

# **Fictions, Conditionals, and Stellar Astrophysics**

Mauricio Suárez

Mauricio Suárez is at the Department of Logic and Philosophy of Science, Complutense University. Correspondence to: Departamento de Lógica y Filosofía de la Ciencia, Facultad de Filosofía, Universidad Complutense de Madrid, 28040 Madrid, Spain. E-mail: [msuarez@filos.ucm.es](mailto:msuarez@filos.ucm.es)

This article argues in favour of an inferential role for fictions in scientific modelling. The argument proceeds by means of a detailed case study, namely models of the internal structure of stars in stellar astrophysics. The main assumptions in such models are described, and it is argued that they are best understood as useful fictions. The role that conditionals play in these models is explained, and it is argued that fictional assumptions play an important role as either background or antecedent conditions. I then expand on the argument for the compatibility of fictions and scientific realism. I argue that realism and antirealism plausibly offer correspondingly different accounts of the semantics of these fictional conditionals.

## **1. The Inferential Role of Scientific Fictions**

This is the third and concluding article in a trilogy defending the inferential role that fictions play in scientific modelling. In the previous articles (Suárez 2009, 2010), I introduced the idea of inferential expediency as the hallmark of scientific fictions, and I argued that fictions are not incompatible with scientific realism. In this article, I begin by distinguishing fiction from idealisation. I then illustrate the distinction by means of a case study in astrophysics, namely stellar structure models. I identify the relevant fictional assumptions in these models, as well as the three empirical quantities that are the object of observational astrophysics. I describe the corresponding ‘fictional conditionals’, which provide these models with inferential expediency. I finish by bringing these lessons to bear on the compatibility of the use of fictions in models with scientific realism. In particular, I argue that the realist may choose to offer a non-truth-functional semantics for these conditionals that preserves the thesis that the practice of model building in science aims at truth.

A number of caveats apply, and it is best to introduce them straightaway. The case study is, of course, open to alternative interpretations. It is employed as an argument in two different senses. First, the case study exemplifies the thesis that fictional assumptions play an inferential role in models. In so doing, it provides an articulation of its main claim. But there may be other ways to articulate this claim. Thus the first function of the case study is illustrative: it helps to display and articulate one way in which fictions play an inferential role in scientific practice. I do claim that this role is typical of fictions; but I do not argue for this ‘typicality’ inductively on the basis of the case studies. An inductive argument for either of these conclusions (the inferential function of fictions, and their typicality) would of course require a much larger inductive basis. Although I believe it possible to provide an indefinite number of cases of this sort, this would be a much larger enterprise than I am able to undertake here. The second function of the case study is hermeneutical. It is intended to furnish a plausible interpretation of modelling practice that does some justice to our knowledge of this and other interrelated aspects of the practice more generally—beyond the specifics of the case. The case study provides an argument for the thesis in this sense too. The fictional construal of some of the assumptions in these models provides an understanding of astrophysicists’ modelling practice. This is not something that a simple textbook example can achieve, and requires some detailed analysis of the assumptions at work in the case study—hence the detailed description in section 2 below. In advancing the claim that fictions have an inferential function, I also aim and hope to provide a sensible and plausible interpretation of that practice—and even though it may not be the only interpretation, it is important to make the case that it *is* a possible interpretation.

An alternative interpretation of the case study, for instance, would employ idealization conditions and *ceteris paribus* clauses, as opposed to fictional conditionals. I believe that nothing much hinges upon this at the methodological level, as long as the idealizations in question are ‘non-controllable’, as defined below. An idealisation is a false description of a system that approximates it to some degree in some respect. A ‘controllable’ idealisation is one that can be corrected according to a prescribed recipe (usually provided by a theory) in order to bring it arbitrarily close to the real case. A non-controllable idealisation, by contrast, cannot be corrected in this manner: There is no theory that stipulates how to bring the description closer to the real case.<sup>1</sup> In this respect non-controllable idealisations and fictions do not differ since fictional assumptions are neither dictated nor stipulated by theory—and most idealisations in science are non-controllable. This means that the differences between idealization and fiction will not reveal themselves at the level of scientific methodology: scientists will follow the same recipes in building models, whether they incorporate non-controllable idealizations or fictional assumptions. In particular, they will not attempt to derive or

explain such assumptions from high-level theory. Nor will there be any apparent difference in the application of such models (i.e. models containing either idealizations of the non-controllable kind, or fictional assumptions) to further systems of interest and study.

The differences will come in rather at the level of scientific epistemology. In particular, non-controllable idealisations are contrary to at least some strong versions of scientific realism. In the tradition of Ernan McMullin (1985), a model cannot be *ad hoc* with respect to a theory if it is to provide confirmatory evidence for that theory. Uncontrollable idealisations are introduced *ad hoc* with respect to the accepted theory of the domain, and cannot be eliminated by means of any recipe provided by the theory—so they do not provide confirmatory evidence for the theory. There is however another form of reliability of the theory for which such models may be good indicators, which we may refer to as instrumental reliability.<sup>2</sup> This is the reliability of the theory merely as an instrument of application: If the idealisation is constructed with the help of a particular theory, then its success does indicate at least that the theory is instrumentally reliable in yielding appropriate models of the phenomena.

In other words, idealisations are relative both to a theory and to the objects in its domain. Thus an idealized description of an object in a theory's domain can confirm the theory in a strong or substantial realist sense if the description has been derived from the theory, or can be appropriately subsumed under such a theory. If, by contrast, the idealized description of the object in the theory's domain is of an uncontrollable kind, then it cannot confirm the theory in these terms, although it may nonetheless increase our degree of confidence in the instrumental reliability of the theory. Scientific fictions, by contrast—it is argued in the present article and in the rest of the trilogy to which it belongs—are not so 'relative to' a theory, nor to an object in the domain, but rather on account of their function in modelling practice. Scientific fictions are normally neither derived from a particular theory, nor employed with the aim to vindicate or confirm such a theory. They do not therefore provide evidence for a theory's truth, or its empirical adequacy; nor do they raise a theory's instrumental reliability. Fictions do of course arise in a particular context of modelling, which, I argue, provides an important framework for the fictions' inferential roles (since it provides the background conditions that make fictional conditionals true). But the context is not normally reducible to a theory or consistent set of theories, being instead made up of a large and heterogeneous number of practices of model building, and accepted claims regarding those practices.

The difference between fictional assumptions and non-controllable idealizations does not typically show in the methodology pursued in their construction. Which one is the applicable notion in which particular case partly depends on the nature of the case. But there is always some interpretive leeway for the philosopher of science to cast the

cases in his or her preferred mode of description. Some cases, such as the simple harmonic oscillator model of a pendulum in classical mechanics, seem clearly controllable idealisations, directly derived from theory. But the status of other models, such as the frictionless ideal inclined plane, is less clear. The model of measurements provided by quantum mechanics is one of the clearest instances of a useful scientific fiction. And so on. The epistemic facts in each modelling case will make more or less plausible one or another interpretation and I claim that the three cases referred to so far (frictionless planes, quantum measurements, stellar interiors) are at least as plausibly rendered as fictions. My additional claim, which I defend in section 4, is that casting these cases in terms of fictions is less unfriendly towards scientific realism than it may appear at first. In particular scientific realism would come out worse were these cases rendered in terms of ‘uncontrollable’ idealisations. The relative counter-intuitiveness of this fact, I argue, rests on an inappropriate understanding of the nature and role of fictions in scientific modelling practice.

## **2. Models of Stellar Structure in Astrophysics**

In this section, I expand on the case study in stellar astrophysics in order to emphasise and illustrate the distinctive feature of fictions as prime vehicles for inferential expediency. In the third section I provide a hermeneutical understanding in terms of fictional conditionals, and in the last two sections 4 and 5 I discuss the semantic and epistemic implications that follow from treating many of these assumptions as fictions.

Within stellar structure models in astrophysics, a star is normally abstractly defined as any body of gas uniformly constituted by a mixture of hydrogen and helium, bound together by self-gravity, and which radiates energy from some internal source (Prialnik 2000, 1). The terms ‘uniform gas’, ‘bound’, ‘self-gravity’, ‘internal source’, and ‘radiation’ are precise technical terms, yet their application to any given star must be understood broadly and tentatively. Thus a star’s boundaries are typically imprecise, and although self-gravity dominates it rarely is the only force acting on the constitutive gas. Radiation is not the only means of stellar energy output, and the internal sources of energy may be very varied. A star is of course not a closed system, but is in constant contact with its environment, the interstellar medium. As the star burns its fuel, it generates elements with higher atomic numbers, such as oxygen, carbon, and nitrogen, which it ejects into the interstellar medium (this is the sense in which stars are said to be the ‘kitchen’ of the universe, where the heavy elements that make life possible are ‘cooked up’).

Thus a typical real star is not very much like the star abstractly defined above (it

rather looks a lot more like the object depicted in Figure 1). The reason why such a highly simplified abstract fiction is accepted instead is related to its ability to swiftly generate the appropriate empirical predictions. The main regularity of observational astrophysics is the so-called Hertzsprung–Russell (HR) law, which establishes a correlation between the luminosity ( $L$ ) and the effective surface temperature ( $T_{eff}$ ) of a main sequence or ‘ordinary’ star.<sup>3</sup> Although often referred to as the ‘empirical’ law of stellar astrophysics, the quantities plotted in a typical HR diagram are not directly observable, but are rather inferred by means of some very simple extrapolations. There are two genuinely observable—or directly measurable by experiment—characteristics: the incident radiation flux from a stellar source into a telescope, also known as its apparent brightness ( $I_{obs}$ ), and the characteristic set of spectral lines of the radiation, or distribution of radiation intensity per wavelength ( $I$ ), also known as the (electromagnetic, not merely optical) spectrum ( $I_\lambda$ ) of its source. In addition, distances ( $d$ ) can be estimated, occasionally directly by measuring parallax (the relative position of a nearby star against the fixed background of more distant stars at two given times of the year—i.e. at two different points of the earth’s elliptical orbit), but more often as a result of complex statistical treatment of parallax to known star clusters, and other galactic objects.<sup>4</sup>

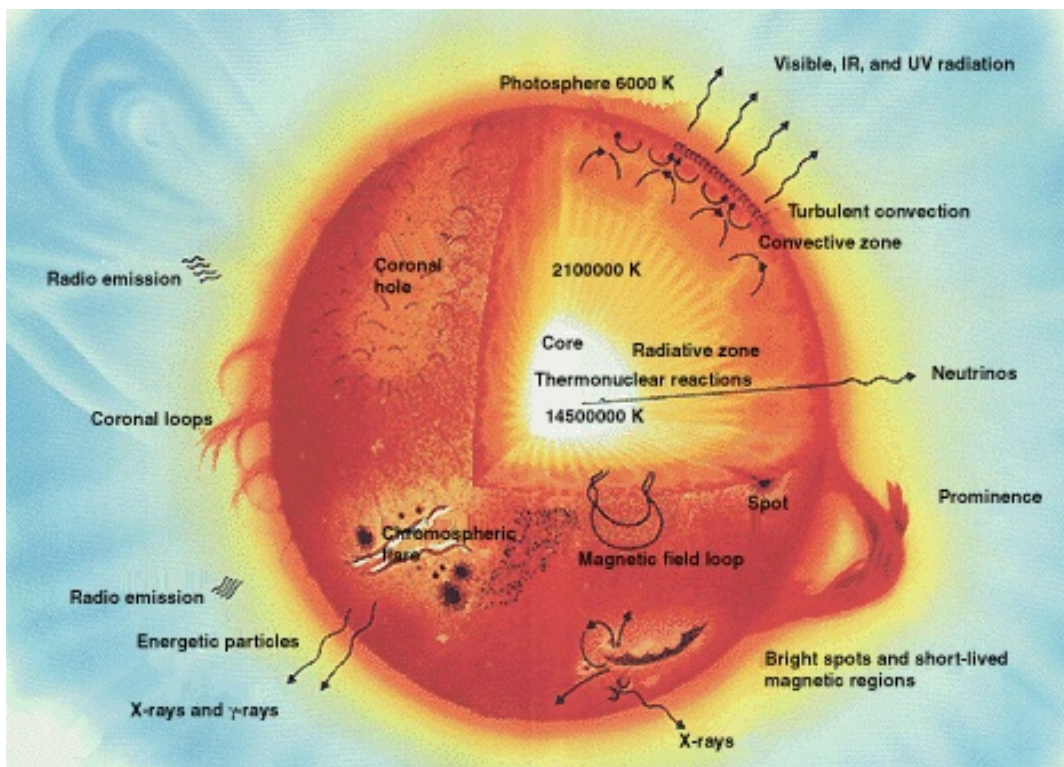


Figure 1: The Internal Structure of a Star

Given precise values for these three quantities ( $I_{obs}$ ,  $I_\lambda$ ,  $d$ ) for any given star source, it is possible to derive its luminosity, its effective surface temperature ( $T_{eff}$ ) and the chemical composition of its photosphere (our outer layer) as follows. The luminosity ( $L$ ) of a star is its ‘energy power’: the amount of energy radiated per unit time, and it is a simple function of distance and observed incident radiation flux, i.e.:  $L = 4\pi d^2 I_{obs}$ . The effective surface temperature of a star may also be derived from its luminosity under the assumption that the star is a blackbody spectrum. We may then estimate the effective temperature as:

$$T_{eff}^4 = L / 4\pi R^2 \sigma,$$

where  $\sigma$  is the Steffan–Boltzmann constant.

Finally, under a similar assumption, the chemical composition of the photosphere is given by a spectral analysis of the emission and absorption lines in its spectrum ( $I_\lambda$ ). Luminosity ( $L$ ), effective surface temperature ( $T_{eff}$ ), and chemical composition are then regarded as the basic observational quantities of stellar astrophysics. The first two are mapped in a typical HR diagram, which is taken to represent the evolution or life cycle of stars, from their birth as objects of high temperature and luminosity to their maturity as lower temperature and luminosity objects, and in some cases their eventual final stage as ‘red giants’ and ‘white dwarfs’ away from the main sequence. In Figure 2, red giants are to the upper right corner of the diagram, indicating a high luminosity but low temperature object, while white dwarfs are in the lower left corner, as low luminosity and high temperature objects. The life cycle of stars takes them to begin in supernova explosions, as high temperature and high luminosity objects on the upper left corner of the main sequence. Then, as they gradually burn their fuel, stars become cooler and less bright, ‘descending’ in the main sequence. Eventually, at least in the case of particularly massive stars, as heavier elements are ejected into the interstellar medium, the outer layers become lighter and cooler, but, since for a while at least, the core continues to produce blackbody radiation, the radiation forces greatly outweigh the bounds of self-gravity, and the star boundaries expand hugely. As the outer layers move still further away from the radiative core, they cool down further, so the spectrum lines correspondingly shift to the red as the surface temperature goes down. A red giant is therefore a comparatively huge mass of lighter and cooler gas, with a still fully functional radiative core. A white dwarf, by contrast, is effectively dead—in its final stage, as the star burns out its fuel completely, radiation forces cease abruptly, and the star collapses under its own gravity into a tiny and very dense cloud of compact gas made up of comparatively heavier elements, and eventually

dies altogether as its internal power engine ceases to emit any radiation.

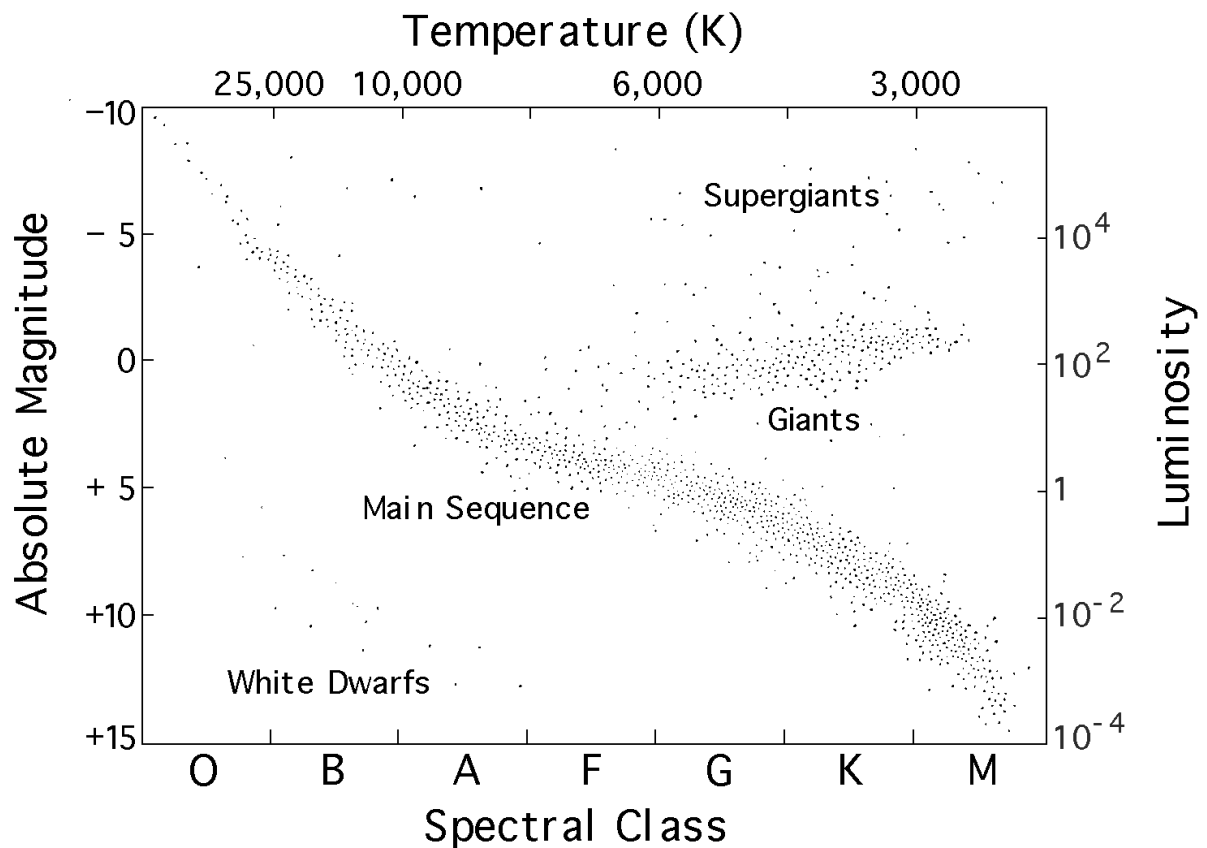


Figure 2: The Hertzsprung-Russell Diagram

The assumption that a star is a blackbody (sometimes called the ‘thermal equilibrium assumption’) is thus by no means trivial. It entails that the energy transfer within a star is entirely radiative (Prialnik 2000, 16ff.)—and in particular it rules out convective or conductive transfer from e.g. hotter to colder parts of the gas. For all we know the gas in a star is unlikely to be fully radiative, and very rarely in the lifetime of a star—and even then only, if anywhere, at its very inner core—are the radiation and gas temperatures identical. The assumption at best holds locally, and probably not at all during the most active and spectacular periods of a star’s life both within and outside the main sequence. From a theoretical point of view the justification for the neglect of convective and conductive energy transfer is the fact that it carries no implications for the temperature distribution within the star. In other words, the assumption that a star is in thermal equilibrium, and is hence a blackbody, can hardly be said to be ‘a little bit away from the truth at the edges’. It is not introduced into the model as a true, or

approximately true assumption—but in order to facilitate the appropriate inferences with respect to the temperature distribution in a star.

There are at least three other similar assumptions in stellar structure models (see Prialnik 2000, 6–8). First, the star is assumed to be isolated in empty space and to be non-interacting with any other stellar object (provisos are made for binary stars). We may refer to it as the ‘isolation assumption’. It implies that the boundaries of the stellar gas are sharp and distinct from the background interstellar medium. However, stellar gas differs at best locally in density from the interstellar medium surrounding it and—at least in young stars—not at all in composition. The sharpness of the boundary is debatable and in fact debated by many astrophysicists working in the field known as coronal stellar astrophysics. The fact that coronas are typically much hotter than the photosphere and subject to huge disturbances caused by the star’s magnetic flux already indicates that the sharp boundary is only in fact sharp if regarded as the source of the detectable radiation, particularly in the visible end of the spectrum (coronas are too light to be significant sources of emitted radiation—they are in fact plasmas). From the point of view of the physics of the whole complex system, the assumption is not even an approximation, since a star is in permanent electromagnetic dynamical interaction with its corona, and vice versa. The isolation assumption also turns out to be a convenient fiction, introduced in order to simplify calculations.

Second, the star is taken to be composed uniformly of, primarily, hydrogen and, secondarily, helium, when as a matter of fact most old stars always contain a large amount of heavy elements. The sun’s internal composition is typically taken as the standard, with up to 91% of its atoms made up of hydrogen, accounting for roughly 70% of its mass fraction, and 9% of its atoms made up of helium, accounting for roughly 27% of its mass fraction. The remaining 3% of the sun’s mass fraction is made up of oxygen, carbon, nitrogen, etc.—but it is usually discounted in stellar models altogether. The sun’s composition is then taken as the benchmark, with the mass fractions for hydrogen and helium at 70–30% assumed as the rule. We may call this the ‘uniform composition assumption’. Stellar models make this assumption in order to derive a very simple set of mathematical relationships between the dynamical properties relevant to the evolution of the stars, known as the stellar structure equations. In fact, once this assumption is in place, the dynamical evolution of a star depends only on its initial mass, which rather simplifies all kinds of models of the stellar interior and the star’s life cycle (Prialnik 2000, 7, 81–86). However, it must be acknowledged that the ‘uniform composition assumption’ does not need to be treated as fiction. Unlike the other assumptions, it does make sense to speak of it as an approximation to what we believe most stars’ chemical composition to be. This is because we do have some established means to account for the presence of heavy elements and slightly diverse ratios of



hydrogen and helium. In fact, bringing such considerations to bear is essential in modelling both older stars and the evolution of any one particular star throughout its life cycle. Stellar models of white dwarfs, in particular, assume that their cores are composed of a carbon-oxygen mixture, with all remaining hydrogen or helium confined to the outer layers (Collins 1989, ch. 5; Hansen and Kawaler 2004, ch. 9). Hence the uniform composition assumption may be regarded as an idealization just as appropriately as a fiction. Only under a very wide sense of ‘fiction’, encompassing all idealisation, distortion or misrepresentation as ‘fiction’, need this assumption be classified as fictional.<sup>5</sup>

Third, a star is assumed bound entirely by self-gravity, and consequently the shape of the star is assumed to be spherically symmetrical. Yet, we actually suspect rotational forces within stars to be huge, and to distort its spherical symmetry—typically along the axis of rotation. And, although the electromagnetic interaction with the interstellar medium is negligible, the internal magnetic forces within the star are not—and can also distort its shape. Hence the ‘spherical symmetry assumption’ is again a misrepresentation of real stars, and must be understood to be a fiction or idealization of sorts. In most cases of regular main sequence stars, the assumption may be treated as an approximate idealization; but in some cases (binary and distorted stars mainly) spherical symmetry is only a fiction brought in to facilitate calculations.

When these assumptions are put together they allow us to build models that are effective in deriving both the internal and the ‘observable’ properties of stars such as their surface temperature ( $T_{eff}$ ) and luminosity ( $L$ ), as follows. The assumptions yield the four equations of stellar structure: hydrostatic equilibrium, continuity, radiative transfer, and thermal equilibrium (Prialnik 2000, ch. 5). These equations provide values for some key properties of stellar interiors, such as pressure ( $P$ ), heat flow ( $F$ ), rate of nuclear energy release per unit mass ( $q$ ), temperature ( $T$ ), etc. Each of these quantities is typically associated with a layer of the stellar interior, and is parametrized with respect to mass—since the uniform composition and spherical symmetry assumptions together imply that the mass increases monotonically with radial distance. The application of these quantities to the outmost layer of the star—the photosphere—then yields values for the observable properties: effective temperature ( $T_{eff}$ ), luminosity ( $L$ ), and mass fraction at the photosphere ( $I_\lambda$ ). In other words, the assumptions are brought in because they jointly facilitate inferences to properties of the star that we would not otherwise be able to draw. Some of these assumptions (e.g. chemical composition) may in addition be close to the truth—but their function within the model nonetheless is to facilitate inference, and in that respect they are all working as fictions.

### 3. Fictional Conditionals

I have discussed four main assumptions that go into models of stellar structure in astrophysics. In a wide enough sense of the term, they are all fictions. But even in a narrower sense that distinguishes fiction from idealisation, two of them (isolation and thermal equilibrium) are best conceived of as fictions in the models of most stars in the main sequence, while a third (spherical symmetry) is best regarded as such in many models of stars off the main sequence. What mainly distinguishes their status is that they do not appear to function as approximate idealizations of even the non-controllable kind. Their primary function is not, or at least not merely, to faithfully represent the target; rather they are in the first instance introduced into the model with the aim to provide inferential shortcuts within the model to the target's properties, including those properties that are 'observational'.<sup>6</sup> Their success is evaluated against pragmatic goals such as the prowess and reliability of the inferences that they condone.

The basic structure of these inferences has three components that may be roughly described as follows. There are first of all some background assumptions that must be in place before the models can be used for any inferential purposes. The most outstanding is distance (of the stars as sources of radiation from the detectors on the earth's surface). Given the expansion of the universe, this is in turn related to age (or lines in the HR diagram describing the life-cycle of the star). Without values of the distance of the source it is impossible to estimate any one of its 'observable' properties: its luminosity ( $L$ ), effective surface temperature ( $T_{eff}$ ), and spectral distribution ( $I_\lambda$ ). I already mentioned that the estimation of distances is in most cases not a simple matter of measuring parallax, but requires considerable knowledge of the large-scale structure and evolution of the universe, including the typical statistical distribution of stars in clusters and galaxies, and the distribution of mass across such clusters and galaxies, which in turn requires knowledge of their life-cycles and dynamical evolution, etc. This sort of knowledge also presupposes some basic rules of inference regarding the observable properties and qualities of interstellar objects and phenomena. For instance, notoriously, it requires the interpretation of redshift in a stellar spectral type as directly caused by its linear speed away from the earth.

Given all this background knowledge, which we may denote as  $B$ , the fictional assumptions discussed above make it possible via the equations of stellar structure to derive precise values of the 'observable' properties of any given stellar source. The input variables into the models typically include a few basic parameters regarding the internal structure or initial state of the stellar gas, such as temperature at the core of the star ( $T_c$ ) and the initial volume of gas coalescing under self-gravity ( $V_i$ ). The model's equations, or inference rules, then generate as output the values for the observable

quantities and properties of the star, such as  $L$ ,  $T_{eff}$ , and  $I_\lambda$ , thus providing a picture of the state of the photosphere that may be compared with actual data. All this is compounded of course by the fact that any estimates of distance, or age, are also estimates of the age of the light that reaches us, so the picture that it portrays must of course be inferred to have been the state of the source way back in the (relative to our present frame of reference) past.

Thus the statements that these models licence take the following form: ‘Given  $B$ , if  $A$  then  $C$ ’, where  $I$  denotes the background conditions; the antecedent,  $A$ , is the conjunction of a number of clauses, including at least the initial values of the parameters  $T_c$  and  $V_i$  and any other inputs into the models; and the consequent is also a conjunction, which includes at least values for  $L$ ,  $T_{eff}$ , and  $I_\lambda$ . We may refer to such conditional statements as ‘fictional conditionals’ since they require a number of fictional assumptions to be in place such as, in our case, the isolation and thermal equilibrium assumptions. The assumptions must be in place, but it is not altogether determined where in the inferential chain they appear—there is some latitude here for philosophical reconstruction, as follows. The simplest account of the inferential practices that these models sanction places the fictional assumptions in the antecedent  $A$  of what I have called fictional conditionals. But in a different model the fictional assumptions appear as part of the background conditions  $B$ . And then, there are all kinds of combinations of these, which may suit certain case studies better than others. In other words, our philosophical representation of scientists’ reasoning is underdetermined by their inferential practices. Their reasoning may typically be reconstructed by either placing the fictional assumptions in the antecedent, or as part of the background conditions. What must be respected is the fact that the fictional assumptions are the inferential grounds for the conclusions regarding ‘observable’ properties in the star. But whether these assumptions are grounds for correct inferences because they are part of the requisite background, or because they are merely entertained hypothetically in the antecedent, is not a fact determined by the scientific practice itself—it is rather up for (hermeneutical and philosophical) grabs.

#### **4. The Semantics of Fictional Conditionals**

The precise representation of the inferential practices of astrophysicists is a matter of interpretation—but not one devoid of important philosophical consequences. Among other things, the precise placing of fictional assumptions in these practices determines the appropriate semantics for the ‘fictional conditionals’ that ground the practices. In particular, it does make a difference to semantics whether the fictional assumptions are

included amongst the background knowledge  $B$ , or instead figure in the antecedent  $A$  of the conditional. A truth-functional conditional, such as material implication, is false only when the antecedent  $A$  is true and the consequent  $C$  is false. All other combinations of truth-values of  $A$  and  $C$  make the conditional true. In particular, it is true whenever both antecedent and consequent are false. In building models of stellar interiors scientists employ equations that yield the right values of the observable quantities  $L$ ,  $T_{eff}$ , and  $I_\lambda$  out of whatever values they ascribe to  $T_c$  and  $V_i$  within some very broad ranges of plausibility. Now, we are not in a position typically to tell the values of the physical properties of the core of the star, in particular, its temperature, chemical composition, pressure, and gas density. These can only be estimated. Astrophysicists will accept whatever values, within broad ranges, yield the right output for the observable quantities. So, a fictional conditional with a false antecedent is acceptable if the consequent is true. But not otherwise: If the output values are incorrect—for instance if they contradict the observations—the modellers will look for different input values that agree with the observations, or will reject the fictional conditionals altogether (by rejecting, or modifying, either the equations—i.e. the rules of inference—or the fictional assumptions themselves).

At first sight this seems an outright contradiction of the truth-functional account, since it appears to entail a rejection of a conditional with a false consequent, whatever the truth-value of the antecedent—including crucially, the case where the antecedent is false. By contrast, astrophysicists seem prepared to accept the very same conditional whenever they believe the description of observable facts in the consequent to be true. Yet, if the conditionals are truth-functional, with a false antecedent, their truth-value cannot depend upon the truth-value of the consequent. Since there are some well-known cases where conditionals with false antecedents and true consequents are false, it must be the case that they are generally false (if they are indeed truth-functional). A standard example is ‘if it is square, it has four sides’, which is true when applied to a square, a rectangle, and a triangle (Edgington 2008). In other words, there appears to be a *prima facie* conflict between the role of fictional conditionals in scientific practice and the truth functional account of their semantics. How can this apparent tension be resolved?

An obvious solution—discussed in the next section—is to relinquish the link between acceptance and truth. Maybe astrophysicists accept or reject conditionals, but not as true or false. However, this move is tantamount to a form of scientific antirealism, and I am in this paper interested in possible ways to preserve realism in the face of the conflict. So let us for the sake of argument continue to assume, with the scientific realist, that the acceptance of a statement, in general, carries a commitment to accept it as true. There are two routes to resolving the tension without giving up on this assumption. For suppose that we assume, following the realist, that the scientists’

acceptance of fictional conditionals of the form ‘given  $B$ , if  $A$  then  $C$ ’ in fact carries a commitment to taking the conditionals as true. One can then simply insist that all fictional assumptions must be placed amongst the background conditions  $B$ . The antecedent of the conditionals thus becomes sparse, and it does not follow that the conditional is false when the antecedent is false—since every statement in the antecedent may now very well be true too. Thus, for instance, we may insist on placing only values for the temperature at the core of the star ( $T_c$ ) and the initial mass of gas ( $V_i$ ) amongst the statements cited in the antecedent  $A$  of the conditional. We have no real reason to suspect these estimates are too far from the truth. The conditional is true then because there is evidence for the truth of the consequent  $C$ , and we are assuming the antecedent  $A$  to be true. Nothing here would offend a standard truth-functional analysis of the fictional conditionals, since a conditional with true antecedent and consequent is truth-functionally always true. The role of the fictional assumptions is exclusively restricted to the background that enables the conditionals to transmit truth across logical consequence. Thus, for instance, the role of the ‘thermal equilibrium’ and ‘isolation’ assumptions is merely to articulate within the models of stellar structure the claims contained in the antecedent and consequent of the conditional in such a way that  $A$  entails  $B$ , and the conditional is true. They do not themselves figure amongst the claims made by  $A$  or  $C$ .

There is yet another option for the realist, which makes it possible for the fictional assumptions themselves to possess a truth-value (namely, falsehood) and be placed firmly in the antecedent of the conditional. This option has several independent advantages: it beefs up the conditions that appear in the antecedent while making suitably thin the assumptions that need to be presupposed as background. The fictional assumptions are, after all, very specific to stellar structure. So it seems desirable to place them (with the exception of chemical composition and possibly, in some cases, spherical symmetry) within the actual antecedent of the conditional. The move has the added advantage (from the realist’s point of view) to characterise all the fictional assumptions as false, on a par with non-controllable idealizations. It then becomes a matter for empirical inquiry to determine the degree to which they are closer or farther from the truth. The success of the conditional thus fails in any way to increase the probability of the fictional assumptions, but this is entirely of a keeping with realism, since the assumptions are false. Nevertheless, we are left with the task to square the realist requirement that science establishes true conditionals with the fact that the truth-values of antecedent and consequent are not always in agreement with a truth-functional understanding of the conditional.

The obvious way out at this point is to give up on truth-functionality (i.e. the thesis that the semantics of conditionals is a function of the truth-values of antecedent

and consequent), at least for the fictional conditionals one finds in scientific modelling practice. Instead we may assume that there are cases where conditionals with false antecedents and true consequents are true (following on ‘if it is a square, it has four sides’ applied to a rectangle), such as those accepted by astrophysicists, and other cases where the same conditionals are rejected on account of their having a false consequent (unlike what we are prepared to do for ‘if it is a square, it has four sides’ when applied to a triangle), such as empirically tested stellar structure models with reliably falsified statements in the consequent. In other words, we may assume that the truth-value of a conditional is not strictly a function of the truth-values of consequent and antecedent. It turns out, however, that this is not such a high price to pay, since there are powerful independent reasons to believe that this is indeed generally the case—i.e. to believe that the semantics of conditionals is not generally truth-functional (Edgington 1997, 2008). Nevertheless, there is the residual worry for the realist—discussed in the final section—namely that we gain a coherent understanding of fictional assumptions as truth-valued but only at the expense of having to accept that the assumptions are outright false.

To sum up, there are two options to preserve the idea that scientists only accept true conditionals. The first option is to suppose that fictional conditionals maybe accepted as true (on the realist understanding of scientific acceptance) whenever the consequent is true, because the fictional assumptions do not appear in the antecedent—which may hence also be true—but only amongst the background conditions. In accepting the conditionals as true, scientists are not entertaining the fictional assumptions even hypothetically, but the semantics of the conditionals is fully truth-functional. The second option is to suppose that fictional conditionals maybe accepted as true because the conditionals are not truth-functional: their truth is not a function of the truth-values of their component propositions. It then becomes possible to understand the divergent practice, even though the antecedent is deemed false on account of containing fictional assumptions. Scientists will evaluate conditionals not just as a function of the truth-values of their component propositions, but also attending to features of the context that go beyond truth-values. Perhaps the right semantics for fictional conditionals is in terms of conditional probability or some other probabilistic option: the antecedent then just makes the consequent probable under the given background conditions, and the conditional probability measures the likelihood of the observational statements in the consequent given the fictional assumptions. Either way, I argue, the acceptance of fictional conditionals is rendered compatible with at least one typical hallmark of scientific realism—the claim that science aims at truth.

## **5. Scientific Realism and Fictional Conditionals**

Scientific realism is a conjunction of many different commitments and it has meant somewhat different things to different authors over the years. I here focus only on one particular claim of realism, and a very general one, namely that science is a cognitive enterprise defined by its search for the truth. That science aims at the truth is an undisputed part of any realism. Many contemporary philosophers, such as Bas van Fraassen (1980) and Arthur Fine (1987), take this tenet to be the essentially distinguishing feature of scientific realism. My claim is that the fact that fictional conditionals are acceptable to scientists as long as their consequents are true (i.e. as long as the models that they are embedded in are able to predict the right values of the ‘observable’ quantities  $L$ ,  $T_{eff}$ , and  $I_\lambda$ ) is compatible with the thesis that science aims at establishing true statements, including true conditionals. So understood, as a very general thesis concerning the aims of science, scientific realism is not incompatible with the inferential role that has been adjudicated to fictional assumptions in this article.

We have so far assumed that in accepting or rejecting these conditionals astrophysicists accept or reject them as true. But the assumption is, of course, a substantial one in the debate between realism and antirealism. Some contemporary forms of antirealism (notably, Van Fraassen’s constructive empiricism) deny that accepting a theoretical proposition  $P$  entails believing it to be true; rather it entails merely the belief that  $P$  is empirically adequate. Any commitment over and above that belief is strictly pragmatic in nature. To accept  $P$  is then to believe in the empirical adequacy of  $P$  and, in addition, at best, to pragmatically commit to working ‘as if’  $P$  (Van Fraassen 1980, 12, 88). Such a programmatic commitment to work ‘as if’ something or other is the case is, of course, a characteristically and traditionally ‘fictionalist’ attitude. So not surprisingly, constructive empiricism has sometimes been taken to be suggesting a form of fictionalism about unobservable entities and their properties (Van Fraassen 1980, 35; Rosen 1994).

In this article, the focus is exclusively on the constructive empiricist understanding of fictional conditionals. It seems clear that the constructive empiricist may maintain the cleavage between acceptance and truth for these statements too, and turn it to his advantage. For example, according to constructive empiricism, scientists’ acceptance of the statement ‘the temperature at the core of the star is  $T_c$ ’ does not require them to believe in the existence of a ‘core’ or its temperature (or indeed any other amongst the core’s properties). It only requires a belief in the observable consequences of the statement, which in the context of the models here discussed, include luminosity ( $L$ ), effective surface temperature ( $T_{eff}$ ), and spectral distribution ( $I_\lambda$ ). Hence, according to this view, accepting a fictional conditional does not carry a commitment to the truth of either background or antecedent conditions. It carries a

commitment only to the truth of the observational statements in the consequent.

Independently of any issues regarding the interpretation of the term ‘observational’ (and whether such statements can really be said to be about genuinely observable properties), this epistemic outlook lets us off the hook with respect to the difficulties regarding the interpretation of conditionals raised in the previous section. By rejecting the link between acceptance and truth, constructive empiricism can accommodate the divergent use of fictional conditionals with false antecedents—i.e. the fact that their acceptance seems to vary with the truth-value of the consequent. This does not require any failure of truth-functionality. For acceptance does not, on this view, require the evaluation of the truth-value of the conditionals in the two different cases—with true and false consequents. Rather, the divergence merely points to different standards of acceptance itself, which do not bear on truth-values, or even judgements of truth-values. If astrophysicists accept one set of conditionals and not the other, this is not, on this antirealist view, because they are judging the conditionals to be true in one case and false in the other. It merely expresses a pragmatic commitment to work ‘as if’ one set of conditionals were true and not the other.

It is not surprising that an antirealist epistemology can be easily made compatible with the inferential role of fictional conditionals. It is perhaps more surprising that a realist epistemology—taken as the general dictum that science pursues the truth—can also be shown, in its own way, to accommodate the role of fictional assumptions in models. A first thing to notice is that, as has already been emphasized, the ‘observable’ facts in the consequents of these fictional conditionals are not typically observable in Van Fraassen’s sense. On the contrary they include statements regarding quantities that can strictly only be inferred from observations against a rich background of theoretical assumptions, including some of those explicitly contained in *B*. This means that the very acceptance of conditionals with true consequents already carries a minimal commitment to realism, understood as the claim that science is a truth-seeking enterprise. Regardless of how we understand the conditionals themselves, the very judgement that the statements in the consequent are true goes beyond what would be strictly demanded by empirical adequacy. The realist certainly has a leg to stand here, but the debate about the nature of the acceptance of the consequents will simply reproduce well-known disputes between realists and antirealists regarding the nature of theoretical statements in general. So, let us put this point aside and focus on the more substantial issue, for our purposes, regarding the status of the conditionals themselves. As we just saw the divergent practice of acceptance amongst astrophysicists is easy to account from the point of view of antirealism. Can the realist too account for it?

In the previous section, I argued that this practice of acceptance itself does not determine whether the fictional assumptions appear amongst the background conditions



that fix the inferential rules that apply to the conditionals, or instead appear in the antecedent of the fictional conditionals themselves. Thus the realist has two options. He or she can then adopt the view that the fictional assumptions are all included amongst the background conditions; thus retaining a truth-functional semantics for the conditionals. The concern with this strategy is that the background conditions include, on this interpretation, most of the substantial assumptions, such as ‘thermal equilibrium’ and ‘isolation’, while the antecedent looks extremely thin—and includes just some very basic quantitative estimates for the values of a couple of key properties of the core. Furthermore, on this account, the fictional assumptions are not themselves playing any role in the search for truth. They are neither judged true nor false; and they are neither justified nor criticized on account of their proximity or otherwise to truth. Their function seems entirely independent of their truth-value, since they merely play the role of setting the context where the conditionals hold. They are then genuinely fictional assumptions that seem inimical to any realist construal. That is, it seems that the success of the conditional can not raise the probability of the background assumptions, including these ‘purely fictional’ ones—thus supporting an antirealist epistemology that takes fictions to never receive any confirmation from the success of the predictions they help generate, contrary to what is the case for ordinary hypotheses.<sup>7</sup> Thus the first alternative preserves the thought that scientists aim at establishing true conditionals, fully understood in terms of a truth-functional semantics, but it does so at the expense of consigning a large number of model assumptions to a set of fictional background conditions that are not amenable to confirmation in realist terms.

The alternative is to insist on interpreting the fictional assumptions as a part of the antecedent, but to render the semantics of the conditionals themselves non-truth-functional. Thus, on this account, scientists not only aim at establishing true conditionals, but they do so in a manner that respectfully builds the ‘fictional’ assumptions within the very antecedent of the conditional, as hypothetical suppositions. Since the fictional assumptions must now receive a truth-value, we must, on account of the ways in which we know them to differ from reality, assign them the value of falsehood. However, the assumptions themselves are subject to confirmatory relations, and can be ultimately tested in the way most theoretical assumptions are tested—via their predictive and explanatory power. The conditionals themselves are accepted as true whenever a false antecedent combines with a true consequent—but of course it is now a matter of empirical inquiry to determine both the truth of the consequent and the degrees and respects in which the antecedent depart from truth. Hence, it is perfectly possible to conceive the whole enterprise of modelling as truth-directed in three different ways. First, scientists look for true conditionals; second, they look for true consequents, which they test against observable predictions; and third, they look for

false antecedents that nonetheless are approximately true, and which may henceforth be said to be approximately explanatory. Thus, the second alternative is closest to the scientific realist tenet that science fundamentally aims at truth—yet it pays the price of having to give up on a truth-functional semantics for the conditionals. Whether this is a price that the realist should ultimately be prepared to pay is a question for further study.

## **6. Conclusion**

In this article, I have introduced in full detail a case study in astrophysics that serves to illustrate the role that fictional assumptions play in the practice of modelling. The practice is open to a certain amount of philosophical reconstruction, and I have argued that scientific realists have two alternative approaches to this practice that, on the face of it, preserve the idea that science aims at truth through its models. First, fictional assumptions may be understood to be part of the required background knowledge—required for what I have called fictional conditionals. Second, fictional assumptions may be interpreted to lie within the antecedent of such conditionals, giving rise to a non-truth-functional semantics for conditionals. In both approaches, it is still the case that scientists will only accept those conditionals they believe to be true. Hence, scientific realism is preserved—in the fundamental and abstract claim that science aims at truth through its modelling practices. While the role of fictions in science does not in any way require scientific realism, it neither fundamentally contradicts it.

## **Acknowledgments**

I thank audiences at Paris, Vancouver (APA Pacific conference), LSE, and Montreal (PSA conference), and Alisa Bokulich, Steve Downes, Catherine Elgin, Arnon Levy, James McAllister, Henrik Zinkernagel, and an anonymous referee of this journal for helpful and detailed comments and reactions. Thanks also to the Harvard Philosophy Department and the London Institute of Philosophy for sponsorship during research visits there. Financial help is acknowledged from the Spanish Government, research project grant FFI2011-29834-C03-01.

## References

- Batterman, R. 2009. "Idealization and Modeling." *Synthese* 169: 427–446.
- Collins, G. W. 1989. *The Fundamentals of Stellar Astrophysics*. New York: W. H. Freeman.
- Edgington, D. 1997. "On Conditionals." *Mind* 104: 235–329.
- Edgington, D. 2008. "Conditionals." In *Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. <http://plato.stanford.edu/archives/win2008/entries/conditionals>, accessed on 19 May 2013.
- Fine, A. 1986. *The Shaky Game: Einstein, Realism and the Quantum Theory*. Chicago, IL: University of Chicago Press.
- Hansen, C. J., and S. D. Kawaler. 2004. *Stellar Interiors: Physical Principles, Structure, and Evolution*. New York: Springer.
- McMullin, E. 1985. "Galilean Idealization." *Studies in History and Philosophy of Science* 16: 247–273.
- Morrison, M. 2005. "Approximating the Real: The Role of Idealizations in Physical Theory." In *Idealization XII: Correcting the Model*, edited by M. Jones and N. Cartwright, 145–172. Amsterdam: Rodopi.
- Prialnik, D. 2000. *An Introduction to the Theory of Stellar Structure and Evolution*. Cambridge: Cambridge University Press.
- Rosen, G. 1994. "What Is Constructive Empiricism?" *Philosophical Studies* 74: 143–178.
- Suárez, M. 1999. "The Role of Models in the Application of Scientific Theories: Epistemological Implications." In *Models as Mediators: Perspectives on Natural and Social Sciences*, edited by M. S. Morgan and M. Morrison, 168–196. Cambridge: Cambridge University Press.
- Suárez, M. 2004. "An Inferential Conception of Scientific Representation." *Philosophy of Science* 71: 767–779.
- Suárez, M. 2009. "Scientific Fictions as Rules of Inference." In *Fictions in Science: Philosophical Essays on Modelling and Idealization*, edited by M. Suárez, 158–178. London: Routledge.
- Suárez, M. 2010. "Fictions, Inference and Realism." In *Fictions and Models: New Essays*, edited by J. Woods, 225–245. Munich: Philosophia Verlag.
- Vaihinger, H. 1924. *The Philosophy of 'As If'*. London: Kegan Paul.
- Van Fraassen, B. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- Winsberg, E. 2010. *Science in the Age of Computer Simulation*. Chicago, IL: University of Chicago Press.

## Notes

<sup>1</sup> The claim that idealizations split into these two forms, at least, is by now entrenched in the literature. Morrison (2005, actually written and circulated about a decade earlier), Suárez (1999), and Batterman (2009) share a very similar understanding of the distinction, and put it to similar epistemic uses.

<sup>2</sup> The term ‘instrumental reliability’ was introduced and applied to idealizations in this way in Suárez (1999). Winsberg (2010, ch. 7) applies the same concept to simulations.

<sup>3</sup> The HR law was independently discovered by the Danish astronomer Ejnar Hertzsprung (1873–1967) and the American Henry Norris Russell (1877–1957). A typical diagram displaying the law is reproduced as Figure 2.

<sup>4</sup> It is worth remarking that these fundamental empirical properties are not observable in Van Fraassen’s (1981) sense of the term. Hence even if these properties lie at the end of some of the inferential chains that start off with fictional assumptions, the sort of antirealist position that would ensue from restricting commitment to such properties is not constructive empiricism. (Other differences with constructive empiricism are emphasized later in the text in discussing the nature of fictional conditionals.) The fact that constructive empiricism may be understood as a kind of fictionalism (e.g. Rosen 1994) does not imply that fictionalism in general requires constructive empiricism, or that it is committed to it.

<sup>5</sup> For the distinction between wide and narrow fictions, or fictionalism more generally, see Suárez (2009). In this article I am developing the fine-grained distinction between fiction and idealization, so I correspondingly adopt a narrow sense of scientific fiction.

<sup>6</sup> On the inferential conception of representation that I have defended (Suárez 2004), the aims of ‘faithfully representing’ and ‘reliably inferring’ are not only unopposed, but may be said to ride upon each other. These two activities are in fact closely linked conceptually—since reliable inference turns out to be a necessary (albeit neither sufficient nor constitutive) condition on representation. The views that I defend in this paper regarding fictional assumptions in models are a corollary of my more general views regarding representation, but they may also be adopted independently of any views regarding scientific representation in general.

<sup>7</sup> The distinction between fictions and hypotheses was often emphasized by Vaihinger (1924) and informs the standard line on fictions that was rehearsed in the first section of the article.