Vaihinger's Fictionalism Meets Binmore's Knowledge-as-Commitment

John A. Weymark

Departments of Economics and Philosophy, Vanderbilt University, VU Station B #35189,
2301 Vanderbilt Place, Nashville, TN 37235-1819, USA
E-mail: john.weymark@vanderbilt.edu

March 2021

Abstract: This article considers how Ken Binmore's use of as-if reasoning is related to Hans Vaihinger's fictionalism. Fictionalism is concerned with the role of idealizations that individuals use to guide their actions and to make sense of the world. Fictionalism employs idealizations that are adopted in spite of being known not to be true. Binmore distinguishes between knowledge-as-commitment and knowledge-as-certainty. With the former, behavior is predicated on the belief that it is impossible that one is wrong, whereas with the latter, behavior is predicated on justified-true-belief. It is argued that by treating knowledge as knowledge-as-commitment, Binmore is employing fictions in Vaihinger's sense. This argument is developed by considering how knowledge-as-commitment is used in Binmore's model of Bayesian decision-making.

Keywords: Ken Binmore; Hans Vaihinger; fictionalism; idealizations; as-if reasoning; Bayesian decision theory.

1 Introduction

Ken Binmore has devoted much of his career to developing a naturalistic approach to rational decision-making and social ethics.[1] Rational choice theory underpins Binmore's ethical theory, whose norms for governing social interactions and the reasons for their adoption are shaped by evolution, both biological and cultural. At a foundational level, *as-if reasoning* plays a fundamental role in Binmore's oeuvre. For rational decision-making, this is reflected in his commitment to a revealed preference methodology in which preferences are revealed or, as he prefers to say, attributed to individuals based on their choice behavior, both actual and hypothetical. Binmore's social ethics makes extensive use of empathetic preferences in which one individual, say A, expresses preferences (in a hypothetic choice situation) between being person

[1] See Binmore (1994, 1998, 2005, 2009, 2020).

B in one alternative or person C in another. In making these comparisons, A engages in as-if reasoning by imagining himself in the situations of B and C, complete with their objective and subjective circumstances.

The publication of *Crooked Thinking or Straight Talk? Modernizing Epicurean Scientific Philosophy* (Binmore, 2020) provides a good occasion to reflect on Binmore's use of as-if reasoning. As the subtitle announces, Binmore describes his methodological approach as a modern version of a scientific philosophy that he attributes to the Greek philosopher Epicurus, an approach that contrasts with what he dismissively describes as that of the metaphysicians. For the most part, Binmore's current methodology is the same as the one that he has employed before describing it as having an Epicurean pedigree, but with some notable changes in emphasis. In particular, Binmore now places great stress on the centrality of knowledge-as-commitment as opposed to knowledge-as-certainty. With knowledge-as-commitment, we "behave *as though* we know things by proceeding as if it is impossible that they could be wrong." (Binmore, 2020, p. 10, emphasis in the original) In contrast, with knowledge-as-certainty, we treat knowledge as justified-true-belief.

In a nutshell, the central tenets of Binmore's methodology are the following:

> Idealization is a fundamental feature of human thought. We build simplified models to make sense of the world, and life is a constant adjustment between the models we make and the realities we encounter. Our beliefs, desires, and sense of justice are bound up with these ideals, and we proceed as if our representations were true, while knowing they are not.

While one can well imagine Binmore writing the quoted sentences, they are in fact taken from the summary that appears on the back cover of Kwame Anthony Appiah's *As If: Idealizations and Ideals* (Appiah, 2017).

Although Binmore often makes brief remarks about the philosophical origins of some of his ideas, he rarely situates them in the literature in any detail. In the case of knowledge-as-commitment, he suggests that Wittgenstein and Hume employed related concepts, but only in passing.[2] Here, I argue that by treating knowledge as knowledge-as-commitment, Binmore is situating himself within a philosophical tradition known as *fictionalism*.

Fictionalism is concerned with the role of idealizations—fictions—that individuals use to guide their actions and to make sense of the world. A feature of fictionalism that sets it apart from other forms of idealization is its use of idealizations that are adopted in spite of being known not to be true. According to fictionalism, descriptions of the world are not be understood literally, but should instead be understood as being "useful fictions". Fictionalism has its origins in the work of Hans Vaihinger, whose 1911 German first edition of *The Philosophy of "As If": A System of the Theoretical, Practical and Religious Fictions of Mankind* (Vaihinger, 1935) is the seminal treatise on fictionalism. In *The Philosophy of "As If"*, Vaihinger introduced the concept of a "useful fiction" and explored its possible meanings and uses at great length.

[2] See Binmore (2011, p. 249) for the claim about Wittgenstein and Binmore (2020, p. 43) for the claim about Hume.

Vaihinger's distinction between real fictions (which embody self-contradictions) and semi-fictions (which do not) is one of the most notable features of his version of fictonalism.

Just like Monsieur Jourdain had been speaking prose without being aware of it in Molière's *Le Bourgeois gentilhomme*, so, too, Binmore has been a practitioner of fictionalism without knowing it. Consider, for example, Binmore's views about models.

> In the absence of an ultimate model of possible realities, we make do in practice with a bunch of gimcrack models that everybody agrees are inadequate. We say that sentences within these models are true or false, even though we usually know that the entities between which relationships are asserted are mere fictions. As in quantum physics, we often tolerate models that are mutually contradictory because they seem to offer different insights in different contexts. (Binmore, 2009, p. 150)

Not only does this passage illustrate fictionalist ideas, it is even uses some of the same terminology.

In this article, I argue that by treating knowledge as knowledge-as-commitment, Binmore is employing fictions in Vaihinger's sense. This argument is developed by considering how knowledge-as-commitment is used in Binmore's model of Bayesian decision-making. Further examples of Binmore's use of fictions are mentioned in my concluding remarks

The plan of the rest of this article is as follows. In Section 2, I provide an introduction to fictionalism. In Section 3, I consider three issues: (i) the role of evolution in idealization, (ii) the degree of idealization used in a model, and (iii) how the distinction between semi-fictions and real fictions is mirrored in the distinction between possible and impossible worlds. In Section 4, I summarize the main features of Bayesian decision theory. In Section 5, I present my argument that Binmore's use of knowledge-as-commitment employs fictions in Vaihinger's sense. Finally, in Section 6, I provide some concluding remarks.

2 Fictionalism

As noted in the Introduction, fictionalism is concerned with the use of idealizations that are knowing false but are nonetheless useful. In this section, I provide a brief overview of the life of Hans Vaihinger and the central tenets of fictionalism.[3]

Vaihinger was born in 1852 in Nehren, in what is now the district of Tŭbugen, Germany. He was a Professor of Philosophy at the University of Halle from 1884 to to 1906. From an early age, he had been troubled with poor eyesight, and it was the deterioration in his sight that led him to resign his position at Halle. For a few years before his death in 1933, Vaihinger was completely blind.

[3] In addition to Vaihinger's own writings, this section draws extensively on the discussions of Vaihinger and his ideas in Fine (1993), Appiah (2017), and Stoll (2020), to which the reader is referred for further details.

Vaihinger began his studies in theology at the University of Tŭbugen in 1870 but soon changed his focus to philosophy and the natural sciences. A leading early influence was the work of the Neo-Kantian Friederich Langen, and this led to Vaihinger becoming a prominent scholar in this tradition. He later founded the journal *Kant-Studien* in 1896. Vaihinger first developed his ideas about fictionalism in his 1877 University of Strassburg habilitation thesis. This work remained unpublished until it appeared as the first part of the first German edition of *The Philosophy of "As If"* in 1911.[4] This volume in its ten editions attracted widespread interest. It was published in an English translation by C. K. Ogden in 1924.[5] In 1919, together with Raymund Schmidt, Vaihinger started the journal *Annalen der Philosophie* (later known as the *Annalen der Philosophie und philosophischen Kritik*) as an outlet for studies that explored the as-if methodology, its scope, and its limitations (Vaihinger and Schmidt, 1919). This journal subsequently became the first incarnation of *Erkenntnis*.

The world is so complex and our cognitive capacities are so limited that, of necessity, individuals employ idealizations as aids to help them understand the world they live in and how they should navigate within it. The latter point is forcefully expressed by Vaihinger when he says:

> It must be remembered that the object of the world of ideas as a whole is not the portrayal of reality—this would be an utterly impossible task—but rather to provide us with an *instrument for finding our way about more easily in this world*. (Vaihinger, 1935, p. 15, emphasis in the original)

Vaihinger places great emphasis on the fact that "*fictions are mental structures.*" "The psyche [i.e, the mind] works over the material presented to it by the sensations," forming and reforming these mental constructs.[6] The nature of these mental processes is something that Vaihinger considers at some length in *The Philosophy of "As If"*.[7]

Vaihinger's appeal to the instrumental value of idealizations provides a general rationale for their use, but it does not capture what is distinctive about his fictionalism. Appiah (2017, p. 56) describes Vaihinger's central thesis as follows:

> [I]dealization involves acting in some respects as if what we know is false is true, this is justifiable because it is useful for some purpose, and the purposes in question are various.

Here, "acting" can be interpreted in an expansive sense that includes adopting beliefs that are not necessarily associated with any action. This quotation highlights a number of fundamental features of Vaihinger's philosophy. First, idealizations are by construction false, at least in some respects. Second, they are used for some purpose.

[4] Vaihinger documents the gradual development of his ideas on the philosophy of "as if" in the autobiographical essay that appears in Vaihinger (1935, pp. xxiii–xlviii).

[5] In this article, quotations from and page references to *The Philosophy of "As If"* are as they appear in the second English edition published in 1935.

[6] The quoted material appears on pages 12 and 157 of Vaihinger (1935), respectively. The emphasis is in the original.

[7] See especially pp. 157–177.

Third, these purposes are various. It is this emphasis on the instrumentality of this kind of idealization that led Vaihinger to coin the phrase "useful fictions" to describe such constructs.

The purposes to which fictional constructs are put to use are of three main kinds. They can be action-guiding, used to predict behavior, or to aid in our understanding of natural phenomena or normative precepts. These purposes need not be mutually exclusive. This trichotomy is nicely illustrated by the uses to which formal models (structures) are put in decision theory. In her introduction to decision theory, Johanna Thoma describes them as follows:

> As an agent, it might help me come to a better decision. But giving formal structure to a decision problem may also help a third party: prior to an action, it may help them predict my behaviour. And after the action, it may help them both understand my action, and judge whether I was rational. (Thoma, 2019, p. 57)

Purposes may be various in a second sense as well; the phenomena that are being considered can take many forms. For example, Adam Smith's assumption that individuals are egoists is a fiction that is used to predict behavior in markets. Similarly, the concept of a point mass is a fiction, but one that is used to help understand how objects interact. Indeed, as its subtitle suggests, much of *The Philosophy of "As If"* is devoted to a demonstration of the ubiquitousness of fictional constructs in the world of ideas, with instances of the employment of fictions in subjects as diverse as ethics, law, and mathematics, among many others.[8]

Fictions are employed by individuals in different roles. In his day-to-day engagement with the world, an individual employs fictions about how the world is structured and the nature of the individuals they interact with. For example, Appiah (2017, pp. 52–53) argues that looking at the world with what Daniel Dennett calls the "intentional stance" is necessary for human interaction. Dennett describes the intentional stance as follows:

> The intentional stance is the strategy of interpreting the behavior of an entity (person, animal, artifact, or whatever) by treating it *as if* it were a rational agent who governed its "choice" of "action" by a "consideration" of its "beliefs" and "desires." (Dennett, 2013, p. 79, emphasis in the original)

This rational agent and all of the terms in quotes are fictions in Vaihinger's sense.

Individuals also employ fictional constructs so as to gain self-understanding and to make moral judgments. For example, Vaihinger (1935, p. 43) argues that complete freedom of choice is a fiction but, nevertheless, this fiction is a necessary one for "we not only make use of this concept in ordinary life in judging moral actions, but it is also the foundation of criminal law."

Similarly, an individual in his role as a scientist uses fictions—models—to understand and predict natural phenomena. Scholars also use fictions when developing normative theories, as is the case when economists offer guidance on the design of

[8] In doing this, Vaihinger developed an elaborate taxonomy of different kinds of fictions.

tax systems using models of how an economy operates and of the behavior of individuals in their economic interactions.

Vaihinger places great emphasis on fictions being by construction knowingly false. For example, Vaihinger (1935, p. 20) argues that Adam Smith realized that to understand the essential features of market interactions, he needed to abstract from secondary motivations such as good-will and habit, and to treat individuals as if their sole motivating factor is egoism, thereby regarding individuals as fictional constructs. But fictions such as these

> are, or at least should be, accompanied by the consciousness that they do not correspond to reality and that they *deliberately substitute a fraction of reality for the complete range of causes and facts*. (Vaihinger, 1935, p. 120, emphasis in the original)

Vaihinger distinguishes between two kind of fictions: real fictions and semi-fictions.

> Ideational constructs are in the strict sense of the term real fictions when they are not only in contradiction with reality but self-contradictory in themselves; the concept of the atom, for example, or the "Ding an sich." To be distinguished from these are constructs which only contradict reality as given, or deviate from it, but are not in themselves self-contradictory (e.g. artificial classes). The latter might be called half-fictions or semi-fictions.[9] (Vaihinger, 1935, p. 16)

Semi-fictions are not only in contradiction with reality, they are also understood to be so. However, a semi-fiction need not be self-contradictory. If a fictional construct is self-contradictory, then it is a real fiction. The examples of Adam Smith's egoist and a point mass illustrate this distinction. With the former, it is recognized that, in reality, individuals have other motives, but it is not self-contradictory to suppose that they do not. With the latter, it is not possible to have an object with mass that has no extension; the concept of a point mass is self-contradictory.[10]

Vaihinger recognizes that there is not a sharp distinction between real fictions and semi-fictions; the distinction is a matter of degree.

> These types are not sharply divided from one another but are connected by transitions. Thought begins with slight initial deviations from reality (half-fictions), and becoming bolder and bolder, ends by operating with constructs that are not only opposed to the facts but are self-contradictory. (Vaihinger, 1935, p. 16)

A distinguishing feature of fictions is that they are not verifiable. In this regard, they are different from hypotheses.

---

[9] "Ding an sich" is Kant's "thing-in-itself".

[10] Concepts like a point mass or a perfect vacuum are examples of what Nagel (1963, p. 25) calls "theoretical terms". A theoretical term denotes something that is not instantiated in reality, but is in some sense the limit of entities that are.

> Fictions are to be distinguished from Hypotheses. The latter are assumptions which are probable, assumptions the truth of which can be proved by further experience. They are therefore verifiable. Fictions are never verifiable, for they are hypotheses which are known to be false, but which are employed because of their utility. (Vaihinger, 1935, p. xlii)

In its action-guiding use, Vaihinger's fictionalism bears some relationship to the pragmatist view that what really matters about, say, beliefs, are the actions that follow from them. However, fictionalism is not simply a version of pragmatism. Appiah (2017, p. 5) captures the essential difference between them when he says that Vaihinger

> thinks that there is a gap between what is true and what is useful to believe. That's why he asserts that most of our thought is best understood as a fiction. If you equated the true and the useful to believe—as pragmatists are sometimes said to do—you would lose exactly the contrast that guided *The Philosophy of "As If"*."

3 Evolution, Degrees of Idealization, and Impossible Worlds

My outline of Vaihinger's approach to idealization has focused on some of its central features. That is sufficient in order to demonstrate its relevance for Binmore's use of knowledge-as-commitment in his formulation of Bayesian decision theory. However, before turning to that task, in this section, I comment on some aspects of Vaihinger's fictionalism that are particularly germane for my discussion of Binmore's methodology.

A natural question to ask is why fictions are useful given that that they are knowingly false. The complexity of the world provides a reason for why thought processes must employ simplified idealizations, but how do they help us in "finding our way about more easily in this world"? Vaihinger does not provide a clear answer to this question. However, in discussing the influences on his philosophical ideas, Vaihinger (1935, p. 25) says that as a result of his exposure to the writings of Johann Gottfried von Herder and Charles Darwin in the late 1860s (when he was still a teenager), "[t]he idea of evolution became one of the fundamental elements of my mental outlook." This suggests that one should look for an evolutionary explanation. While Vaihinger appeals to natural selection on occasion in *The Philosophy of "As If"*, he never systematically explores the role that evolution plays in providing a foundation for the use and usefulness of fictions.

Appiah (2017, pp. 48–53) addresses this issue. He argues that the as-if reasoning employed when one adopts Dennett's intensional stance is evolutionarily adaptive, but that the reasons for why it is so are not clear. He contends that looking at the world with this stance appears to be part of our evolved nature because of its value in facilitating human interaction, but if we only have fictionalized accounts of the world at our disposal, then we can have no true theory to explain why our fictions

pick out the features that they do. Appiah's remarks apply more generally to all as-if reasoning.

In his classic article on economic methodology, Milton Friedman (Friedman, 1953, pp. 21–22) also appeals to natural selection arguments to justify the use of as-if reasoning in order to make predictions about behavior, knowing that by doing so one is knowingly making false assumptions. Friedman argues that one is justified in treating (i) a firm as if it maximizes expected returns in full knowledge of the relevant data and economic laws and (ii) an expert billiard player making shots as if he knows and rapidly applies the relevant physical laws governing the motions and interactions of objects. For if a firm or a billiard player did not appear to behave in these ways, "natural selection" will ensure in the case of the firm that it will not survive without external support, and in the case of the billiard player that he will not be considered an expert.

Friedman makes these observations in the context of his argument that the success of a theory lies in the accuracy of the predictions that follow from its assumptions, and not with how well the assumptions correspond to reality.[11] In the natural and social sciences, theories are described using formal models. Models are fictions in Vaihinger's sense. Models are used not only for predictive purposes; they are also used to gain an understanding of natural and social phenomena. Why and how they are useful for this purpose is a matter of considerable debate. Of particular relevance here is the question addressed by Allan Gibbard and Hal Varian (Gibbard and Varian, 1978, p. 665) about models as used by economic theorists: "In what ways can a model help in understanding a situation in the world when its assumptions, as applied to that situation, are false?"

Gibbard and Varian distinguish between models that are chosen because they approximate reality and those that are caricatures. In their view,

> [c]aricatures … seek to "give an impression" of some aspect of economic reality not by describing it directly, but by rather emphasizing—even to the point of distorting—certain selected aspects of the economic situation. (Gibbard and Varian, 1978, p. 665)

> A caricature involves deliberate distortions [in order] to isolate one of the factors involved in the situation, or to test for robustness under changes of caricature. … [T]he model will be chosen not for the sake of a good approximation, but to distort reality in a way that illuminates certain aspects of reality. (Gibbard and Varian, 1978, p. 676)

As with Vaihinger's observation that fictions differ in the degree to which they falsify reality, Gibbard and Varian recognize that caricatures also come in degrees. How much detail a modeler incorporates in a caricature depends on how much detail is needed to gain the level of understanding that he is seeking. But whatever that level

[11] Friedman's views have been the subject of much critical commentary. A number of the major reservations about his arguments are set forth in Nagel (1963). See also Gibbard and Varian (1978).

is, he acknowledges that the model that is employed has been deliberately chosen to be false in some respects.

The question of how idealized a fiction is or should be is important not only for understanding natural and social phenomena; it is also an important consideration when developing a normative theory. As an illustration of this point, consider the motivations that John Rawls in his *A Theory of Justice* (Rawls, 1971) attributes to individuals when designing principles of justice that are to be applied to the basic structure of a society. In order to elucidate what justice requires, Rawls supposes that when making their decisions about what kind of career to pursue and how much effort to undertake, individuals are, at least in part, motivated by external rewards—command over resources, social status, etc. Appiah (2017, p. 166) suggests that "Rawls may not be idealizing enough" about the facts of human nature (which are at least to some degree shaped by the kind of society they live in) in developing his normative political theory. He notes that Jerry Cohen (Cohen, 2008) has argued that someone who is committed to Rawls' project should instead suppose that such an individual be idealized as being someone who cares more for the intrinsic rewards of work and by how much he contributes to society than by what financial rewards he obtains. Rawls' idealization may be more realistic than that of Cohen, but Cohen and his adherents would argue that fit with reality need not trump other considerations.

Vaihinger's distinction between semi-fictions and real fictions have a counterpart in the distinction between possible and impossible worlds. These two kinds of worlds have been shown to be useful in a number of branches of philosophy (Nolan, 2013). A possible world is a complete description of the way the world might be, one of which is our actual world. A "world" is thought of as being an all-encompassing description of everything that ever exists and everything that ever happens. In contrast, an impossible world is one that is like a possible world in many respects, but contains within its description contradictions or metaphysical impossibilities.

Even if a world is impossible, it is nevertheless possible to reason about what is true in such a world and what is not.

> One way to distinguish different impossible contents is to agree that they are not true at any possible world, but that they are true at different *impossible* worlds. Then the sets of worlds associated with those contents can be different, even if the set of *possible* worlds associated with each impossible content is the same (it is the null set, since they are not true at any possible world). (Nolan, 2013, p. 364, emphasis in the original)

Nolan (2013, p. 365) illustrates this point with the problem of logical omniscience. If an individual's beliefs are characterized by a set of possible worlds, then his belief-set must be a subset of the possible worlds in which p is true if he believes p. But then he must also believe anything entailed by p, which because of limitations on human powers of inference is not credible. However, this inference is not valid if beliefs are characterized by a set of possible and impossible worlds. For example, if an individual believes p and q is entailed by p, there may be impossible worlds

in his belief set that do not include q. As a consequence, believing p does not entail believing q; logical omniscience fails.[12]

When employing one of Vaihinger's fictions, we use an idealization that is in Gibbard and Varian's sense either an approximation of reality or a caricature of it, both of which come in degrees. A fiction that has no connection with reality is useless. Regardless of whether the fiction is a mental construct being used by to guide someone's actions or a model constructed to help understand some phenomenon, the fiction should in some respects be recognizable as being sufficiently close to reality to have some relevance. The concept of similarity between worlds, both possible and impossible, provides a way of addressing this issue.[13] As Nolan (2013, p .363) says:

> If we understand nearness of worlds as a matter of relevant similarity, we can put the thought by saying some close impossible worlds are relatively "well behaved" —they are for the most part like possible worlds, albeit with some impossibilities true according to them.

4 Bayesian Decision Theory

Bayesian decision theory is the approach to decision-making under uncertainty developed by Leonard Savage in his 1954 monograph, *The Foundations of Statistics* (Savage, 1972). Before turning to Binmore's use of fictions in his version of Bayesian decision theory, it is first necessary to summarize the main features of this theory.[14]

A *decision problem* consists of a set of *states of the world* Ω, a set of consequences C, and a set of *acts* A. An act $a \in A$ is a function $a\colon \Omega \to C$ that assigns the consequence $a(\omega)$ to state ω. For Savage (1972, p. 9), a state of the world is "a description of the world, leaving no relevant aspect undescribed." A consequence can be anything that happens in some state as a result of choosing an action. An *event* is a subset of Ω—a collection of states. An individual has a preference $\succeq$ on A (interpreted as "weakly preferred to") that is assumed to satisfy the Savage axioms.

An individual who satisfies the Savage axioms behaves as if he assigns beliefs—his credences—to the states and has a utility function on the set of consequences representing his tastes (or, as they are sometimes called, desires) such that acts are ranked according to their expected utilities. Formally, a *belief* is a probability function $p\colon \Omega \to [0,1]$ that specifies the subjective probability with which each state occurs and *tastes* are described by a utility function $U : C \to \mathbb{R}_+$. Acts are ranked according to their expected utilities if the preference $\succeq$ can be represented by a utility

[12] Appiah (2017, pp. 109–110) suggests that when we think of somebody as being logically omniscient, we only do so in certain respects. This requires being able to think in terms of mutually inconsistent models of the world that are applied in different contexts. As noted in the Introduction, a similar view is expressed when Binmore says that we hold inconsistent models because they are illuminatingly applied in different contexts.

[13] Binmore has reservations about the application of the concept similarity between worlds to decision problems. See Binmore (2011, p. 254).

[14] See Binmore (2009, Chap. 7) for a more extended treatment.

function V on the set of acts that has an expected utility form. Formally, for any act $a \in A$,

$$V(a) = \sum_{\omega \in \Omega} p(\omega)U(a(\omega)),$$

and for any two acts $a, a' \in A$,

$$a \succeq a' \leftrightarrow V(a) \geq V(a') \leftrightarrow \sum_{\omega \in \Omega} p(\omega)U(a(\omega)) \geq \sum_{\omega \in \Omega} p(\omega)U(a'(\omega)).^{15}$$

Information arrives over time. A Bayesian assumes that an individual has prior beliefs about the likelihoods of the states in Ω and updates these beliefs on learning new information using Bayes' Rule. Thus, on learning that event E has occurred (i.e., that the true state is in E), the prior probability assigned to any event F is replaced by its posterior probability, which is the conditional probability of F given E as computed using Bayes' Rule.

If an individual has consistent subjective beliefs, then there is never any real learning going on as time unfolds. Prior to making any decisions, he can anticipate what is implied by the occurrence of any event and can compute his posterior probabilities accordingly. Consequently, in principle, this individual could plan at the outset what to do in every future contingency rather than deferring these decisions until the times they need to be made. Savage (1972, p. 16) colloquially contrasts these two points of view with the proverbs "Look before you leap" and "You can cross that bridge when you come to it". He describes the problem that an individual faces in planning his whole life as a *grand-world* problem (Savage, 1972, p. 84).

Long before Savage, Edith Wharton in her story *The Last Asset* rather vividly articulates the advantage of planning ahead. Her character, Sam Newell, advises his interlocutor to

> [g]et your life down to routine—eliminate surprises. Arrange things so that, when you get up in the morning, you'll know exactly what is going to happen to you during the day—and the next day and the next. ... It saves a lot of wear and tear to know what's coming. For a good many years I never did know, from one minute to another, and now I like to think that everything's cut-and-dried, and nothing unexpected can jump out at me like a tramp from a ditch. (Wharton, 1904, p. 151)

As Savage (1972, p. 16) acknowledges,

> [c]arried to its logical extreme the "Look before you leap" principle demands that one envisage every conceivable policy for the government of his whole life (at least from now on) in its most minute details, in light of the vast number of unknown states of the world, and decide here and now on one policy. This is utterly ridiculous ... because the task implied by making such a decision is not even remotely resembled by human possibility.

[15] To simplify the discussion, the formal statement of the expected utility criterion assumes that the set of states is finite. It is straightforward to restate this criterion in terms of a non-finite state space, but at the cost of introducing some measure theory. In this restatement, beliefs are defined on measurable events.

Nevertheless, if the decision problem being faced is sufficiently simple and can be considered in isolation, Savage proposes adopting the "Look before you leap" perspective. It is this kind of problem that Savage calls a *small-world problem*, and it is only in small worlds that Savage advocates his subjective utility theory. The states in a small-world problem are events in the grand-world problem obtained by partitioning the latter's states. Savage (1972, p. 16) offers no general principle for identifying a small world, noting that this "may be a matter of judgment and experience."

Savage (1972, pp. 100–104) recommends that if an individual's unexamined preferences are not consistent with his axioms, then he should engage in "thoughtful reflection" in order to determine whether his considered judgments are consistent. He suggests that in small-world decision problems, they will be.

Consistency with the Savage axioms implies that an individual makes his choices as if he has consistent beliefs, but the origin of these beliefs in a small world is left unspecified by Savage. Binmore offers a way of arriving at consistent prior beliefs that is similar in spirit to Savage's "thoughtful reflection".[16] This is done using a back-and-forth process of checking for the consistency of the posterior probabilities and then adjusting the prior probabilities if they are not consistent.

5 Knowledge-as-Commitment and Decision Theory

In this section, I present my argument that by treating knowledge as knowledge-as-commitment in his model of Bayesian decision-making for small worlds, Binmore makes use of fictions in Vaihinger's sense. Fictions are employed in two ways. First, the Bayesian decision-maker uses fictions when making his decisions. Second, this decision-maker is himself a fiction.

Binmore (2020, p. 41) describes himself as "a Bayesian who does not believe in Bayesianism." By saying that he is a Bayesian, Binmore is endorsing the view of decision-making under uncertainty developed by Savage for small worlds. More precisely, Binmore supposes that when making choices in a small world, an individual behaves as if he is choosing in accordance with a preference over acts that satisfy Savage's axioms. As a consequence, he can be thought of as being a subjective expected-utility maximizer. As is the case with Savage, Binmore rejects Bayesianism, which requires accepting expected utility theory even when the world is not small—that is, in a "large world" (Binmore, 2007, p. 26).[17]

Binmore (2011, p. 252, emphasis in the original) distinguishes between knowledge and belief: "one *knows* things about the model or world that underlies an analysis. One has *beliefs* about the various states of the world that may arise within the model." For Binmore, knowledge amounts to committing to a model, such as some instantiation of Savage's model of decision-making under uncertainty in small worlds.[18] He says that his view

[16] See Binmore (2007) and Binmore (2009, Chap. 9).

[17] Possible ways of extending Bayesian decision theory to large worlds are considered in Binmore (2007) and Binmore (2009, Chap. 9).

[18] See Binmore (2009, Sec. 8.5.2) and Binmore (2011).

> represents a radical departure from the orthodox view of knowledge as justified true belief. With the new interpretation, knowledge need neither be justified nor true in the senses usually attributed to these terms. It won't even be classified as a particular kind of belief. (Binmore, 2009, p. 150)

In the case of an individual making decisions, Binmore (2009, pp. 150–151) suggests that "we ask … what choice behavior on her part would lead us to regard her acting *as though* she knew some fact within her basic model" and that she "reveals that she knows something if she acts as though it were true in all possible words generated by her model." Furthermore, this is the case even if this involves entertaining contradictions.

> What if a possible world occurs that embodies an in-your-face contradiction of something that [the individual decision-maker] knows? Knowledge-as-commitment requires ignoring the contradiction and continuing to uphold what was previously held to be known. (Binmore, 2011, p. 250)

As these quotations indicate, Binmore treats a Bayesian decision-maker as making use of fictions about the nature of the world. They are fictions in Vaihinger's sense because they are not only useful in helping him find his way in the world; they are also knowingly false. I take it that consistency with the Savage axioms in a small-world problem implies that there are no self-contradictions and, hence, that these fictions are only semi-fictions in Vaihinger's sense.

However, there need not be consistency between the models that an individual uses in the different small-world problems that she might encounter. He might well make use of fictions for one small-world problem that are inconsistent with the fictions used for a different one small-world problem, in which case, the collection of fictions taken as a whole is self-contradictory—it is a real fiction. Just as a physicist might use contradictory models to study different problems without feeling the need to revise his models so as to eliminate any inconsistencies, an individual decision-maker may feel no need to revise the models used in different small-world problems so that they are mutually consistent.[19]

No actual decision-maker is completely committed to what he regards as the facts in his idealization of the world. Thus, Binmore's decision-maker is himself a fiction, one that I contend is a real fiction. Binmore's modeling of a Bayesian decision-maker (an agent) accords well with the following description provided by Appiah (2017, p. 73).

> An agent's degrees of belief and desire are characterized by the behavior to which they would lead in a—conceptually related—fictional agent of a certain idealized kind, and not by the behavior of that actual agent.

But, as Appiah (2017, pp. 83–84) notes, such a fictional agent is supposed to be able to carry out the requisite computations instantaneously and error-free (or at least choose as if they are doing so). No actual agent can do this—it is impossible,

[19] Recall the quotation from Binmore (2009) in the Introduction about physicists using inconsistent models.

just as it is impossible for Friedman's expert billiard player to calculate the requisite physical equations of motion when planning a shot. In both cases, the agents are real fictions in Vaihinger's sense. And like Friedman, natural selection can be appealed to for the success of this way of modeling behavior.[20] Moreover, the description of a fictional agent distorts reality to such an extent that it is best thought of as being a caricature in Gibbard and Varian's sense, not as an approximation to reality.

Binmore's view that knowledge in decision theory is to be interpreted as being knowledge-as-commitment is not standard. For example, an alternative interpretation of knowledge is employed by James Joyce in *The Foundations of Causal Decision Theory* (Joyce, 1999). According to Joyce (1999, p. 75), when making a decision A is a small-world situation

> on the basis of less than fully considered beliefs and desires we thereby commit ourselves to the view that our fully considered beliefs and desires would sanction the choice of A from among the alternatives listed in the [small-world decision problem].

In other words, we commit to making consistent choices, just as Binmore's Bayesian decision-makers do.

Joyce (1999, p. 76, emphasis in the original) says that

> we can think of a rational agent's attitudes toward the states, outcomes, and acts in a small-world decision problem as her *best estimates* of the attitudes that she would hold regarding those states, outcomes, and acts in the grand-world context.

This quotation highlights what distinguishes Joyce from Binmore with respect to the nature of the knowledge being appealed to. For Joyce, the idealizations employed in a small-world decision problem are not knowingly false; they are not fictions in Vaihinger's sense.

Binmore (2009, p. 151, emphasis in the original) regards his approach to knowledge as "the *de facto* norm in scientific enquiry." For example,

> [m]athematicians say that a statement is an axiom to indicate that they are committed to behaving as if it were true in the context under study. When arguing by contradiction, they even commit themselves temporarily in this way to treating statements as true that they plan to refute. Binmore (2011, p. 257)

Similarly, Binmore (2009, p. 151) says that when a physicist is asked about the ontology of an electron, he is "simply being invited to clarify the working hypotheses built into [his] model" rather than being asked for "a disquisition on the ultimate nature of reality."

Just as is the case with scientific models, the accumulation of contradictions with a model of decision-making may at some point require abandoning it in favor of a different model (e.g., one with different prior probabilities). Speaking of scientific models, Binmore (2009, pp. 151–152) says that

[20] See, for example, Binmore (2009, Sec. 1.6).

> when the data no longer allows [someone] to maintain her commitment to a particular model [she] throws away her old model and adopts a new model—freely admitting as she does so that she is being inconsistent.

Applied to a model of decision-making, we have a counterpart to a scientific revolution.

One way in which a contradiction may arise in a decision problem is for a decision-maker to encounter a zero probability event. As Binmore (2009, p. 152) notes, "any attempt by [someone] to massage her system of beliefs into consistency would fail when she found herself trying to condition on a zero-probabiilty event." Binmore's response to this problem is to not entertain the possibility of zero-probability events, and instead to treat them as a limiting cases of events with small positive probabilities. This resolution is predicated on the view that anything can be inferred from a contradiction (Binmore, 2011, p. 251). While this is true with classical logics, it is not true with logics that make use of impossible worlds. This suggests that zero probabilities can be accommodated by allowing for impossible worlds.[21]

Binmore recognizes that the properties that are appropriate for a formal model of knowledge or of possibility depend on the kind of knowledge or possibility being considered.[22] Furthermore, they differ for small and large worlds. Similarly, a different logic may be needed in order for a decision-maker to deal with zero-probability events than the logic underlying Binmore's small-worlds Bayesianism.

6 Concluding Remarks

In order to demonstrate that Binmore is a practitioner of fictionalism and that he makes use of fictional constructs of the kinds first introduced by Hans Vaihinger, I have focused on the roles that knowledge-as-commitment play in his version of Bayesian decision theory. In these concluding remarks, I briefly note two of the other ways in which Binmore employs fictions.

There has been a great deal of controversy about whether common knowledge of rationality in a finite game of perfect information implies that the game can be solved by backward induction. According to Binmore (2011), the answer to this question depends on how knowledge of rationality is interpreted. If it is interpreted as being knowledge-as-commitment, he argues that this inference is valid. With knowledge interpreted in this way, as a result of player A making a decision that places player B at a decision node that would be reached with probability zero had player A played

[21] I am not aware of any explicit use of impossible worlds to model belief revision when a zero-probability event is encountered. However, there is an extensive literature concerned with belief revision when belief with probability zero is a feature of the model. Some of this literature specifically deals with the problem of extending Bayesian updating to zero-probability events. Basu (2018) provides a brief overview of some of the approaches to belief revision that are most relevant for the Bayesian case. He also offers some possible solutions to this problem.

[22] See Binmore (2007, Sec. 4), Binmore (2009, Chap. 8), and Binmore (2011, Sec. 5).

rationally, B does not then regard A as being irrational. Rather, he maintains his belief that A is rational, knowing that this belief is false. In other words, continuing to play as if A is rational involves B regarding A's rationality as a fiction.

An empathetic preference involves an observer imaging himself in the positions of other people complete with their objective and subjective circumstances. In Binmore's social contract theory (Binmore, 1994, 1998, 2005), empathetic preferences provide a foundation for the interpersonal utility comparisons that are employed in the fairness norms that are shaped by the forces of biological and cultural evolution so as to fairly share the gains made possible by social cooperation. Binmore's use of empathetic preferences for this purpose is borrowed from John Harsanyi (e.g., Harsanyi, 1977, pp. 51–52).

The use of empathetic preferences raises a metaphysical problem that is nicely articulated by Mongin (2001).

> [H]ow much of the observer's identity is preserved by ... empathetic identification of the nondeductive sort? Is there enough left, as it were, to warrant the claim that it is the observer who makes the preference judgement? The point has been put forward that if i must effectively enter j's or k's mental state to make extended preference judgements, it cannot be i, after all, who makes them. There would be something self-destructive in the way identification works.

Put another way, if there are attributes that are essential to the observer's identity, then it is impossible for the observer to fully identify with other individuals so as to form an empathetic preference (Adler, 2014, p. 132).[23]

Binmore's use of knowledge-as-commitment allows him to sidestep this metaphysical problem. He can do this by having the observer act in a hypothetical choice situation in which he is to choose between occupying the positions of two individuals with all of their attributes *as if* he really is able to completely identify with either of them, and this is in spite of knowing that complete identification is impossible. In other words, such an observer employs a real fiction in Vaihinger's sense because he is comparing impossible situations.

The examples I have chosen to illustrate Binmore's use of idealizations that are knowingly false establish that he is an adherent of fictionalism, whose methodology originates in Hans Vaihinger's *The Philosophy of "As If"* published over a century ago. It is possible to find many other examples of fictions in Vaihinger's sense in Binmore's writings. Considering his body of work from this perspective, I believe, sheds new light on Binmore's contributions to the foundations of rational decision-making and social ethics.

[23] Greaves and Lederman (2018, p. 642) believe that, with this indentification-based interpretation of what an empathetic preference is, the observer would not be able to make sense of the identifications needed to make interpersonal comparisons.

Acknowledgements

I dedicate this article to Ken Binmore. Ken has been a good friend ever since we first met at the Public Choice Institute held at Dalhousie University during the summer of 1984. I presented this article as a keynote lecture to the online Joint Meeting of Asian Network for the Philosophy of the Social Sciences (ANPOSS), the European Network for the Philosophy of the Social Sciences (ENPOSS), and the Philosophy of Social Science Roundtable (POSS-RT) hosted by Hitotsubashi University, March 4–7, 2021. I am grateful to the participants on this occasion for their comments.

References

Adler MD (2014) Extended preferences and interpersonal comparisons: A new account. Economics and Philosophy 20:123–162

Appiah KA (2017) As If: Idealization and Ideals. Harvard University Press, Cambridge, MA

Basu P (2018) Bayesian updating rules and AGM belief revision. Journal of Economic Theory 179:455–475

Binmore K (1994) Game Theory and the Social Contract, Volume 1: Playing Fair. MIT Press, Cambridge, MA

Binmore K (1998) Game Theory and the Social Contract, Volume 2: Just Playing. MIT Press, Cambridge, MA

Binmore K (2005) Natural Justice. Oxford University Press, New York

Binmore K (2007) Rational decisions in large worlds. Annales d'Économie et de Statistique 86:25–41

Binmore K (2009) Rational Decisions. Princeton University Press, Princeton, NJ

Binmore K (2011) Interpreting knowledge in the backwards induction problem. Episteme 8:248–261

Binmore K (2020) Crooked Thinking or Straight Talk? Modernizing Epicurean Scientific Philosophy. Springer, Cham, Switzerland

Cohen GA (2008) Rescuing Justice and Equality. Harvard University Press, Cambridge, MA

Dennett DC (2013) Intuition Pumps and Other Tools for Thinking. W. W. Norton, New York

Fine A (1993) Fictionalism. Midwest Studies in Philosophy 18:1–18

Friedman M (1953) The methodology of positive economics. In: Essays in Positive Economics, University of Chicago Press, Chicago, pp 3–43

Gibbard A, Varian HR (1978) Economic models. Journal of Philosophy 75:664–677

Greaves H, Lederman H (2018) Extended preferences and interpersonal comparisons of well-being. Philosophy and Phenomenological Research 96:636–667

Harsanyi JC (1977) Rational Behavior and Bargaining Equilibrium in Games and Social Situations. Cambridge University Press, Cambridge

Joyce JM (1999) The Foundations of Causal Decision Theory. Cambridge University Press, Cambridge

Mongin P (2001) The impartial observer theorem of social ethics. Economics and Philosophy 17:147–179

Nagel E (1963) Assumptions in economic theory. American Economic Review, Papers and Proceedings 53:211–219

Nolan DP (2013) Impossible worlds. Philosophical Compass 8:360–372

Rawls J (1971) A Theory of Justice. Harvard University Press, Cambridge, MA

Savage LJ (1972) The Foundations of Statistics, second revised edn. Dover, New York, first edition published in 1954

Stoll T (2020) Hans Vaihinger. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy, spring 2020 edn, Metaphysics Research Lab, Stanford University

Thoma J (2019) Decision theory. In: Pettigrew R, Weissberg J (eds) The Open Handbook of Formal Epistemology, PhilPapers Foundation, London, ONT, Canada, pp 57–106

Vaihinger H (1935) The Philosophy of "As If": A System of the Theoretical, Practical and Religious Fictions of Mankind, second English edn. Kegan Paul, Trench, Trubner, London, translated by C. K. Ogden.

Vaihinger H, Schmidt R (1919) Programm der zeitschrift. Annalen der Philsosophie 1:iii–vi

Wharton E (1904) The last asset. Scribner's Magazine 36:150–168