuals who belong to two or more disadvantaged groups, and that these disadvantages are not accounted for fully by pointing to their membership of those two groups taken individually. For example, Black women in America are disadvantaged in ways that we can't account for by looking only at the ways in which Black people in general are disadvantaged, and the ways in which women in general are disadvantaged. We might call this *the intersectionality thesis*. One central objection to this thesis turns on a claim of impossibility: there is no way to design an empirical test of the intersectionality thesis in any specific social case.

In recent work, Cailin O'Connor, Liam Kofi Bright, and Justin Bruner respond to that objection by formulating a series of such empirical tests (O'Connor et al., 2019). And they do that using an interesting formal method. Whereas Condorcet appealed to an analytic mathematical result—that is, one established by proof—O'Connor, et al. use the increasingly popular, powerful, and flexible method of computer simulation. They approach the problem by providing a model of a society in which individuals bargain with one another for some finite resource. In the model, the individuals are divided according to two features: gender and race.[8] These individuals bargain with one another in two separate domains, such as the marketplace and the workplace; and, for each of these two domains, only one of the two identities is salient for interactions—perhaps gender identity is salient in the marketplace, while racial identity is salient in the workplace. They then use computer simulations to find out how the favoured bargaining strategies of individuals of different identities evolve over time, given certain assumptions about what determines how those strategies change from

[8]For ease of modelling, they assume that there are only members of two genders and two races in the population, though that isn't essential to the final result.

one time to the next, and to find out what amounts of resource individuals from different groups receive. They claim that certain assumptions about how these strategies evolve account for intersectional disadvantage.

On the most minimal version of their assumptions, an individual's favoured bargaining strategy changes based purely on their previous strategy's success. Running the simulation with this assumption, we see that members of the intersectional group can be disadvantaged in acquiring this resource simply because the intersectional group is small. On a more moderate version of the assumptions, an individual's favoured strategy is determined not only by past success but by learning from the success of others, but only others with the same identity. Again, intersectional disadvantage arises, and this time it's more pronounced. They consider a number of further assumptions about the evolution of strategies and describe the consequences for intersectional disadvantage.

What results is a series of empirically testable predictions. If A, B, and C are the social forces driving intersectional disadvantage, we'd expect to see P; if D, E, and F are the social forces, we'd expect to see Q; and so on. We can now test the hypothesis that only A, B, and C are the relevant social forces that drive intersectional disadvantage by testing whether P. If we find that P doesn't hold, we might take the hypothesis to be refuted. And similarly for the hypothesis that only D, E, and F are driving such disadvantage. Thus, the formal methods used by O'Connor, et al. have provided a range of hypotheses that make the intersectionality thesis precise enough to be tested.[9] And, what's more, they have shown that we can generate testable hypotheses concerning the forces that govern a complex system like our society using computer simulations, which allow us to draw out predictions from hypotheses that we would not be capable of drawing out using analytical mathematical methods.

8 Logical rocket boosters

Sometimes, the results of applying formal methods in philosophy can look almost like magic. This can happen when we use powerful mathematical results as logical 'rocket boosters' that take our formalised philosophical assumptions and boost them to a conclusion that seems to lie very far from the starting point. To illustrate this, I'll draw from my own area of research, namely, accuracy-first epistemology (Joyce, 1998; Greaves & Wallace, 2006; Pettigrew, 2016).

Above, I introduced Bayesian epistemology, in which we represent a person's beliefs by their credence function and assume that a rational person's credence function is a probability function. The first claim here is a

[9] For an earlier application of formal methods to explicate the intersectionality thesis, see (Bright et al., 2016).

modelling claim—it tells us that we can faithfully represent degrees of belief as numerical credences. The second claim is normative—it tells us what rationality requires of our beliefs represented as credences. What allows us to make this second normative claim? Why does rationality require that my credence in a proposition A and my credence in its negation $\neg A$ sum to my credence in $A \vee \neg A$, for instance, as Bayesianism demands?

Accuracy-first epistemology answers this by showing that I'll go wrong in a particular way if my credences aren't probabilities; and, moreover, I won't go wrong in that way if they are. The central philosophical assumption of this argument is *veritism*, which says that my credences are better, epistemically speaking, the more accurate they are, where a higher credence in a true proposition is more accurate than a lower one, while a lower credence in a false proposition is more accurate than a higher one. In accuracy-first epistemology, we introduce numerical measures of accuracy to measure how well a credence function is doing, epistemically speaking, according to veritism. Now, of course, there are very many numerical measures of accuracy that will satisfy veritism as I've just stated it. Are we allowed to use any of them? Accuracy-first epistemology says no. It says that there is some restricted class of legitimate numerical measures of accuracy.

According to the most popular current characterization of these legitimate accuracy measures, they are all and only the strictly proper ones, where a measure of accuracy is strictly proper if each probabilistic credence function expects itself to be most accurate. Why is this the correct characterization? Very briefly, here's an argument adapted from Jim Joyce (2009). For each probability function P, there is some evidence E you might obtain such that the only rational response to E is to adopt P—in particular, you might learn that P is the probability function that records the objective chances. Now, if P were to expect some other credence function Q to be at least as accurate is it expects itself to be, it would be rational to ditch P in favour of Q. But in that case P could never be rationally required of anyone. So: accuracy measures must be strictly proper.

Let's grant Joyce's argument. Now, from the assumption that our accuracy measure is strictly proper, we can establish the so-called *accuracy dominance theorem* (Savage, 1971; Predd et al., 2009; Pettigrew, ms): if a credence function Q is not a probability function, there is an alternative probabilistic credence function P that is more accurate than Q however the world turns out. This is the flaw that accuracy-first epistemology identifies for any credence function that violates the probability axioms. And it's possible to prove the converse, which says that no probabilistic credence function suffers this flaw.

The proof of this result uses reasonably sophisticated mathematical machinery. One version of it goes via a representation theorem that shows that, to each strictly proper inaccuracy measure, there corresponds a mathe-

matical object called a Bregman divergence, and then draws on the already-known mathematical properties of Bregman divergences (Bregman, 1967; Savage, 1971; Predd et al., 2009).

I won't say more about the philosophical virtues of this argument. What I want to highlight instead is the boost given to the philosophical premises by appealing to the mathematics that goes into the dominance theorem. From Joyce's assumption that every probabilistic credence function might at some point be rationally required, this result shows us that no non-probabilistic credence function is ever rationally permissible. That's a big logical leap. What warrants it is a series of powerful mathematical results.

9 Summing up

Inevitably, I've had the space to mention only a handful of examples of how formal methods have been used in philosophy. However, I hope to have covered the many roles that formal methods might play in philosophy: they help us make our premises and our conclusions precise; they allow us to analyse the inferential strength of our arguments, and improve those arguments; they can provide models of the phenomenon we study, and they allow us to investigate those phenomena by proving results about those models or by simulating the evolution of those models over time, sometimes revealing important and counterintuitive features of them. In many ways, then, formal methods are simply a natural extension of the method of analytic philosophy itself, which also seeks to make our philosophical thinking precise and rigorous, to make our arguments as strong as possible, and to give us a picture of the part of the world we think about that is clear enough that we can explore its features reliably.

References

Anders, J. (2012). Aquinas and quantifier mistakes. *International Journal for Philosophy of Religion*, *71*(2), 137–143.

Arrow, K. J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*(3), 211–233.

Benzmüller, C., & Paleo, B. W. (2014). Automating Gödel's Ontological Proof of God's Existence with Higher-Order Automated Theorem Provers. In T. Schaub (Ed.) *ECAI 2014: Proceedings of the 21st European Conference on Artificial Intelligence*, (pp. 93–98). IOS Press.

Binmore, K. (1994). *Playing Fair (Game Theory and the Social Contract; vol. 1)*. Cambridge, Mass.: MIT Press.

Binmore, K. (1998). *Just Playing (Game Theory and the Social Contract; vol. 2)*. Cambridge, Mass.: MIT Press.

Bregman, L. M. (1967). The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, *78*(384), 200–217.

Bright, L. K., Malinsky, D., & Thompson, M. (2016). Causally Interpreting Intersectionality Theory. *Philosophy of Science*, *83*(1), 60–81.

de Condorcet, N. (1785). *Essay sur l'Application de l'Analyse à la Probabilité des Décisions Rendue à la Pluralité des Voix*. Paris.

Dennett, D. C. (2003). *Freedom Evolves*. New York: Viking Penguin.

Dietrich, F., & Spiekermann, K. (2013). Epistemic Democracy with Defensible Premises. *Economics and Philosophy*, *29*(1), 87–120.

Earman, J. (1992). *Bayes Or Bust?: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, Mass.: MIT Press.

Easwaran, K. (???). *An Opinionated Introduction to Bayesianism*. Routledge.

Estlund, D. (1994). Opinion Leaders, Independence, and Condorcet's Jury Theorem. *Theory and Decision*, *36*, 131–162.

Fitelson, B. (2013). Gibbard's Collapse Theorem for the Indicative Conditional: An Axiomatic Approach. In M. P. Bonacina, & M. E. Stickel (Eds.) *Automated Reasoning and Mathematics*, vol. 7788 of *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer.

Fitelson, B., & Zalta, E. N. (2007). Steps Towards a Computational Metaphysics. *Journal of Philosophical Logic*, *36*, 227–247.

Frege, G. (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle a. S.: Louis Nebert.

Friederich, S. (2018). Fine-Tuning. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 ed.

Gibbard, A. (1981). Two Recent Theories of Conditionals. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.) *Ifs: Conditionals, Belief, Decision, Chance and Time*. Reidel.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik Physik*, *38*, 173–198.

Gödel, K. (2004). Appendix A: Notes in Kurt Gödel's hand. In J. H. Sobel (Ed.) *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge University Press.

Greaves, H., & Wallace, D. (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, *115*(459), 607–632.

Hempel, C. (1945). Studies in the Logic of Induction. *Mind*, *64*, 1–26, 97–121.

Hodges, W. (2001). *Logic: an introduction to elementary logic*. Penguin.

Hosiasson-Lindenbaum, J. (1940). On Confirmation. *Journal of Symbolic Logic*, *5*(4), 133–148.

Joyce, J. M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, *65*(4), 575–603.

Joyce, J. M. (2009). Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In F. Huber, & C. Schmidt-Petri (Eds.) *Degrees of Belief*. Springer.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251.

Kanazawa, S. (1998). A brief note on a further refinement of the Condorcet Jury Theorem for heterogeneous groups. *Mathematical Social Sciences*, *35*, 69–73.

Kuhn, T. (1969). *The Stucture of Scientific Revolutions*. Chicago: University of Chicago Press.

Lewis, D. (1970). Anselm and Actuality. *Noûs*, *4*(2), 175–188.

List, C., & Goodin, R. E. (2001). Epistemic Democracy: Generalizing the Condorcet Jury Theorem. *Journal of Political Philosophy*, *9*, 277–306.

Magnus, P. D., Button, T., Loftis, J. R., Trueman, R., Thomas-Bolduc, A., & Zach, R. (2019). *forallx*. Open Source.

Nicod, J. (1924). *Le problème logique de l'induction*. Paris: Alcan.

O'Connor, C. (ta). Methods, Models, and the Evolution of Moral Psychology. In M. Vargas, & J. M. Doris (Eds.) *Oxford Handbook of Moral Psychology*. Oxford, UK: Oxford University Press.

O'Connor, C., Bright, L. K., & Bruner, J. P. (2019). The Emergence of Intersectional Disadvantage. *Social Epistemology*, *33*(1), 23–41.

Okasha, S. (2011). Theory Choice and Social Choice: Kuhn versus Arrow. *Mind*, *120*(477), 83–115.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.

Pettigrew, R. (ms). Accuracy-first epistemology without additivity. Unpublished manuscript.

Portoraro, F. (2019). Automated Reasoning. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2019 ed.

Predd, J., Seiringer, R., Lieb, E. H., Osherson, D., Poor, V., & Kulkarni, S. (2009). Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions of Information Theory*, *55*(10), 4786–4792.

Priest, G. (2001). *An Introduction to Non-Classical Logic*. Cambridge University Press.

Raatikainen, P. (2020). Gödel's Incompleteness Theorems. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2020 ed.

Restall, G. (2004). *Logic: an introduction*. Routledge.

Rinard, S. (2014). A New Bayesian Solution to the Paradox of the Ravens. *Philosophy of Science*, *81*(1), 81–100.

Savage, L. J. (1971). Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association*, *66*(336), 783–801.

Scott, D. (2004). Appendix B: Notes in Dana Scott's hand. In J. H. Sobel (Ed.) *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge University Press.

Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge, UK: Cambridge University Press.

Sprenger, J., & Hartmann, S. (2019). *Bayesian Philosophy of Science*. Oxford University Press.

Sugden, R. (1986). *The Economics of Rights, Co-operation and Welfare*. Cambridge, UK: Cambridge University Press.

Tal, E. (2020). Measurement in Science. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2020 ed.

Talbott, W. J. (2008). Bayesian Epistemology. In E. N. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Titelbaum, M. G. (???). *Introduction to Bayesian epistemology*. ???

Zach, R. (2019). Hilbert's Program. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2019 ed.