Università degli Studi di Genova
North Western Italian Philosophy Consortium (FINO)

Ph.D THESIS

Epistemic Values and Scientific Theories

Supervisor:
Jan Sprenger

Reviewers:
Gustavo Cevolani
Jacob Stegenga

Candidate:
Cristina Sagrafena

Academic Year 2022/2023

Contents

Introduction

In 1977, Thomas Kuhn published an essay, destined to be very influential, titled ‘Objectivity, Value Judgement, and Theory Choice’. As we will see in section 1.2, the main claim of the essay is that, although the five epistemic criteria of accuracy, consistency (both internal and external), scope, simplicity and fruitfulness are the objective basis for theory choice, they do not uniquely determine it. One of the reasons for this is that scientists can disagree about the importance to be given to the criteria when they pull in different directions.

As Kuhn puts it, “accuracy may, for example, dictate the choice of one theory, scope the choice of its competitor” (1977, p. 322), so the question is: how should accuracy be traded off against scope? The main point of Kuhn’s argument is that there is not a single answer to this question. Rather, scientists can reach different acceptable answers, depending on which criterion they give more importance to.

Such subjective differences in applying epistemic criteria led Kuhn to famously claim that: "the criteria of choice whit which I began function not as *rules*, which determine the choice, but as *values*, which influence it" (p. 331, emphasis is mine).

The idea that criteria should be better understood as values was favorably received at Kuhn’s time. For instance, in McMullin 1982, we read: "these criteria clearly operate as *values* do, so that theory choice is basically a matter of *value-judgement*" (p. 16, emphasis is mine). And, nowadays, it seems to be fairly uncontroversial (e.g. Reiss and Sprenger 2020, section 3).

More generally, since, at least, the publication of Kuhn’s essay, the issue surrounding the relationship between epistemic values and scientific theories, with a special focus on theory choice, became a central topic in philosophy of science. Important names in this regard, each of which proposed their own list of values, are, for instance, the aforementioned McMullin (2008), Quine and Ullian (1978), Longino (1990), Laudan (1990) (see sections 1.3).

Still today, the topic retains its influence, intertwining with the theme of scientific objectivity. A popular idea of scientific objectivity, in fact, states that science is objective in the extent in which its core - i.e. the gathering of evidence and the assessment and acceptance of scientific theories - is free from contextual values, and guided only by epistemic ones.

Broadly speaking, epistemic values are assumed to be indicative of the truth (or acceptability) of a theory, while contextual ones are moral, personal, social, political and cultural values. A famous exemplification of improper use of contextual values concerns the condemnation, by the Third Reich, of a large part of contemporary physics, such as the theory of relativity, on the base that its inventors were Jewish.

More precisely, the idea of scientific objectivity as freedom from contextual values has a normative component, expressed as follows:

Value-Free Ideal (VFI): scientists should strive to minimize the influence of contextual values on scientific reasoning, e.g. gathering evidence and assessing/accepting theories.

Instead, the question whether **VFI** is actually attainable is the subject of

Value-Neutrality Thesis (VNT): scientists can - at least in principle - gather evidence and assess/accept theories without making contextual value judgements.

Criticisms to these two claims led to interesting reflections and results concerning theory choice and epistemic/contextual values (for a complete exposition of the debate surrounding **VFI** and **VNT**, see Reiss and Sprenger (2020), section 3).

For instance, against **VNT**, Longino (1996) argues that there is no such thing as pure epistemic values in that the use of values like accuracy, simplicity, consistency, and so on is not politically neutral. To see this, she confronts the Kuhnian values with feminist ones such as novelty, ontological heterogeneity, mutuality of interaction, and so on. And she argues that using the former instead of the latter in gendered contexts - i.e., contexts that are structured by gendered power asymmetries - can contribute to keep hidden the asymmetric power relations in favor of the male. To prove her point, she proposes different examples. In one of these, she juxtaposes the Kuhnian criterion of external consistency, i.e. the consistency of the theory with accepted theories in other fields, and the opposite feminist value of novelty, which recommends theories that differ in significant ways from the accepted ones. Now, using external consistency in contexts in which there is need for theoretical frameworks other than those that have functioned in gender oppression by making gender invisible, surely perpetuates this invisibility.

The longstanding issue in philosophy of science concerning the relationship between scientific theories and epistemic values is the frame of this thesis, too. Indeed, here, I present three essays, each of which constitutes a chapter, dealing with such an issue from a formal point of view, i.e., by using logical patterns, and mathematical tools borrowed from economics and statistics. More specifically, the thesis has the following structure.

Chapter 1's starting point is Samir Okasha's 2011 paper, *Theory Choice and Social Choice: Kuhn versus Arrow*, and the literature that sprung up around it. In that paper, Okasha applies results from social choice, i.e. the study of how to reach collective decisions, to epistemology. Particularly, he looks at scientific theory choice on the base of Kuhn's epistemic criteria from the point of view of Kenneth Arrow's impossibility theorem (1951). And, by this move, he concludes that there is no acceptable algorithm for theory choice, in that there is no theory choice rule that satisfies analogues of Arrow's conditions. In other words, rational theory choice is impossible. The severity of such a conclusion urges Okasha to look for escape routes from impossibility. After different failed attempts, he found one in Amartya Sen's work for social choice, which Okasha applies to theory choice. In a nutshell, if theory choice criteria permit inter-criterion comparisons of the kind 'the accuracy of theory T_1 is less than the simplicity of T_2', then there are theory choice rules that meet analogues of Arrow's conditions. However, from considerations made by Stegenga (2015) and Okasha himself (2015), it turns out that 'large scale' theory choices, i.e. choices among 'key theories' which imply a change of paradigm, based on Kuhn's criteria cannot escape from impossibility. The reason is that such criteria do not seem to allow for inter-criterion comparisons. In this chapter, I propose two counterarguments to such a conclusion. The first one suggests, by making an example, that inter-criterion comparability may be possible when Kuhn's criteria are used in the aforementioned kind of choices. The second one, instead, proposes to use Kuhn's values to evaluate the prior probability and the likelihood. The latter, in fact, are 'criteria' of the 'Bayesian theory choice rule', which is not subject to Arrovian impossibility thanks to Sen's escape route. However, an objection to the second counterargument may arise. According to Stegenga (2015), a rule that translates the information provided by Kuhn's criteria in prior probabilities suffers from Arrovian impossibility since it should satisfy Arrow's conditions. If this is true, then the second counterargument proposed vanishes as Arrovian impossibility obtains anyway in using Kuhn's criteria to determine the prior probability and the likelihood. I show that this is not the case as the algorithm Stegenga has in mind has a different mathematical form from Arrow's social choice rule. Namely, the former outputs a list or real number, while the second a ranking of theories. So, we cannot ask whether the rule Stegenga considers should satisfy or not Arrow's conditions. Finally, I consider the question, triggered by Stegenga's observation, whether the Arrow-style rule, associated with Stegenga's algorithm, suffers from Arrovian impossibility. To answer it, I introduce an algorithm which I call 'Prior Probability Rule' (PPR). PPR assumes as input the information about the theories provided by (some of) Kuhn's criteria and outputs an order of the theories by their decreasing prior probability assigned to them by Stegenga's algorithm. I highlight that an observation Morreau (2015) made against Okasha applies also to our algorithm, which lead us to suspend the judgement about its impossibility.

In chapter 1, the second counterargument is ultimately based on the relationship epistemic values entertain with the prior probability and the likelihood of the 'Bayesian theory choice rule'. In **chapter 2**, I keep following this path, and I explore the relationship between Bayesianism and an epistemic value not considered in chapter 1, i.e. 'explanatory power' which refers to the goodness of an explanation. Specifically, I focus on an inference where the explanatory value of a hypothesis for a set of phenomena obtains special status, that is the 'Inference to the Best Explanation' (IBE) or 'Abduction'. And, I consider how IBE relates to a problem which has been haunting the Bayesian theory of confirmation since Glymour first described it in 1980, namely, the Problem of Old Evidence (POE). POE states that Bayesian confirmation theory cannot account for a shared intuition of scientific reasoning according to which a theory H can be confirmed by a piece of evidence already known, i.e. by an old piece of evidence. Different dimensions of POE have been highlighted, and different solutions have been proposed to each of them. Here, I consider only two dimensions, i.e. the dynamic and static dimension. In the former, we want to explain how the discovery that H accounts for E confirms H. In the latter, we want to understand why E is and will be a reason to prefer H over its competitors. My aim, in chapter 2, is threefold. First, I point out a technical shortcoming of the second of two recent solutions to the dynamic dimension of POE, proposed by Eva and Hartmann (2020). Second, I highlight that these solutions can be read in terms of Inference to the Best Explanation (IBE). On this base, I further gauge the weaknesses and strengths of the two models. Namely, I show that Eva and Hartmann endorse a particular understanding of IBE, and I stress that it is still an open question whether it is the one descriptively used in the scientific practice. Moreover, I contend that, while one condition of their first model is not expression of such an understanding, the only condition used in their second model is. Third, I focus on the static dimension of POE which, now, has to be expressed in IBE terms. To solve it, I rely on the counterfactual approach, and on the Bayesian IBE, a probabilistic version of IBE in which explanatory considerations help to evaluate the terms in Bayes' theorem. However, it turns out that the problems of the counterfactual approach recur even when it is used to solve the static POE in IBE terms.

So, by focusing on making explicit that cases of confirmation by old evidence are instances of abduction or IBE, chapter 2 is about a sense of abduction in which the explanatory power of a hypothesis is used for *justifying* the hypothesis itself. Thus, the proper place of this sense of abduction is the so called *context of justification*. However, in the philosophical literature, the term 'abduction' refers also to the use of explanatory power in *generating* hypotheses, and, as such, it belongs to the *context of discovery*. Abduction so intended will be the protagonist of **chapter 3**. It was especially defended by Charles S.

Peirce and Norwood R. Hanson (1958, 1960, 1965), who both believed that it could be a logic, with its own precise pattern. However, it is now commonly thought that their (very similar) logics make better sense if understood, not as logics of generation, but as logics of adoption of new hypotheses which appear to be worthy candidates for further investigations. A view of abduction very much in this spirit is endorsed by Gerhard Schurz (2008) who, indeed, thinks that abduction is "a *search strategy* which leads us, for a given kind of *scenario*, in a reasonable time to a most promising explanatory conjecture which is then subject to further test" (p. 205, emphasis in the original). This general definition encompasses different patterns, which Schurz detects in various contexts ranging from common sense to philosophy and science. In chapter 3, I carry on such a descriptive analysis in the context of the ongoing Covid-19 pandemic. The latter is, in fact, a fertile ground for search strategies as Schurz describes them, given the novelty it puts us in front of.

Chapter 1

Theory Choice and Social Choice: Two Proposals to Escape from Arrovian Impossibility for 'Large Scale' Theory Choices Based on Kuhn's Criteria

1.1 Introduction

Samir Okasha (2011) applies Kenneth Arrow's impossibility theorem (1951) for social choice to scientific theory choice on the base of the five criteria identified by Thomas Kuhn (1977, p. 321). From this move, he concludes that there is no acceptable algorithm that relies on them. That is, there is no theory choice rule that satisfies the four Arrovian conditions.

However, this conclusion seems to be in contrast with Kuhn's own views (1970, 1977). According to Okasha, Kuhn contends that the fact that different criteria can pull in different directions is a reason to expect a plurality of acceptable algorithms, each of which weights the criteria differently.

Okasha sees the only promising escape route from impossibility in enlarging the information that the theory choice rule takes in, following Amartya Sen's work for social choice (1970, 1977, 1986). Briefly, if criteria of theory choice permit some form of inter-criterion comparability (e.g. if it is meaningful to say statements like 'the accuracy of theory T_1 is less than the simplicity of T_2'), then the rule that trades them off satisfies analogous Arrovian conditions.

But such an escape route raised some concerns. In fact, Stegenga (2015) highlights, and Okasha himself (2015) agrees, that it has a narrow domain of application so that impossibility can be rarely escaped. This is due to the fact that the majority of criteria do not allow for inter-criterion comparisons. Specifically, it seems that this is the case for large scale theory choices that rely on Kuhn's criteria, i.e. choices among key theories that implies a change of paradigm.

The aim of the chapter is to propose two solutions to overcome these concerns. The first one investigates if it is really the case that when Kuhn's criteria are used in the aforementioned kind of choices, then no inter-criterion comparability is possible. By proposing an example, we will see that the contrary may be true.

The second one proposes to use Kuhn's values to evaluate the prior probability and the likelihood, which are, in turn, 'criteria' of the 'Bayesian theory choice rule'. In fact, the latter is not subject to Arrovian impossibility because of Sen's escape route (Okasha 2011, pp. 107-108). To this end, some meanings of Kuhn's criteria will be considered and it will be shown how they relate to the prior probability and the likelihood. It will turn out that while accuracy is reflected by the likelihood, (some of) the other criteria bear on the prior. Since the latter can pull in different directions, the same moral Okasha draws from Kuhn's work applies here. That is, there are many acceptable algorithms to the evaluate of the prior probability.

At this point, an objection may arise. Stegenga (2015) argues that an algorithm that somehow translates the information provided by Kuhn's criteria in prior probabilities should satisfy Arrow's conditions. Thus, it suffers from the same impossibility. If this is true, then the second solution proposed vanishes as Arrovian impossibility obtains anyway in using Kuhn's criteria to determine the prior probability and the likelihood. However, I will highlight that Stegenga's claim is flawed as the algorithm he has in mind has a different mathematical form from Arrow's social choice rule. Namely, the former outputs a list or real number, while the second a ranking of theories.

So, we cannot ask whether conditions defined for the Arrovian social choice rule should apply to Stegenga's algorithm.

That said, Stegenga's observation raises an interesting question: does the Arrow-style rule, associated with Stegenga's algorithm, suffer from Arrovian impossibility? To answer this question, I will consider an algorithm which I will call 'prior probability rule' (PPR). The latter assumes as input the information about the theories provided by (some of) Kuhn's criteria and outputs an order of the theories by their decreasing prior probability assigned to them by Stegenga's algorithm. We will see that an observation Morreau (2015) made against Okasha applies also to our algorithm, which will lead us to suspend the judgement about its impossibility.

To show my claims, I will follow this structure. In section 1.2, I will take into consideration Okasha's argument by closely following his 2011 paper, where this argument is presented. Section 1.3 will be dedicated to expound the two solutions to save from impossibility large scale theory choices relying on Kuhn's criteria. Section 1.4 will consider Stegenga's objection. In this regard, I will explain why it does not undermine the second solution proposed, and consider if the Arrow-style rule associated with Stegenga's algorithm is subject to Arrovian impossibility. Section 1.5 concludes.

1.2 Okasha's argument

Okasha's reflection starts by considering a famous argument made by Kuhn in the Postscript to *The Structure of Scientific Revolution* (1970), and developed in the essay 'Objectivity, Value Judgement, and Theory Choice' (1970). That is, even if the five epistemic criteria are the objective basis for scientific theory choice, they do not uniquely determine the latter. One of the reasons is that scientists can disagree about importance of the criteria, i.e. about their weights, when they pull in different directions. Okasha (2011, p. 86) expresses this point by saying that, according to Kuhn, there are *many* acceptable algorithms for theory choice, each of which weights the criteria differently.

However, a totally different argument can be made, namely that there is *no* acceptable algorithm for theory choice based on Kuhn's epistemic criteria[1], i.e. there is no algorithm that meets reasonable conditions. Okasha arrives at this conclusion by applying Kenneth Arrow's impossibility theorem (1951) for social choice to theory choice. Let us briefly see what his argument consists of.

An important part of social choice theory deals with the analysis of aggregating individuals' preferences over a set of alternatives (e.g. election candidates) into a single social preference. The basic model of preference aggregation is as follows. We start with a set $N = \{1, 2, \ldots, n\}$ of individuals ($n \geq 2$) and a set $X = \{x, y, z, \ldots\}$ of social alternatives. Each individual $i \in N$ has a *weak preference ordering* R_i over the alternatives, namely a complete and transitive binary relation on X[2]. For any x and $y \in X$, xR_iy means that individual i weakly prefers alternative x to y, that is either xP_iy (i strictly prefers x to y), or xI_iy (i is indifferent between x and y)[3]. Informally, a weak preference ordering is a ranking of the alternatives from best to worst with ties permitted. A list of weak preference orderings, each for every individual in society, is called a *profile*, and it is denoted $<R_1, R_2, \ldots, R_n>$. We would like to construct a single social weak preference ordering R that somehow depends on the information provided by the profile.

Arrow proposes to approach this problem by considering a function, a 'social choice rule'[4], that takes as input a profile of weak preference orderings, and outputs a social weak preference ordering R. Then, he asks which conditions this function should satisfy. He comes up with the following four conditions:

- Unrestricted domain (**U**), according to which the domain of the social choice rule should include every list $<R_1, R_2, \ldots, R_n>$ of n weak preference orderings of X.
- Weak Pareto (**P**) which says that for any profile, and any alternatives x and y, if every individual i strictly prefers x to y, so should society.

---

[1] Notice that Okasha's argument is not restricted to Kuhn's list of criteria, but it may also be applied to other lists, like, for instance, McMullin's one (1982).

[2] Completeness requires that for any $x, y \in X$, xR_iy or yR_ix. Transitivity requires that for any $x, y, z \in X$, if xR_iy and yR_iz, then xR_iz.

[3] xP_iy if xR_iy and not yR_ix, and xI_iy if xR_iy and yR_ix.

[4] Arrow called his function 'social welfare function' but, since this expression is often used in another sense today, Okasha calls it 'social choice rule'.

- Non-dictatorship (**N**), according to which there must not be any dictator, that is a person d such that if, for any alternatives x and y and for any profile $<\ldots, R_d, \ldots>$, d strictly prefers x to y, so should society.
- Independence of irrelevant alternatives (**I**) that says that the social choice between alternatives x and y must depend only on individuals' preferences between x and y, and not on their preferences over other alternatives. This condition is captured by saying that if each individual's preference for x over y is the same in two profiles, then the social choice rule, when applied to both profiles, must yield the same social preference for x over y.

Remarkably, Arrow showed that, so long as there are at least three alternatives to choose among, these four conditions are incompatible, that is there is no social choice rule that jointly satisfies them.

Now, according to Okasha, Kuhn's weighting problem and Arrow's aggregation problem are formally analogous. To see this, just consider each Kuhnian criterion as an individual with their weak preference ordering over the competing scientific theories. That is, each criterion ranks the theories according to how well they satisfy it. Then, consider a 'theory choice rule' that, given a profile of weak preference orderings, each for every criterion of theory choice, yields an overall ranking of the competing theories. Next, we ask whether such a theory choice rule satisfies the Arrovian conditions rephrased for theory choice. The answer seems to be affirmative (Okasha 2011, pp. 92-93). Here is why.

- Unrestricted domain (**U**) states that the domain of the theory choice rule should include every list $<R_1, R_2, \ldots, R_5>$ of the five criteria's weak preference orderings of X. That is, there should be no *a priori* restriction on the preferences that criteria are allowed to have. This condition seems to be reasonable: since there is no reason to think that trade-offs (or correlations) between criteria must *always* obtain, then *a priori* we should include in the domain all the possible rankings. It may be that, in the *actual* situation, such trade-offs or correlations are the case. But, not knowing this in advance, as it is not something that always happens, we cannot exclude even one possible ranking.

- Weak Pareto (**P**) is an obvious condition. Indeed, it says that if theory T_1 does better than theory T_2 by each of Kuhn's criteria, then it should be preferred overall.

- According to Non-dictatorship (**N**), there must not be any criterion such that if T_1 is ranked above T_2 by that criterion, then T_1 is automatically above T_2 in the overall ranking. Such a condition makes good sense as long as we agree that all the criteria are relevant to theory choice.

- Independence of irrelevant alternatives (**I**) says that the overall ranking of T_1 and T_2 should depend only on how criteria rank T_1 and T_2, and not on how they rank other theories. This condition is captured by saying that if each criterion's preference for T_1 over T_2 is the same in two profiles, then the theory choice rule, when applied to both profiles, must yield the same preference for T_1 over T_2 in the overall ranking. **I** has an intuitive appeal for theory choice as it has for social choice. Indeed, it is highly intuitive that, say, in an election to determine whether the Labour or Tory candidate is socially preferable, individual's preferences involving the Liberal candidate should not matter. And the same goes for theory choice.

Since the four Arrow's conditions apply to theory choice as well, we have an analogous impossibility result: if there are at least three alternatives to choose among, there exists no theory choice rule that satisfies the four conditions. In other words, rational theory choice is impossible.

Given the severity of such a conclusion, Okasha looks for escape routes. After having considered some possible solutions that he rules out (section 5), he turns to Amartya Sen's 'informational basis' approach for social choice (1970, 1977, 1986). The latter is a complex framework of which I will here present the gist. For a more complete exposition of Sen's work, see List 2013, section 4. Sen's reflection starts by observing that the input of Arrow's social choice rule, i.e. a profile of weak preference orderings, provides poor information. Firstly, weak preference orderings are ordinal, namely they contain no information about intensity of preference. From the fact that individual i prefers x to y to z, we cannot say if their preference for x over y is greater or less than their preference for y over z. Secondly, weak preference orderings do not permit interpersonal comparisons. For example, we do not know if individual 1 is better off in alternative x than individual 2. To incorporate such richer information, Sen generalizes Arrow's framework by proposing to start with a profile of utility functions, one for each individual in society, denoted $<u_1, \ldots, u_n>$. An individual's utility function assigns to each alternative a real number which represents how much utility that alternative brings to the individual in question. Then, he defines a function, the social welfare functional (SWFL), which has the same output of the Arrovian social choice rule, a weak social preference ordering R. Its input, however, is different: it is no longer a profile of weak preference orderings, but a profile of utility functions which encodes more information. In fact, while a utility function on X induces a weak preference ordering, the converse is not true. How much of this information is used depends on the assumptions we impose on the SWFL. Let us see why. Sen imposes on the SWFL the four Arrovian conditions rephrased in terms of utility functions - **U′**, **P′**, **N′**, **I′** (see List 2013, section 4.2) -, but we do not have Arrovian impossibility by default. To this end, we have to impose on the SWFL an additional condition called **ONC** ('ordinal measurability with no interpersonal comparability'). **ONC** captures the fact that the SWFL takes into consideration only ordinal and not interpersonally comparable information as the Arrovian social choice rule does. It does so by considering two profiles $<u_1, u_2, \ldots, u_n>$ and $<v_1, v_2, \ldots, v_n>$ as informationally equivalent[5] whenever, for each $i \in N$, $v_i = \varphi_i(u_i)$, where φ_i is some positive monotonic transformation, possibly different for each individual. So, when applied to these profiles, SWFL yields the same weak social preference ordering. When $<v_1, v_2, \ldots, v_n>$ is obtained from $<u_1, u_2, \ldots, u_n>$ in the aforementioned way, the only meaningful information is ordinal and not interpersonally comparable. As a matter of fact, this is the only information that remains invariant across the two profiles as the individual utility functions of $<u_1, u_2, \ldots, u_n>$ are arbitrarily monotonically transformed. Sen then asks what happens if we substitute **ONC** with weaker conditions. There are two ways in which it can be weakened: (i) measure utility on a more-than-ordinal scale; (ii) permit interpersonal comparisons. To obtain (i), we restrict the trans-

---

[5] Two profiles are informationally equivalent when they contain the same information.

formations that are held to preserve information. Thus, we might hold utility to be measured on a *cardinal* scale so that only positive linear transformations of the form $v_i = \mathrm{a}u_i + \mathrm{b}$, $\mathrm{a} > 0$ preserve information. This means that utility differences are meaningful. As an alternative, we might consider utility to be measured on a *ratio* scale, namely solely transformations of the kind $v_i = \mathrm{a}u_i$, $\mathrm{a} > 0$ preserve information. This means that utility ratios are meaningful, as utility has a natural zero point. Finally, we might hold that utility is measured on an *absolute* scale, that is, only the identity transformation preserves information. This means that actual utility numbers are meaningful. Once a scale for utility has been chosen, we have to make a decision about interpersonal comparability. Utility is *non-comparable* when each individual can apply a transformation, from the permissible class, independently from the other. Conversely, utility is *fully comparable* when each individual has to apply the same transformation. Depending on the scale, some form of *partial* comparability is also possible. For instance, if utility is measured on a cardinal scale and it is *unit* comparable[6] then individuals' positive linear transformations must have the same slope (a) but can have different intercepts (b). By these moves, numerous alternatives to Arrow's **ONC** can be obtained, such as: cardinal-scale utility with no comparability (**CNC**), cardinal-scale utility with full comparability (**CFC**); ratio-scale utility with full comparability (**RFC**); ratio-scale utility with no comparability (**RNC**); absolute-scale with full comparability (**AFC**). Now, Sen proves that impossibility can be avoided only if some interpersonal comparison is allowed. Thus, if **ONC** is replaced with **CNC** or **RNC**, then we fall back into impossibility. However, if **ONC** is replaced with **CFC**, **RFC** or **AFC**, then there are SWFLs that satisfy the analogous Arrovian conditions plus an alternative to **ONC**. Thus, we see that Arrow's impossibility is mostly due to the meagre information his social choice rule assumes. Conversely, given enough information, social choices that satisfy the analogues of the Arrovian conditions are possible.

After having exposed Sen's social choice framework, Okasha applies it to theory choice (Okasha 2011, section 7). Accordingly, we

---

[6] Utility is unit comparable when it is meaningful to say statements like 'individual 1 prefers a switch from x to y more that individual 2 prefers a switch from x to y'.

start with a profile of 'utility functions'[7] which are real-valued representations of the weak preference orderings of each criterion. Then, we define a 'theory choice functional' that takes such a profile as input and outputs an overall ranking of the theories. Now, we need to know which conditions the theory choice functional should satisfy, namely which profiles it should consider as informationally equivalent. This issue needs to be addressed both for measurement scales and 'inter-criterion' comparability. As for the first point, Okasha notices that for some criteria of theory choice an ordinal scale might be appropriate. An example of this circumstance is Kuhn's criterion of fruitfulness for which differences in fruitfulness cannot be compared, i.e. they are not meaningful. However, in certain contexts we can go beyond ordinal measurement. For instance, in the statistical model selection domain, the simplicity of a hypothesis is the number of free parameters it contains. That is, the simplicity of hypothesis '$y = ax^2 + b$' is the integer 2, while the one of '$y = ax^2 + bx + c$' is the integer 3, and so on. So, here simplicity is measured on an absolute scale as the actual numbers are meaningful, thus only the identity transformation preserves information. As for inter-criterion comparisons, which are needed to escape from impossibility, Okasha (*ibid.*, p. 104) observes that they seem unlikely. In fact, he says, it is hard to see which the basis of judgements like 'the accuracy of T_1 is less than the simplicity of T_2' might be. However, one thing to notice is that when all criteria are absolutely measurable, then inter-criterion comparability becomes straightforwardly meaningful in the technical sense of invariance under permissible transformations. In fact, if the utility functions that represent simplicity and accuracy cannot be transformed without loss of information, then neither can statements like 'the accuracy of T_1 is less than the simplicity of T_2'. Thus, we see that when all criteria are absolutely measurable, the theory choice functional that trades them off satisfies **AFC**. However, there are many algorithms that satisfy the analogous Arrovian conditions plus an alternative to **ONC** that allows for inter-criterion comparability. Thus, the informational

---

[7] To stress the formal analogy between Sen's framework and the theory choice problem, Okasha refers to the real-valued representations of the weak preference orderings of each criterion as 'utility functions', even if they are not. However, as he himself (2011, p. 102, n 18) mentions, Kelsey (1987) points out that the functions in an SWFL do not necessarily have to be considered as utility functions.

basis approach leads us back to Kuhn's argument: there are many acceptable algorithms for theory choice.

1.3 Narrow domain of application of Sen's escape route: two solutions

A general moral can be drawn from the application of Sen's escape route to theory choice considered at the end of the previous section. That is, criteria have to be measured on a scale that implies a technically meaningful form of inter-criterion comparability[8]. As seen before, this is the case when all criteria are measured on an absolute scale.

Such a way of applying Sen's escape route to theory choice raised some concerns. Specifically, Stegenga (2015, p. 272) highlights that it has a narrow domain of application so that impossibility can be rarely escaped. This is due to the fact that the majority of criteria are not measured on an absolute scale. Indeed, Stegenga argues (*ibid.*, section 4) that ubiquitously criteria provide ordinal information about the theories. He bases his claim essentially on two considerations. First, some criteria are *always* measured on an ordinal scale, like fruitfulness, as already noticed by Okasha. Second, *mostly* some criteria are measured on an ordinal scale, as it is the case for simplicity. In fact, he agrees with Okasha that in some contexts simplicity can provide more than ordinal information, like in the statistical model selection domain. However, this is a narrow domain that does not cover many interesting cases in science where there are no parameters to count in the theories. Examples are the choice between an eccentric or epicyclic structure of solar motion or whether genes are composed of proteins or of DNA.

Actually, Stegenga's observations about simplicity being in some cases only ordinally measurable and in some others more than ordinally do not take Okasha (2011, p. 282) by surprise: "I agree with

---

[8] Actually, Okasha (2011, p. 104) mentions another way to apply Sen's escape route to theory choice that does not require a technically meaningful form of inter-criterion comparability, namely when criteria are each measured on their own ratio-scale (i.e. **RNC**). This, in fact, permits a limited form of inter-criterion comparability, and avoids Arrovian impossibility as long as 'utilities' are non-negative. However, since the following discussion can be applied also to this case, for brevity, I did not consider it.

this; indeed I said as much in my own paper". As a matter of fact, in Okasha 2011 (p. 103), we read:

> "It may be that for the 'large scale' theory choices that Kuhn was interested in, ordinal comparisons are all that can be achieved. But it seems clear that in other more humdrum cases, particularly where the problem may be formulated statistically, we have much more than ordinal information at our disposal".

Thus, from these considerations, it seems that when we go beyond the statistical domain (e.g. statistical model selection), and we try to choose among different key theories[9], e.g. different models of solar motion, Kuhnian criteria can provide only ordinal information. So, a question arises: is there a way to escape from impossibility for large scale theory choices relying on Kuhnian criteria? It seems to me that two answers can be given, which I will consider in turn, but before let me explain why we should care about this rescue.

As Okasha (2011, p. 94) observes, choices among key theories *tend* to be binary (e.g. geocentrism versus heliocentrism, oxygen versus phlogiston, and so on), and so no Arrow's theorem applies to them. In fact, as pointed out in section 1.2, Arrovian impossibility obtains if there are at least three alternatives to choose among. But, as Okasha's observation suggests, this may not always be the case. Moreover, even if pragmatically large scale theory choices are just binary, we may be interested in how to achieve a rational evaluation of different key theories of the same domain based on Kuhn's criteria. Arguably, this is what motivated Stegenga's worry that Sen's escape route does not apply to choices among keys theories.

Let us go back now to the two answers to the aforementioned question. The first one investigates whether it is really the case that for large scale theory choices, the criteria isolated by Kuhn are measured only on an ordinal scale. For example, let us consider fruitfulness, the ordinally measurable criterion *par excellence*. In particular, let us

---

[9] By this term, I refer to the theories at the foundations of the paradigm, i.e. the applications of these theories in the solution of important problems, along with the new experimental or mathematical techniques employed in those applications (see Bird 2018, section 3). Thus, choosing among these theories involves a paradigm change. This is why such choices are on a large scale, as Okasha says in the quote above.

focus on one meaning of fruitfulness that may be employed in such choices, i.e. a fruitful theory correctly predicts surprising phenomena (e.g. W. Salmon 1990, p. 199). (I specify that I take into consideration one particular meaning of fruitfulness because Kuhnian criteria are ambiguous, namely they can be understood in different ways. Indeed, this is the other reason why, together with the weighting problem, these criteria do not uniquely determine theory choice (Kuhn 1977, p. 324)[10]. One way to measure such a criterion could be the following. Let us consider a set X = $T_1, \ldots, T_n$ of competing theories and a set Y = $e_1, \ldots, e_m$ of occurred surprising phenomena. Given that the more surprising the phenomenon is, the less probable it is (e.g. Earman 1992, p. 64), we can measure the surprising character of the latter, i.e. $S(e_i)$, in the following way:

$$S(e_i) = 1 - P(e_i) \tag{1.1}$$

Thus, we can define the fruitfulness of a theory as the weighted sum of the phenomena's probability, namely:

$$F(T_k) = \sum_{i=1}^{m} (1 - P(e_i)) f(T_k e_i), f(T_k e_i) = \begin{cases} 1 & \text{if } T_i \text{ predicts } e_i \\ 0 & \text{if } T_i \text{ does not predict } e_i \end{cases} \tag{1.2}$$

So, ultimately, the fruitfulness of a theory is measured by the sum of probabilities representing the surprising character of the phenomena it predicts. Now, since probabilities are measured on an absolute scale as the actual numbers are meaningful[11], fruitfulness as defined by $F(T_k)$ is measured on an absolute scale too. Moreover, a similar reasoning might lead us to conclude that an absolute scale is appropriate also for other meanings of the Kuhnian criteria which we may employ, along with the aforementioned understanding of fruitfulness, to choose among theories. If this is the case then, as pointed out in section 1.2, inter-criterion comparability becomes technically meaningful and impossibility can be escaped.

---

[10] However, Okasha argues that the ambiguity problem collapses into the weighting one (see Okasha 2011, p. 85).

[11] Probabilities are measured on an absolute scale if we stick to the convention of assigning probability one to the sure event. Otherwise, they are measured on a ratio scale.

Some results in the literature seem to support the possibility of solutions on these lines. For instance, Bovens and Hartmann (2003, p. 612) propose a probabilistic measure of coherence of an information set, i.e., a set of propositions that we have acquired through gathering of information. And they propose to apply this measure to Kuhn's internal consistency, which they interpret as a sort of coherence of an information set (p. 602). This interpretation is based on a specific view of theories. But if this view applied to key theories, and if some other assumptions were met (see p. 628), then we could measure internal consistency using probabilities, meaning that it would be absolutely measurable. However, Bovens and Hartmann warn us that there could be theories that cannot be compared using their measure. So, it could be the case that some key theories could not be compared in terms of internal consistency. In other words, internal consistency would not yield a complete preference ordering, and thus it would not fit Arrow and Sen's formal apparatus in the first place. Be it as it may, Bovens and Hartmann's important result still suggests that a cardinal measure of consistency is possible.

That said, it is still a work in progress to ascertain whether other meanings of the Kuhnian criteria may be measured on an absolute scale. Thus, in the meantime, I will assume that they can provide only ordinal information, and I will see if another answer is possible. Indeed, it is. It relies on two claims: (i) the Bayesian theory choice functional, whose 'criteria' are the prior probability and the likelihood, is not subject to Arrovian impossibility; (ii) thus, one can use Kuhn's values to determine the terms in Bayes' theorem. Let us see in more details these two points.

According to orthodox Bayesianism, the hypothesis' probability in light of the new evidence should be revised following the principle of conditionalization. That is, given a set $X = \{T_1, \dots, T_n\}$ of mutually exclusive and jointly exhaustive hypotheses[12], when we learn evidence E for certain:

$$P'_E(T_j) = P(T_j|E) = \frac{P(T_j)P(E|T_j)}{P(E) = \sum_{i=1}^{n} P(T_i)P(E|T_i)} \tag{1.3}$$

---

[12] Respectively, a set of hypotheses that cannot be true at the same time and where at least one of them must be true.

$P'_E(T_j)$ is the hypothesis' probability *after* we have learned E, $P(T_j|E)$ is the hypothesis' probability *before* E is actually learned, called 'posterior probability'. The latter is expressed by means of Bayes' theorem, in which P(T) is the prior probability of the hypothesis, while P(E|T) is the likelihood. These two terms assign to each hypothesis a real number belonging to the interval [0; 1] measuring respectively the truth (or acceptability) of the hypothesis before the evidence, and how probable the evidence is in light of the hypothesis[13]. However, for our purpose, we do not need the entire Bayes' theorem, but only its numerator, i.e. $P(T_j)$ x $P(E|T_j)$. In fact, given competing theories $T_1, \ldots, T_n$ and a body of evidence E, we are not interested in the exact value of their posterior probabilities, but only in their comparison. Thus, we can ignore the denominator of Bayes' theorem, which does not depend on any particular theory. In what follows, I will call the quantity $P(T_j)$ x $P(E|T_j)$ 'Bayesian algorithm for theories' comparison'.

To see how Bayesianism translates in terms of Sen's social choice framework, consider $P(T_i)$ and $P(E|T_i)$ as 'utility functions'. Then, encode the information they provide about the theories in the profile <$P(T_i)$, $P(E|T_i)$>. Next, consider the function that takes such a profile as input and outputs the ranking of the theories generated by the quantity [$P(T_i)$ x $P(E|T_i)$]. This is our theory choice functional. Let us call it 'Bayesian theory choice functional' (BCF) and ask if it satisfies the four Arrovian conditions. BCF meets conditions **P′**, **I′**, **N′**. Respectively, if T_1 has a higher prior probability and a higher likelihood than T_2, then it is ranked higher by BCF. Whether T_1 or T_2 is ranked higher by BCF is exclusively determined by the priors and likelihood of those two theories. Neither the prior nor the likelihood can dictate over the other. Conversely, BCF does not satisfy condition **U′**. Its domain does not contain all possible profiles, i.e. pairs of real-valued functions: as highlighted before, the prior probability and likelihood functions can only take on values in the interval [0; 1]; moreover, it is required that $\sum(T_i) = 1$.

Thus, BCF jointly satisfies conditions **P′**, **I′**, **N′**. Why is that? Two reasons can be given. First, Arrow's theorem, which in Sen's framework requires condition **ONC**, does not apply because BCF does not satisfy condition **U′**. Second, even if BCF does not meet **U′**, Arrow's

---

[13] For a complete exposition of orthodox Bayesian approach to scientific inference, see the introduction of Stephan Hartmann and Jan Sprenger's book *Bayesian Philosophy of Science* (2019).

theorem applies nonetheless, as the restricted domain retains enough variety for what is essentially Arrow's result. And BCF escapes it thanks to the fact that it satisfies condition **AFC** because probabilities are measured on an absolute scale as the actual numbers are meaningful. Now, Okasha proves (2011, Appendix) that, with the domain restriction appropriate to BCF, Arrow's theorem applies. Thus, the reason why BCF jointly satisfies **P′**, **I′**, **N′** is that it escapes Arrow's result by satisfying **AFC**.

As for the second point, we have to investigate how Kuhn's criteria relate to the terms in the Bayesian algorithm for theories' comparison. But, in order to do that, first, we need to consider more closely what these five criteria consist of. This is not an easy thing to do because, as mentioned before, these criteria are ambiguous. Thus, in what follows, I will present them relying on what I think are the most common understandings of these criteria.

Accuracy, Kuhn (1977, p. 321) says, is the agreement between the consequences deducible from a theory and the results of existing experiments and observations. As for consistency, this criterion has two dimensions: an internal and an external one (*ibid.*, pp. 321-322). According to the former, a theory should not have internal contradictions, while the latter states that it should be compatible with other accepted theories. Regarding simplicity, Kuhn points out that a simple theory brings order to phenomena that otherwise would be individually isolated and, as a set, confused (*ibid.*, p. 322). However, on this base, there are still lot of ways in which a theory can be simple. Longino (1996, p. 43) believes that the most common understanding of simplicity is ontological. In fact, every theory stipulates an ontology, i.e. entities and processes, and the simpler theory is the one that stipulates the fewer number of them. Newtonian mechanics then, which applies to all that bodies characterized by extension, hardness, impenetrability, mobility and inertia, is simpler than Aristotle's one which stated that there were four (sublunary) elements, each with distinctive properties. Let us now consider fruitfulness. Kuhn (1977, p. 322) says that a fruitful theory discloses new phenomena or previously unnoted relationship among those already known. Again, this criterion has many aspects. As we have seen earlier in this section, one sort of fertility involves the correct predictability of surprising phenomena. As such, this understanding of fruitfulness comprises two criteria: correct prediction, i.e. accuracy, and the surprising character of what is predicted. Finally, as of scope, Kuhn (*ibid.*) explicates it by saying that: "a theory's consequences should extend far beyond the particular observations, laws, or subtheories it was initially designed to explain". This item also can be characterized in various ways. One of these could be the diversity of phenomena that can be explained by the basic principles of a theory. Again, Newton's mechanics exemplifies this criterion, since many theretofore different phenomena (from falling bodies to orbiting planets) fell into the range of the three laws of motion.

Let us see now how these criteria relate to the terms in the Bayesian algorithm for theories' comparison. To address this issue, the first thing to notice is that, as Longino (1996, p. 41) points out, Kuhn's items overlap with philosophers' lists of 'epistemic" values, the constant being accuracy (e.g. Quine and Ullian 1978; Longino 1990; Laudan 1990). That is, they are considered as counting for the truth (or acceptability) of hypotheses. On this base then, those Kuhnian criteria which do not make reference to the agreement with the evidence are suitable to be reflected by the prior probability, which, as seen just now, represents the probability of the hypothesis' truth (or acceptability) *before* the evidence. Internal and external consistency, simplicity, and scope are of this kind. Whether a theory is more internally or externally consistent than another, simpler or if it has a broader scope does not depend on its agreement with observations or experiments. Rather, it has to do with *inner* features of the theory such as how it is formulated, how it relates to accepted theories, how many principles it postulates to explain a given phenomenon, and how broad is the range of such principles. However, this is not the case for accuracy and fruitfulness. Accuracy, indeed, *is* the agreement between the consequences deducible from a theory and the evidence, and its proper place is the likelihood. To see this, consider the following historical example. Copernican cosmology (T_1) implies the Venus' phases (evidence E), whereas Ptolemaic system (T_2) the opposite ($\neg$E). Then, Galileo observed the phases. So, according to the accuracy criterion, since T_1's consequence agrees with the observation, but T_2's one does not, T_1 is more accurate than T_2. Now, in Bayesian terms, since E is observed, we have to apply the principle of conditionalization with respect to E. Given that T_1 implies E, $P(E|T_1) = 1$, whereas

T_2 implies the negation of E, so $P(E|T_2) = 0$. That is, when the theory's consequence agrees with the observation, the likelihood is 1. Conversely, when the theory's consequence does not agree with the observation, the likelihood is 0. Let us now consider fruitfulness. As it was highlighted before, the understanding of fruitfulness considered can be seen as a combination of two criteria: accuracy and the surprising character of what is predicted. The accuracy part, as argued just now, is reflected by the likelihood. The surprising character of the evidence, on the other hand, is captured by a low value of P(E), where E is the body of surprising phenomena that the theory in question correctly predicts. And it can vary from theory to theory depending on which surprising phenomena the theory under consideration correctly predicts. The Bayesian algorithm for theories comparison cannot capture this sense of fruitfulness. As we have seen, in the algorithm E is not the body of surprising phenomena, which changes for every theory, but rather it is the total body of evidence under consideration, which means that E and consequently P(E) are the same for every theory: thus, P(E) can be ignored to compare theories. That said, my aim is not to provide an escape from Arrovian impossibility for large scale theory choices based on *all* possible interpretations of Kuhn's criteria. Rather, it is to show one way out of the impossibility, however incomplete. For this reason, in what follows, my discussion will not consider fruitfulness.

To sum up, so far, we have seen that four of the Kuhnian criteria can bear on the prior probability, i.e. consistency (in its two dimensions), simplicity, scope. Accuracy, on the other hand, is reflected by the likelihood.

Now, since the four criteria that bear on the prior can pull in different direction, we can say for the evaluation of the prior probability the same thing Okasha says about Kuhn's argument. That is, there can be many acceptable algorithms to determine the prior probability, each of which weights the criteria differently. This consideration leads us directly to the next section.

1.4 Stegenga's objection

In the second part of section 1.3, we have seen a way to escape from impossibility for large scale theory choices based on (some of) the

Kuhnian criteria. The strategy consisted of two steps. First, it was highlighted that the Bayesian theory choice functional is not subject to Arrovian impossibility. Second, it was argued that (some of) the Kuhnian values could be used to determine the prior probability and the likelihood, which are, in turn, criteria of the Bayesian theory choice functional. Particularly, it was seen that accuracy is reflected by the likelihood, while consistency (in its two dimensions), simplicity, and scope have a bearing on the prior probability. Since these latter criteria can pull in different directions, we can conclude that there are many acceptable algorithms to determine the prior probabilities of the theories on the basis of Kuhn's criteria.

An objection to the conclusion just reached may arise from an observation made by Stegenga (2015, p. 273), who wonders what happened to Kuhnian criteria Okasha started with, since they disappear in his discussion about Bayesianism:

> "perhaps the Kuhnian theoretical virtues help *determine* the values of $P(T_i)$ and $P(E|T_i)$. But such a determination would require some sort of algorithm to translate the application of the virtues into probabilities. Based on Okasha's own analogy, such an algorithm would face an Arrow's analogue, since Arrow's axioms – unanimity[14] , non-dictatorship, etc. – would obviously apply to such an algorithm" (emphasis in the original).

If this is true, then the second strategy to save from impossibility theory choices among key theories relying on Kuhn's criteria no longer works. For, Arrovian impossibility obtains anyway when we use these criteria to evaluate the terms in the Bayesian algorithm for theories' comparison. However, this is too quick. Firstly, for reasons provided in section 1.3, only one of the Kuhnian criteria can help to determine the likelihood, i.e. accuracy. Thus, no Arrow's analogue arises as Arrow's theorem applies to a set N of two or more individuals. Conversely, Arrovian impossibility might arise for the determination of the priors on the base of simplicity, internal and external consistency, and scope. The latter, in fact, constitute a four-person society in Arrow's framework. Hence, Stegenga's objection should be aimed at an algorithm that assigns prior probabilities to the theories

---

[14] 'Unanimity' is another way to call the Pareto condition.

on the base of the four Kuhnian criteria just mentioned. From now on, I will call this algorithm 'Stegenga's algorithm'. That said, the real problem with Stegenga's claim is that for an algorithm of the kind he has in mind, we cannot ask whether it satisfies or not Arrow's conditions. For, these conditions are defined for a function that takes as input a profile of social-alternative-rankings and outputs an overall ranking of the social alternatives. Now, while the input of Stegenga's algorithm is analogous, i.e. a profile of theory-rankings, its output is different, namely it is a list of real numbers representing the prior probabilities of the theories. In other words, Arrow's conditions are defined for a function that has a different mathematical form from the one Stegenga has in mind. Thus, it does not make sense to ask if conditions defined for the former apply to the latter. Therefore, contrary to what Stegenga says, Arrow's theorem tells us nothing about the impossibility of algorithms that assume as input theory-rankings, one for each of Kuhn's criteria and outputs a list of prior probabilities. It may be that some kind of impossibility arises for such algorithms, but it is not an Arrovian impossibility, so our second strategy still stands[15] .

Nonetheless, Stegenga's observation raises an interesting question, namely if the Arrow-style rule, associated with Stegenga's algorithm, suffers from Arrovian impossibility. I will tackle this issue in what follows. Let us rank order the theories by their decreasing prior probability assigned to them by Stegenga's algorithm. Then, consider an algorithm that outputs such an order, and that assumes as information the theory-rankings of each of the four Kuhnian criteria that can bear on the priors. I will call this algorithm 'prior probability rule'

---

[15] In what follows, I will consider whether the Arrow-style rule derived from Stegenga's algorithm is subject to Arrovian impossibility, and I will conclude that it is still an open question. Now, one might be tempted to argue that, since, such a rule is derived from Stegenga's algorithm, the latter has the same fate. That is, it is yet to be seen if Stegenga's algorithm suffers from a sort of Arrovian impossibility. However, this is not obvious: it has to be proved that what holds for the Arrow-style rule derived from Stegenga's algorithm is somehow inherited by the latter. For the moment, the only sure thing is that we cannot ask whether Stegenga's algorithm satisfies or not Arrow's conditions, and so if it is subject or not to Arrovian impossibility. That said, if such a proof can be made, then also the judgement about our second solution should be suspended. But, as I will highlight at the end of this section, if it turns out that the meanings of Kuhn's criteria we are considering are absolutely measurable, as I suggest in section 1.3, then the Arrow-style rule derived from Stegenga's algorithm is not subject to Arrovian impossibility, even if we have yet to understand why.

(PPR). PPR has the same mathematical form of Arrow's social choice rule, so, now, it is sensible to ask whether it satisfies analogous Arrovian conditions. Thus, does it? Let us see if this is the case by taking them into consideration one by one.

P: such a condition translated for our rule states that if theory T_1 does better than theory T_2 by each of the four Kuhnian criteria, then it should be preferred overall by PPR. This condition seems satisfied. As seen in section 1.3, the Kuhnian criteria at hand are considered 'epistemic', i.e. as conducing to the truth of the hypothesis. Thus, if T_1 does better than T_2 by each of them, then it should surely have a higher prior probability, i.e. a higher probability to be true before the evidence, according to Stegenga's algorithm. But, PPR yields an order of the theories by their decreasing prior probability assigned to them by Stegenga's algorithm. So, T_1 should be preferred overall to T_2 by PPR.

N: according to this condition, there should not be any criterion such that if T_1 is ranked above T_2 by that criterion, then T_1 is placed before T_2 by PPR in the overall ranking. This condition is met as long as we concede that all the four Kuhnian criteria are relevant for the determination of the prior probability in Stegenga's algorithm. In fact, again, PPR outputs a ranking the theories induced by the prior probabilities assigned to them by Stegenga's algorithm. Thus, if we agree that Stegenga's algorithm should not have a dictatorial criterion, then neither PPR.

I: this condition translated in terms of our algorithm would say that how PPR orders T_1 and T_2 in the overall ranking should depend only on how criteria rank T_1 and T_2, and not on how they rank other theories. Again, this condition seems reasonable because it can be argued that a rational evaluation of the prior probabilities, according to Stegenga's algorithm, should not depend on irrelevant factors. But, since PPR yields an order of the theories by the prior probabilities assigned to them by the latter, then the same thing can be said for PPR.

U: according to this condition, there should be no *a priori* restriction on the possible profiles that PPR can assume as input. This con-

> dition seems to hold for the same considerations Okasha provided for the theory choice rule (see section 1.2). That is, there is no reason to think that trade-offs (or correlations) between two of the Kuhnian criteria must always obtain. The only difference is that now the targets of such a consideration are the four Kuhnian criteria at hand, and not all of them.

Thus, it seems that there is no acceptable prior probability rule, i.e. a probability rule that jointly satisfies **P**, **I**, **N**, and **U**. However, this is too hasty. For, an objection to Okasha's argument made by Morreau (2015) can be relevant also for the conclusion just reached. Let us see then what this objection consists of, and how it relates to our discussion.

Morreau argues that Arrow's theorem does not apply to the theory choice rule because the latter does not meet condition **U**. More precisely, this condition translated for such a rule would say that the domain of the theory choice rule should contain all possible profiles, i.e. all the possible combinations of any ordering at all of the alternatives. It makes sense to impose this condition on the rule only if: (i) each criterion is capable of preferences corresponding to any logically possible ordering of the theories; (ii) each criterion is capable of these preferences independently of all the others. Otherwise, **U** would ask the rule to handle profiles that will never arise. Okasha states the applicability of **U** just focusing on point (ii) (see section 1.2). But he overlooks that some of Kuhn's criteria do not meet point (i). That is, they are 'rigid' (Morreau 2015, p. 248) in that they can order the alternatives in just one way, and the presence of only one rigid criterion is enough for **U** not to apply. As Morreau (*ibid.*, pp. 247-250) rightly observes, the notion of rigidity can be fully understood by comparing a non-rigid criterion with rigid ones. Of the first kind is accuracy, supposing that it can be intended as the extent to which a theory agrees with available empirical data. Now, different orderings of the alternatives on the base of accuracy are possible depending on which data are available. One example of the second kind is one meaning of simplicity that Kuhn (1977, p. 324) mentions discussing the choice between Copernican and Ptolemaic systems, i.e. simplicity as the computational labour required to predict the positions of the planets. This criterion is rigid because there is nothing on which the order it assigns to the theories depends in the same way in which the order accuracy assigns to theories depends on available data. Indeed, it is about features of the theories it is comparing which cannot change without the theories losing their identity. Whether it is easier to predict the positions of planets using Copernican or Ptolemaic system depends on the geometry of these models which cannot change. In fact, these theories could not have had a different geometry, and still remain the theories that they are. Thus, this criterion can order the theories in just one way, and in this sense, it is rigid. So, what do we have to conclude from this? First, notice that condition **U** is known to be unnecessarily strong, hence we do not need a full variety for Arrow's theorem to apply. Still, we need to have a lot of variety. But, the domain restriction considered before is severe, which makes Morreau concludes that Arrow's theorem tells us nothing about the possibilities of deriving an overall ranking of the theories.

Now, Morreau's observation may apply also to condition **U** of our algorithm, PPR. For example, consider again simplicity as the number of principles postulated by the theory to explain a given phenomenon, and scope as the range of application of these principles. Intuitively, it seems that such features cannot be different without the theory being a different theory: Newton's mechanics would have been a different theory, had it postulated a different number of laws, and had not such laws applied to the lunar and sublunar motion. As mentioned before, even the presence of only one rigid criterion makes the domain restriction severe. So, we should conclude, along with Morreau, that Arrow's theorem does not have any consequence at all on the possibilities of prior probability rules.

Unfortunately, Morreau's optimism is misplaced, as Okasha (2015) points out. The main reason for this is that it is still an open question whether Arrovian impossibility can be obtained with a weaker domain assumption than **U**, consequence of the rigid criteria. In fact, while it was proved that impossibility can be derived with weaker domain assumptions (see Gaertner 2001), until now, there does not seem to be any result in the literature that settles the issue for domains restricted by rigid criteria. Actually, Morreau (2015, p. 250, n. 12) is aware of this, so his claims about Arrow's theorem not affecting theory choice are misleading. Rather, in light of this unresolved mathematical problem, the right reaction is to suspend the judgement, and the same goes for PPR[16].

Finally, consider that if it turns out that the meanings of Kuhn's criteria we are considering are absolutely measurable, then not only does PPR not satisfy **U**, but it also does not satisfy **I**. To see this, remember what was said about condition **I** in section 1.2: it is captured by saying that if each criterion's preference for T_1 over T_2 is the same in two profiles, then the theory choice rule, when applied to both profiles, must yield the same overall ranking for T_1 and T_2. Thus, in the generalized Sen's framework, **I** is a combination of **I′** and **ONC**. Now, if the aforementioned meanings of Khun's criteria are actually absolutely measurable, then PPR satisfies **AFC**, and not **ONC**. Thus, we are left with an open question related to the one above, which relies on a reasoning similar to the one made in section 1.3 to understand why BCF jointly satisfies conditions **P′**, **N′**, **I′**. That is, for similar considerations of before PPR jointly satisfies **P′**, **N′**, **I′**. But, why is that? Two answers are possible. The first one is that Arrow's theorem, which in Sen's framework requires condition **ONC**, does not apply because PPR does not satisfy condition **U′**. The second one is that, even if PPR does not meet **U′**, Arrow's theorem applies nonetheless, and PPR escapes it thanks to the fact that it satisfies condition **AFC**. In other words, the Arrow-style rule derived from Stegenga's algorithm is not subject to Arrovian impossibility, but we do not know why yet. And, to solve this dilemma, we first have to answer our first open question.

---

[16] Notice that a typical proof of an Arrow-style impossibility theorem requires the domain to be unrestricted with respect to at least a triple. That is, there has to be at least an unrestricted triple, i.e. some three alternatives that can be ordered in all possible ways (Morreau 2019, section 4.1). But, if (some of) the meanings of Kuhn's criteria considered are rigid, we do not have the unrestricted triple we need for an Arrow's theorem to apply. So, it could be thought that the right reaction is not to suspend the judgement, but to state that Arrow's theorem is not worrisome for PPR. However, the requirement of the unrestricted triple seems to be a sufficient, but not a necessary condition for impossibility. For example, Morreau (2015, footnote 12, p. 250 and appendix, pp. 259-262) and Okasha (2011, pp. 289-290 and footnote 6, p. 290) present Arrow-style impossibility theorems that do not require the condition of the unrestricted triple.

1.5 Conclusion

I have proposed two ways to save from impossibility large scale theory choices relying on Kuhn's criteria. The first one showed that a particular meaning of fruitfulness can be measured on an absolute scale. This led us to conjecture that such a scale might be appropriate also for other meanings of the Kuhnian criteria which can be employed, along with the understanding of fruitfulness at hand, to choose among theories. If this will turn out to be the case, then a technically meaningful form of inter-criterion comparability becomes possible and impossibility can be escaped.

The second proposal suggested to use Kuhn's values to determine the prior probability and the likelihood of theories. These are, in fact, criteria of the Bayesian theory choice functional, which escapes from Arrovian impossibility as it satisfies condition **AFC**. To show how this can be done, I considered some meanings of Kuhn's criteria and highlighted how they relate to the terms in the Bayesian algorithm for theories' comparison. It turned out that accuracy is reflected by the likelihood, while the prior is evaluated on the base of four criteria (scope, simplicity, internal and external consistency) that can pull in different directions. Because of this, Okasha's interpretation of Kuhn's argument suggests that there are many acceptable algorithms for the assessment of prior probabilities. At this point, we have considered an objection. Stegenga (2015) claims that an algorithm that translates the information provided by Kuhn's criteria in prior probabilities satisfies Arrow's conditions so that it is destined to the same impossibility. We noticed that, if true, this consideration could jeopardize the second solution proposed because Arrovian impossibility obtains also if we use Kuhn's criteria to determine the terms in Bayes' theorem. However, I pointed out that Stegenga's claim is not worrisome. In fact, the algorithm he has in mind has a different mathematical form from Arrow's social choice rule in that it outputs a list of real number instead of a ranking of theories. So, we cannot ask whether conditions defined for the Arrovian social choice rule apply to Stegenga's algorithm.

In any case, Stegenga's observation raised an interesting question, namely if the prior probability rule (PPR) associated with Stegenga's algorithm, suffer from Arrovian impossibility. We have seen that an observation Morreau (2015) made against the applicability of condition **U** to Okasha's theory choice rule applied also to condition **U** of our algorithm. In fact, we have seen that some of the Kuhn's criteria that can bear on the prior are rigid in Morreau's sense. And, since it is still an open question whether Arrovian impossibility obtains with a domain restriction caused by rigid criteria, the judgement about the impossibility of our algorithm was suspended.

Chapter 2

The Old Evidence Problem and the Inference to the Best Explanation

2.1 Introduction

In the previous chapter, we have seen that the second solution consisted of using Kuhn's epistemic values to evaluate the prior probability and the likelihood, which are criteria of the Bayesian theory choice rule. Continuing on this path, this chapter will be dedicated to explore the relationship between Bayesianism and another epistemic value that goes by name of 'explanatory power', which refers to the goodness of an explanation. Specifically, I will take into consideration the 'Inference to the Best Explanation' (IBE), in which the explanatory power of a hypothesis for a body of evidence deserves a special status. And I will explore how IBE relates to a problem which has been bedeviling the Bayesian theory of confirmation since Glymour (1980) first described it, that is, the Problem of Old Evidence (POE). So, let us see of what this relationship consists.

POE has been seen as a major descriptive flaw of the Bayesian confirmation theory, which struggles to account for a shared intuition of scientific reasoning according to which a theory H can be confirmed by a piece of evidence already known, i.e. by an old piece of evidence.

Different dimensions of POE have been highlighted (Eells 1985, 1990), and, accordingly, different solutions have been proposed to them. Here, I will consider only two dimensions, i.e. the dynamic and static one. In the former, we want to explain how the discovery that H accounts for the old evidence confirms H in the sense that it raises the subject's degree of belief in it. In the latter, instead, we want to understand why the old evidence is and will be a reason to prefer H over its competitors.

The chapter has three different aims. First, I will point out a technical shortcoming of the second of two recent solutions to the dynamic dimension of POE, proposed by Eva and Hartmann (2020). Second, I will highlight that these solutions can be read in terms of Inference to the Best Explanation (IBE). In fact, by using a different formal apparatus, both models show that learning that H is the *only* available hypothesis that adequately explains the old evidence confirms H.

By making such a reading explicit, I will further gauge the weaknesses and strengths of the two models. More precisely, I will show that Eva and Hartmann have in mind a specific understanding of IBE's core idea that explanatory considerations have a confirmational import. On this base, pending the question if this is indeed the IBE's formulation descriptively used, I will point out that, while one condition of their first model is not expression of such a formulation, the only condition employed in their second model is.

As for the third aim, I will highlight that the explicit realization that real cases of confirmation by old evidence are instances of IBE sheds some light also on the static dimension of POE which now has to be expressed in IBE terms. To solve the static dimension of POE so expressed, I will rely on the counterfactual approach (Howson 1984, 1985, 1991), and on the Bayesian IBE (Lipton 2001; Okasha 2000). The latter is a probabilistic version of IBE, according to which explanatory considerations help to evaluate the terms in Bayes' theorem. However, we will see that the problems that haunt the counterfactual approach recur even when it is used to solve the static POE in IBE terms.

To show my claims, I will follow this structure. In section 2.2, I will introduce the old evidence problem for Bayesian confirmation theory, its dynamic and static dimension, and the solutions proposed to the them. In section 2.3, I will present Eva and Hartmann's two novel solutions to the dynamic POE, and the shortcoming of their second model. Moreover, after having made explicit the connection of the two solutions with IBE, I will further assess them in light of such a connection. Finally, section 2.4 will focus on the static POE in IBE terms. After having introduced the Bayesian IBE, I will point out how the latter can be employed, along with the counterfactual approach, to solve the IBE's reading of the static POE. Finally, I will explain how the problems of the counterfactual approach can undermine also the solution at hand. In section 2.5, the sums of the work will be drawn.

2.2 The old evidence problem, and its dimensions

As said in the introduction, the old evidence problem states that the Bayesian confirmation theory cannot account for the intuition according to which a theory can be confirmed by an old piece of evidence.

The most famous instance of confirmation by old evidence is the one of the Mercury perihelion shift (Glymour 1980). The advance of Mercury's perihelion was an anomalous piece of evidence, as it was not explained by the available scientific theories, like Newtonian mechanics. Then, Einstein realized that his General Theory of Relativity (GTR) accounted for this phenomenon. When such a relationship between the theory and the evidence was discovered, the evidence was already known: the nature of Mercury's perihelion had been object of intense study by astronomers for many decades. However, according to many physicists, the Mercury perihelion shift confirms GTR because the latter resolves that observational anomaly.

From this example, we can derive a general pattern (Hartmann and Sprenger 2019, pp. 131-132):

1. We start with an anomalous piece of evidence E.
2. At some point, it is discovered that theory H can account for E.
3. E is an old piece of evidence: at the time in which the relationship between H and E is developed, the scientist is already certain or close to certain that E is real.
4. E confirms H because the latter resolves the observational anomaly E.

Now, if we formalize this schema in a Bayesian fashion, we obtain:

$$P(H|E) = P(H)P(E|H)/P(E) = P(H) \tag{2.1}$$

Indeed, since E is already known, the scientist's degree of confidence in it is maximal, i.e. $P(E) = 1$. From this, it follows that also $P(E|H)$ is equal to 1, as the following theorem holds: If $P(E) = 1$, and $P(H) \in (0; 1)$, then $P(E|H) = P(E|\neg H) = 1$.

Given that the posterior probability of the theory, $P(H|E)$, is equal to its prior probability, $P(H)$, E does not confirm H, according to the notion of 'confirmation as increase in firmness' (Hartmann and Sprenger 2019, pp. 50-55). The latter, in fact, states that E confirms H if and only if $P(H|E) > P(H)$. This is the relevant sense of confirmation in this example, as confirming evidence raises the rational agent's degree of belief in the theory. Thus, as announced, Bayesian confirmation theory cannot explain the scientific intuition according to which a theory can be confirmed by a piece of evidence already known. From here, the problem of old evidence.

Different dimensions of POE have been highlighted (Eells 1985, 1990). Here, I will take into consideration only two of them, i.e. the dynamic and static dimension, and I will briefly present the solutions proposed to them.

2.2.1 The dynamic dimension of POE

In the dynamic POE, we find ourselves in a moment in time in which H and its relationship with E are discovered. What we want to know is how the discovery that H accounts for E raises the subject's degree of belief in H.

To understand this better, let us take into consideration again the Mercury perihelion shift's example. It took Einstein some time to find out that GTR (H) entailed the anomalous shift (E), i.e. $X = H \vdash E$ (Brush 1989, Earman 1992). By learning X, Einstein increased his degree of belief in H, as X was a surprising fact, in line with the scientific intuition that surprising evidence has more confirmational value. So, we can take the inequality $P(H|X, E) > P(H|E)$ to represent Einstein's actual degree of belief respectively after and before learning X. How this actual degree of belief is reached is precisely what we want to understand in the dynamic POE.

Notice that the confirming evidence is not E itself but learning the *logical* fact X. Modeling this situation in a Bayesian fashion implies an abandonment of the logical omniscient of the Bayesian agent who otherwise cannot learn the logical fact X, always part of her background knowledge. Such an abandonment is the common core of the classic approaches to dynamic POE which, however, differ in relevant respects. Let us see briefly what they consist of.

The first one I will consider is due to Garber (1983) who, as mentioned, abandons the logical omniscience of the Bayesian agent and adds a new class of *atomic* sentences of the form X to the language L of propositions about which the agent can have degrees of belief. A new language L' is therefore formed. This addition prompts the question: which is the relationship between the new logical sentences X and the preexisting sentences of the language, particularly H and E? Garber answers this question by coming up with the following basic constrains, which are nothing more than an instantiation of a basic form of modus ponens, i.e. 'if I am certain of $\mathrm{H} \vdash \mathrm{E}$ (X) and of H, then I should also be certain of E':

G1 $P(E|H, X) = 1$.

G2 $P(H|E, X) = P(H, X)$.

By **G1** and **G2**, Garber proves the following results:

GT1 There exists at least one probability function on L' such that every non-trivial atomic sentence of the form X has a non-extreme probability value $P(X) \in (0;1)$.

That is, it is possible to be genuinely uncertain about X in a coherent way.

GT2 For any atomic sentence of the form X there are infinitely many probability functions satisfying (i) $P(E) = 1$, (ii) $P(H|X, E) > P(H|E)$.

GT2 is the crucial result: it is always possible to find a probability function that solves POE in that the logical sentence X confirms H (i), even if E is old evidence (ii).

A similar point is reached by Niiniluoto (1983). These results, however, just show the existence of probability functions that solve POE. What we need to know is the conditions that capture real cases of confirmation by old evidence and that lead the scientist to conclude that $P(H|X, E) > P(H|E)$.

This gap is filled by Richard Jeffrey (1983). Again, Jeffrey abandons the Bayesian agent's logical omniscient and adds two atomic sentences to the language L. The first one is Garber's X, while the second one is $Y = \mathrm{H} \vdash \neg \mathrm{E}$. Then, he introduces a bunch of conditions from which he derives what we want to show, namely that $\mathrm{P}(\mathrm{H}|X, \mathrm{E}) > \mathrm{P}(\mathrm{H}|\mathrm{E})$. That is:

J1 $\mathrm{P}(\mathrm{E}) = 1$.

J2 $\mathrm{P}(\mathrm{H}), \mathrm{P}(X), \mathrm{P}(Y) \in (0, 1)$.

J3 $\mathrm{P}(X, Y) = 0$.

J4 $\mathrm{P}(\mathrm{H}|X \vee Y) \geq \mathrm{P}(\mathrm{H})$.

J5 $\mathrm{P}(\mathrm{H}, \neg \mathrm{E}, Y) = \mathrm{P}(\mathrm{H}, Y)$.

The adequacy of the proof relies on the plausibility of its conditions. **J1**, **J2**, **J3**, and **J5** seems to be plausible. **J1** captures the fact that we are dealing with the old evidence problem. **J2** encodes the standard assumption that to begin with we are not certain about the truth of H and $\mathrm{H} \vdash \pm \mathrm{E}$. **J3** is a consistency requirement: we believe that H is consistent. Finally, **J5** is an instantiation of modus ponens as **G2** is. Thus, the burden of the proof is carried by **J4**.

Earman (Earman 1992, p. 127) highlights that this condition has some strange technical consequences. However, Sprenger (2015, p. 390) points out that its real problem is philosophical in that "it conflates an evidential virtue with a methodological one". Let us see why. **J4** says that learning that a theory makes precise predictions about some phenomena, even if the content of the predictions is not known yet, increases the probability of the theory. For example, Newton was convinced that his theory of gravitation (G) would have borned on the phenomenon of the tides, even though it was not clear at the time if G would have entailed it or not. Nonetheless, this led Newton to accept G as a working hypothesis (Jeffrey 1983, pp. 148-149). It may be that Newton followed a Popperian methodological rule preferring theories with high empirical content, i.e. theories that make precise predictions, and to develop them further, as they will help us to solve scientific problems. However, this does not mean that the plausibility of a theory increases with its empirical content. As a matter of fact, Popper (2002, pp. 268-269) though the other way around: theories with high empirical content will have a low probability because, since they make many predictions, they will have a higher risk of being falsified. This is why Jeffrey's proof does not seem to be very compelling.

Earman (1992, pp. 128-129) proposes alternative derivations of $P(H|X, E) > P(H|E)$, that use conditions **J1**, **J2**, **J3** and **J5**, but replace **J4** with, respectively, two new conditions.

The first one is the following:

E1 $P(H|X) > P(H|\neg X, \neg Y)$.

That is, learning that H implies E is more favorable to H than learning that H gives no definite predictions about E. But Sprenger (2015, p. 390) highlights: "this condition just seems to beg the question". In fact, as said before, what we want to show in the dynamic problem of old evidence is that the probability of the theory after learning X is higher than the probability of the theory before that learning, so when H gives no definite predictions about E. Thus, this conditions simply imposes something very similar to what we want to show without offering any independent motivation.

The second alternative is the following:

E2 $P(X \vee Y) = 1$.

But, as Earman (1992, p. 129) himself acknowledges, this condition is too strong: it requires that the scientist, upon formulating H, is certain that it either implies E or ¬E, which is an unrealistic assumption.

Thus, all in all, the approaches considered so far struggle to work or because they are incomplete (Garber and Niiniluoto's ones) or because they rely on very problematic assumptions (Jeffrey and Earman's ones).

Hartmann and Fitelson (2015, p. 714) highlight another reason why these solutions are not adequate. That is, they do not allow for the natural possibility that explanatory facts, that can also be non-deductive, can provide the basis for confirmation by old evidence. Thus, they propose a new interpretation of X and Y. Namely:

- X: H *adequately explains* E.

- Y: H's best competitor (H') *adequately explains* E.

Then, they derive P(H$|X$) > P(H) from the following qualitative constrains (assuming that P(E) = 1):

HF1 P(H$|X, \neg Y$) >P(H$|\neg X, \neg Y$).

HF2 P(H$|X, \neg Y$) >P(H$|\neg X$, Y).

HF3 P(H$|X$, Y) >P(H$|\neg X$, Y).

HF4 P(H$|X$, Y) $\geq$ P(H$|\neg X$, $\neg Y$).

Prima facie these conditions seem plausible. According to **HF1** and **HF2**, H's probability is higher supposing that it adequately explains E and H′ does not, than supposing that H does not adequately explain E, along with H′ (**HF1**), or while H′ adequately explains E (**HF2**). Finally, **HF4** makes two exclusive claims: H's probability is strictly higher, given that both H and H′ adequately explain E, than whether neither H nor H′ adequately explain E; H's probability, given the supposition that both H and H′ adequately explain E, is equal to the one H would have supposing that neither H nor H′ adequately explains E. Both of these claims seem to be compelling: one may be willing to rank P(H$|X$, Y) strictly higher than P(H$|\neg X$, $\neg Y$), as $X \wedge Y$ implies that H adequately explains the old evidence E, whereas $\neg X \wedge \neg Y$ implies that H does not adequately explain E; on the other hand, it could be argued that in both suppositions there is no difference between H and H′ with respect to explaining E, and so there should not be a difference between the two in terms of probability.

However, Eva and Hartmann (2020) point out that Hartmann and Fitelson's proposal is incomplete, as its conditions (**HF1-HF4**) are jointly sufficient to guarantee P(H$|X$) > P(H), but they are not sufficient to guarantee P(H$|X \wedge \neg Y$) > P(H). That is, they allow for the possibility that $X \wedge \neg Y$ can disconfirm H. And this is implausible: learning that H adequately explains E and H's best competitor (H') does not should always make us more confident in the truth of H. In fact, in this situation, I become more confident that H is the *only* way I can possibly account for the evidence and thereby become more confident that H *has to be true*.

2.2.2 The static dimension of POE

In the static dimension of POE, we find ourselves in a moment in time in which belief changes caused by the discovery of H and its relationship to E already happened. However, we want to say why E is and will be a reason to prefer H over its competitors.

The standard approach to the static POE states that the confirmational relation between H and E has to be evaluated relying on a counterfactual degree of belief function where E is not taken for granted[1]. So, we are giving up the actual degrees of belief in E.

This means that the conditional probabilities, $P(E|H)$ and $P(E|\neg H)$, are not equal to 1, as it would turn out if E is taken for granted, as seen before. Thus, $P(E|\pm H)$ describes the degrees of belief we would have in E if we did not know that E and H were the case. Namely, they describe our degrees of belief in E supposing H and H'.

This allows for a meaningful comparison between the likelihoods to establish confirmation judgements. Let us see why. First of all, notice that the denominator of Bayes' theorem can be neglected as it is the same for all the theories we are considering – remember, we want to say why evidence E confirms H more than other theories. Thus, we have that $P(H|E) > P(\neg H|E)$ if and only if $P(H)P(E|H) > P(\neg H)P(E|\neg H)$. Now, if P is an 'impartial' prior probability distribution, i.e. $P(H) = P(H')$, then $P(H|E) > P(\neg H|E)$ if and only if $P(E|H) > P(E|\neg H)$.

Going back to the Mercury perihelion shift example, we have the latter confirmation judgement. For, $P(E|H) = 1$ because GTR implies the Mercury perihelion shift, whereas $P(E|H') \ll 1$, as Newtonian mechanics and other theories do not make definite predictions about E.

The main flaw of the counterfactual approach is that its novel interpretation of probability raises many philosophical and technical problems (e.g. Eells 1990, p. 208). For instance, it is not always the case that we are able to say which our degree of belief in H given E would have been, if we had not known E. This can happen for differ-

---

[1] The counterfactual approach to the static POE is primarily due to Colin Howson (1984, 1985, 1991). However, subsequently, Howson (2017) abandons the counterfactual approach presented in this section, by arguing that static POE can be easily solved in an *objective* Bayesian setting. That said, my focus throughout the article is how the old evidence problem emerges in the *subjective* Bayesianism, and the solutions proposed to it

ent reasons: in some cases, H would not even have been formulated had E not been known; in some other cases, if the person's degree of belief in E had been less than one, the person would be dead now, as their knowledge of E saved their life at some point on the past. Moreover, such a modification of Bayesian confirmation theory would inherit the well-known difficulties relating the proper interpretation of counterfactual conditionals.

2.3 Eva and Hartmann's two novel solutions to the dynamic POE: the Inference to the Best Explanation's perspective

In a recent article, Eva and Hartmann (2020) propose two novel solutions to the dynamic POE. Their common denominator is the observation that, in the real cases of confirmation by old evidence, hypothesis H receives strong confirmation by old evidence E because it is the *only* available hypothesis that adequately explains E. Indeed, both models show that learning such a fact confirms hypothesis H. To put it in other words, what confirms H is learning that H is the best explanation of E. That is, real cases of confirmation by old evidence are instances of the Inference to the Best Explanation (IBE).

However, the two authors never make an explicit connection to IBE. I contend that such a connection allows to further gauge the weaknesses and strengths of the two models. Showing the latter point is the main purpose of this section (subsection 2.3.3). But, in order to do that, first, I will present the two models, and point out a shortcoming of their second model (subsection 2.3.1). Then, I will briefly explain what IBE is (subsection 2.3.2).

2.3.1 Eva and Hartmann's two models

The first model aims to overcome the incompleteness of Hartmann and Fitelson's model (see end of subsection 2.2.1), by finding plausible extra conditions, consistent with **HF1-HF4**, under which $X \wedge \neg Y$ does not disconfirm H.

There's no need to go very far, as this condition is a slight strengthening of condition **HF4**, namely **HF4***: $\mathrm{P}(\mathrm{H}|X, Y) = \mathrm{P}(\mathrm{H}|\neg X, \neg Y)$.

According to Eva and Hartmann, this assumption captures the idea that, typically, hypotheses receive significant confirmation by the old evidence only when they are the only hypotheses that adequately explain the relevant evidence. It does so - I believe - by telling us that when H's best competitor explains E as well as H does, H's probability is equal to the one it would receive if neither H nor H' adequately explain E, which does not confer significant confirmation.

Then, they prove that **HF1-HF4*** are jointly sufficient to guarantee: (1) $P(H|X) > P(H)$; (2) $P(H|\neg Y) > P(H)$; (3) $P(H|X \wedge \neg Y) > P(H)$. Thus, we have the desired result: (3).

More specifically, the aforementioned idea is justified by the observation that it is what happens in the real cases of confirmation by old evidence. Indeed, in the Mercury perihelion shift's example, GTR received such a strong confirmation because it adequately explained E, and none of its competitors did. In fact, the competing hypotheses - Le Verrier's unobserved planet 'Vulcan', and Von Seeliger's ring of a particular matter around the sun (e.g. Crelinsten 2013) - were not serious competitors to GTR when Einstein found out it implied E. If there are competing theories which are also capable of adequately explaining the old evidence, then the degree of confirmation conferred on the hypothesis by the old evidence would intuitively be far weaker, and possibly negligible. For instance, if $H \in S$, where S is a class of mutually incompatible theories, and it is showed that all the theories in S adequately explain some old evidence E, this proof will not do much to increase our confidence in H, since it does nothing to distinguish H form its competitors.

The point just exposed suggests that scientists are primarily concerned with the proposition A: 'H is the *only* available hypothesis that adequately explains E'. Accordingly, we need to show that the agent increases her confidence in H because she becomes more confident in A, *without necessarily* becoming certain of the truth of any individual proposition, i.e. 'H adequately explains E' and 'H is the only available hypothesis that adequately explains E'. Formally, by using Jeffrey's conditionalization, we want to show: $P^*(H) = P(H|A)P^*(A) + P(H|\neg A)P^*(\neg A) > P(H)$.

In order to do this, Eva and Hartmann use only one minimal constraint **A1**: $P(H|A) > P(H|\neg A)$. That is, H is more likely to be true assuming that it is the only available hypothesis that adequately explains the old evidence than assuming that it is not.

From **A1**, it straightforwardly follows that, when the scientist becomes more confident in *A*, P*(H) > P(H)[2].

When the scientist learns *A* for certain, then Jeffrey's conditionalization reduces to strict conditionalization, and we need to prove: $P^*(H) = P(H|A) > P(H)$, which, again, straightforwardly follows from **A1**, since the latter is equivalent to $P(H|A) > P(H)$.

A last interesting remark Eva and Hartmann (p. 492) makes about their second model is that it is closely related to the so called 'no alternative argument' (NAA). Here are the reasons. Roughly, NAA's pattern is the following:

Premise 1: hypothesis H has some desirable features F.

Premise 2: despite significant effort, the scientific community has been unable to find any alternatives to H that share those desirable features F.

Conclusion: hence we have at least one good reason in favour of H.

In regard of this pattern, Eva and Hartman mention that Dawid, Hartmann, and Sprenger (2015) provide a Bayesian analysis of NNA and that they identify conditions under which the premises of the argument (especially premise 2) can provide legitimate confirmation to H. Now, in the special case in which the features F of the premises denote the ability to adequately explain some existing body of evidence, then the conjunction of the premises corresponds roughly to proposition *A*. So, assumption **A1** can be seen as a particular instantiation of the conclusion of NNA.

In drawing the sums on the two models, Eva and Hartmann point out that they are continuous in that they both show that learning the proposition 'H is the only available hypothesis that adequately explains E' confirms H. This proposition, in the first model, is expressed by $X \land \neg Y$, and, in the second model, by *A*.

However, in theses final remarks, while remaining faithful to their first model, Eva and Hartmann (pp. 491-492) distance themselves from their first one, by arguing that it cannot be correct, as it assures

---

[2] for the proof see Eva and Hartmann (2020), footnote n. 4, p. 491

that (1) and (2), i.e. that both individual propositions X and $\neg Y$ always confirm H. But, as they have stressed, the realization that H accounts for E might not have confirmatory significance if we are in a situation in which E has already been explained by all of H's serious competitors. By the same token, learning that none of H's competitors can explain E might have no confirmatory significance if we already know that H also fails to explain E.

That said, one could also reasonably distance themselves from Eva and Hartmann's second model on the base that its only condition **A1** begs the question, being too close to what, ultimately, needs to be explained, i.e. $P(H|A) > P(H)$[3]. Now, Eva and Hartmann motivate the simplicity of their model by saying that it models only those aspects of the scientist's cognitive state that are necessary to explicate the origin of confirmation by old evidence. However, perhaps, a better trade-off between the descriptive accuracy of the model and its informativeness is needed.

2.3.2 What is Inference to the Best Explanation?

The core idea of Abduction or, as it is more commonly called nowadays, Inference to the Best Explanation is that explanatory considerations have a confirmational import. Such an idea has been cashed out in a variety of ways. Here, following Douven (2017), I will consider three of them, which are all inference rules, starting with the following:

ABD1 given evidence E and candidate explanations $H_1, \ldots, H_n$ of E, infer the (probable) truth of *that* H_i which best explains E.

The main problem with **ABD1** is that it does not appear to be normatively adequate as its reliability is based on conditions hard to justify. In fact, in order for **ABD1** to be reliable, we need two necessary conditions:

1. In most of the cases, the best explanation relative to the hypotheses we have considered must also be the best relative to the hypotheses we might have conceived. That is, the best *absolute* explanation of the evidence has to be among the candidate

---

[3] My supervisor, Jan Sprenger, pointed this out to me

hypotheses we have come up with. Otherwise, **ABD1** would lead us to consider probably true, and so to believe, "the best of a bad lot" (van Van Fraassen 1989, p. 143).

2. In most of the cases in which the best explanation of the evidence is also the best absolute explanation, the best explanation is probably true.

However, 1 can be fulfilled only when we assume a predisposition on the agent part to hit the best absolute explanation among the ones she has considered. But, as van Fraassen points out (*ibid.*, p. 144), it is *a priori* implausible to suppose we have such a form of privilege.

The most promising response to the 'argument of the bad lot' point outs that the rule is asymmetric (e.g. Kuipers 2000, p. 171). Namely, it has an absolute conclusion – the hypothesis is probably true – on the basis of a comparative premise – the best explanation of the data is relative to the available explanations of the data. This discrepancy can be avoided in two ways: or by making the premise absolute as well, or by making the conclusion comparative.

According to the first path, the probable truth of the best explanation is not to be inferred only when the latter is the best explanation with respect to the candidate explanations, but also when it is a satisfactory (Musgrave 1988) or good enough (Lipton 1993) explanation. Thus:

ABD2 given evidence E and candidate explanations $H_1, \ldots, H_n$ of E, infer the (probable) truth of *that* H_i which best explains E, provided H_i is satisfactory/good enough *qua* explanation.

The main problem with **ABD2** is that it relies on concepts, such as the satisfactoriness of an explanation or its being good enough, of which we lack a full understanding.

Conversely, as announced before, the second option derives a comparative conclusion from a comparative premise:

ABD3 given evidence E and candidate explanations $H_1, \ldots, H_n$ of E, if H_i explains E better than any of the other hypotheses, infer that H_i is closer to the truth than any of the other hypotheses.

ABD3 requires, instead, an account of closeness to the truth. But many accounts of this kind are available today (e.g. Niiniluoto 1998).

The bright side of the latter two definitions is that, despite the shortcomings, their reliability is not based on an implausible form of privilege as **ABD1**'s one.

2.3.3 Weaknesses and strengths of the two models form IBE's point of view

Armed with what said in the previous two sections, it is now time to make explicit the connection between IBE and Eva and Hartmann's two models, and to further assess the latter in light of such a connection.

In order to do that, the first thing to understand is which of the three formulations of IBE considered in subsection 2.3.2, if any, comes out from Eva and Hartmann's discussion.

As pointed out in subsection 2.3.1, the two authors contend that in the real cases of confirmation by old evidence, hypothesis H receives strong confirmation by old evidence E, because H is the only available hypothesis that adequately explains E. At a closer look, it can be seen that such an idea is expressed both in terms of *confirmation as firmness*, and in terms of *confirmation as increase in firmness*[4].

According to the former, learning that H is the only available hypothesis that adequately explains E (e.g. $X \wedge \neg Y$) strongly confirms H because $P(H|X \wedge \neg Y) = t$, where t is a high value. This is how the idea is captured by condition **HF4***: $P(H|X, Y) = P(H|\neg X, \neg Y)$. Here is why. Remember that such a condition captures the aforementioned idea by, presumably, telling us the following: when H's best competitor explains E as well as H does, H's probability is equal to the one H would receive if neither H nor H′ adequately explain E, which does not confer a significant confirmation. Thus, ultimately, **HF4*** captures the idea at hand by saying that $(X \wedge Y)$ does not strongly confirm H, as $P(H|X, Y) = t$, where t is not a significantly high number. Namely, $(X \wedge Y)$ does not strongly confirm H according to the sense of confirmation as firmness. So, when the idea according to which hypothesis H receives strong confirmation by old evidence E because H is the only available hypothesis that adequately explains E is captured by

---

[4] For the distinction about these two senses of confirmation, see Hartmann and Sprenger 2019, variation 1.

HF4*, the sense of confirmation in it is expressed in terms of confirmation as firmness. As for the idea of confirmation as increase in firmness, instead, as seen in section 2.2, learning that H is the only available hypothesis that adequately explains E ($X \wedge \neg Y \equiv A$) strongly confirms H because P(H|$X \wedge \neg Y \equiv A$) > P(H). And, as we have seen in subsection 2.3.1, this is just what the two models show.

Now, IBE's three formulations assign to explanatory considerations firm confirmation judgements (see subsection 2.3.2): the best explanation (and good enough explanation) is probably true (**ABD1**, **ABD2**); the best explanation is closer to the truth than any of the other hypotheses (**ABD3**). Thus, the best way to see if one of these formulations is implicitly endorsed by Eva and Hartmann is to stick to their idea in terms of confirmation as firmness. Namely, learning that H is the only available hypothesis that adequately explains E strongly confirms H because the probability of H in light of such a learning is a high value. This can be read as saying that if H is the best explanation in that it is the only good enough explanation, than its probability is equal to a high value. This inference rule is really similar, although not equal, to **ABD2**.

Thus, summing up, if we want to integrate Eva and Hartmann's novel contributions with IBE, the following should be endorsed: learning that H is the best explanation of E, in the aforementioned sense, implies that E strongly raises H's probability so that the latter assumes a high value.

With that in mind, let us now see the assessment of the two models in light of such an explicit connection. A remark that can be made to both models is that, as highlighted just now, they seem to rely on an inference rule which is very similar to **ABD2**. But, so far, it is still an open question which of the three IBE's formulations is descriptively used, or if some further rule is used or whether some version is used in some context and another version in others (Douven 2017, section 2). That is, there is an empirical descriptive question that needs to be solved: do scientists rely on a version of IBE very similar to **ABD2**, as Eva and Hartmann implicitly think?

As for the first model alone, it seems to me that condition **HF4*** is not expression of IBE's core idea that explanatory considerations have a confirmational import. Let us see why.

If H is the best explanation of E in the sense that it is the only hypothesis that adequately explains E (or the only good enough explanation of E), than H receives a high probability. So, such a high probability is given to H because: (i) it is an adequate explanation itself; (ii) it is the only adequate explanation of E.

Now, the first supposition of condition **HF4***, i.e. $X \wedge Y$, tells us that H is an adequate explanation of E itself but that it is not the only adequate explanation of E, as H′ is an explanation as adequate as H. The second supposition of **HF4***, i.e. $X \wedge \neg Y$, on the other hand, tells us that H is not an adequate explanation at all, and so it cannot be the only adequate explanation of E. Thus, in the first supposition, we have one of the two reasons why H receives a high probability. In the second one, none of them.

Consequently, the probability of H given the first supposition should be higher than its probability given the second one. Namely: $P(H|X, Y) > P(H|\neg X, \neg Y)$.

This is not to say that when H's best competitor explains E as well as H does, H does not receive weak or negligible confirmation, but that, no matter how insignificant, H's confirmation in light of X and Y is higher that its confirmation in light of $\neg X$ and $\neg Y$.

Regarding the second model, the situation is different. In fact, condition **A1**: $P(H|A) > P(H|\neg A)$ is expression of IBE's core idea. Indeed, Eva and Hartmann's IBE implies that when H is the best explanation of E, then H receives a high probability. Conversely, when H is not, it does not receive such a high probability.

Therefore, the explicit connection between IBE and Eva and Hartmann's two solutions allows us to make a more complete assessment of the two models. That is, in addition to the problems the two models already have, one condition of their first model is not appropriate to model cases of confirmation of old evidence intended as instances of IBE, whereas the only condition of their second model is.

2.4 Bayesian IBE and the Inference to the Best Explanation's perspective on the static dimension of POE

By focusing on the origin of confirmation by old evidence, i.e. on how H is confirmed at the moment in which H and its relationship to E are discovered, Eva and Hartmann tackle only the dynamic dimension of POE. However, the explicit realization that real cases of confirmation by old evidence are instances of IBE sheds some light also on the static dimension of POE which now has to be expressed in IBE terms.

Before expounding on this point, let us consider a particular version of IBE, i.e. the Bayesian IBE, which will come in handy to solve the static POE from the IBE's perspective. More specifically, the latter is the product of a response to an incompatibility between IBE and Bayesianism claimed by some philosophers.

2.4.1 Incompatibility between Bayesianism and IBE, and Bayesian IBE

The confirmational role that IBE assigns to explanatory considerations (see subsection 2.3.2) directly suggests a comparison with Bayesian confirmation theory, the dominant view on confirmation.

In this regard, some philosophers have stated an incompatibility between IBE and Bayesianism. For example, C. Salmon W. (2001) stresses that the Bayesian confirmation theory is not guided by the concept of explanation. In fact, he concedes that the prior probability in Bayes' theorem can be identified with the goodness, or "loveliness" (Lipton 2001, p. 105), of an explanation, that is with the degree of understanding the explanation at hand provides. But, such a loveliness is a consequence of the prior probability of the hypothesis and not the other way around.

The prior probability of the hypothesis is determined by considering how the latter fits our background knowledge. This can contain all sort of information: theories known at the time, frequencies of the data, the evidential record available. On this base, one can determine the prior probability of the hypothesis considering, for example, its external consistency, or its ad hoc or non-ad hoc character.

It may well be that such epistemic virtues coincide with explanatory ones, but the prior probability of the hypothesis is evaluated relying on their inferential character, and not on their explanatory character. According to Salmon, we do not say that a given hypothesis deserves a low or high prior probability because it is a bad or good explanation in that it scores badly or well on the aforementioned virtues. Rather, we say that a given hypotheses deserves a low or high prior probability because it scores badly or well on the aforementioned virtues. Period.

Moreover, Salmon continues, the prior probability is not enough to make a choice among hypotheses. We need the posterior probability of the hypothesis of which prior probability is just a component, as Bayes' theorem tells us. At least, he says, this is what scientists implicitly do when they choose their hypotheses.

On the normative side, there is van Fraassen's criticism (1989), according to which a Bayesian agent that uses IBE as a rule, is liable to diachronic Dutch Book (Teller 1973). In fact, such an agent adopts an explicit strategy that consists in adding bonus points to the hypotheses that explain the evidence particularly well *after* conditionalization. On this base then, the bookie can construct a series of bets that leave the Bayesian agent with a certain loss.

Reactions to these criticisms have come respectively from Lipton (2001, 2004) and Okasha (2000), who endorse a 'Bayesian IBE'. Namely, explanatory considerations may act as a heuristic by helping to determine, even if roughly, the probabilities in Bayes' theorem needed for the transition from prior to posterior probability.

More precisely, *contra* Salmon, Lipton (2001) makes different claims, all of which are directed to support his heuristic endeavour which he expresses by saying that "the Bayesian and the Explanationist should be friends" (p. 94).

One step in this direction is to show that the inferential character of the virtues, used to estimate the prior probability of the theories, is a symptom of their explanatory character. This is what Lipton calls "the guiding challenge" (pp. 107-109), which he resolves by saying that the best explanation of the match between inferential and explanatory virtues is that scientists select hypotheses on the base of their explanatory virtues.

The second step consists of arguing that the loveliness is not related solely to the prior probability, but to all the components of Bayes' theorem. For example, explanatory considerations might help to evaluate the likelihoods because lovelier explanations tend to make what they explain likelier. Moreover, they could help to determine the priors in different ways. Firstly, priors are generally determined by earlier conditionalization, where the assessment of the likelihood is essential. But, as pointed out just now, such an assessment might be helped by explanatory considerations. Secondly, explanatory virtues such as scope, mechanism, precision, unification, simplicity, could be used by the Bayesian to estimate the prior probability, if we take for granted the success of the guiding challenge. Finally, explanatory considerations might help to determine why certain bits of evidence enter the Bayesian process of conditionalization. In fact, we can come to see that a datum is relevant for the hypothesis precisely because the hypothesis would explain that datum.

Contra van Fraassen, Okasha (2000), instead, highlights that van Fraassen's way to represent IBE in probabilistic terms is idiosyncratic. For, it does not capture the phenomenology of IBE where there is no hint of a two-stage process. We do not first respond to the evidence, and then take explanatory considerations into account. Rather, we use explanatory considerations to decide how to respond to the evidence – just one process.

This suggests that the best way to represent IBE in probabilistic terms is to use explanatory considerations in the process that realizes conditionalization. That is, the better the explanation, the higher its prior and/or likelihood are, and, in any case, given the same body of evidence E, the best explanation of E will end up having the higher product of these two probabilities.

Indeed, as Lipton, Okasha underlines that explanatory considerations help to determine the likelihoods because better explanatory hypotheses tend to give a higher likelihood to the evidence. Differently from Lipton, and agreeing with Salmon, Okasha argues that explanatory considerations help to determine the prior probability in the sense that the goodness of an explanation is a consequence of its plausibility, i.e. of its prior probability.

2.4.2 The IBE's perspective on the static dimension of POE

As seen in subsection 2.2.2, in the static dimension of POE, we find ourselves in a moment in time in which belief changes caused by the discovery of H and its relationship to E already happened. Still, we want to say why E is and will be a reason to prefer H over its competitors.

In IBE terms, we want to explain why E confirms the best explanation H more than the other theories that are not best explanations of E. Indeed, this is what happens: (i) if Eva and Hartmann's IBE is descriptively accurate; (ii) given IBE's "self-evidencing character" (Lipton 2001, p. 96).

In fact, according to (i) when H is the best explanation of E in the sense that it is the only hypothesis that adequately explains E, than it receives a high probability. Conversely, the other explanations do not receive such a high probability. And, according to (ii), the datum explained by the hypothesis, in turns, confirms the hypothesis precisely because it is explained by the hypothesis.

In the same subsection 2.2.2, we saw that the standard approach to the static POE urges us to give up the actual degrees of belief in E so to allow a meaningful comparison between the likelihoods to establish confirmation judgements. That is, $P(H|E) > P(\neg H|E)$ if and only if $P(H)P(E|H) > P(\neg H)P(E|\neg H)$.

As seen in subsection 2.4.1, the heuristic conciliatory attempts prove that better explanations are the ones with higher priors and/or likelihoods, and, anyhow, the ones with the higher product of these two quantities. Thus, we have $P(H|E) > P(\neg H|E)$, i.e. E confirms the best explanation H more than the competing theories that are not the best explanations of H.

That said, the problems connected with the counterfactual interpretation of probability are still present when the aforementioned approach is used to solve the static POE in IBE terms. For instance, in the Mercury perihelion shift's example, it is unlikely that we could say which our degree of belief in ¬H given E would have been, if we had not known E.

Indeed, Le Verrier's unobserved planet 'Vulcan', and Von Seeliger's ring of a particular matter around the sun were generated to account for the anomalous shift (see Crelinsten 2013, pp. 51-54). That is, these hypotheses would not have been formulated had E not been known. The same cannot be said for Einstein's GTR. The latter, in fact, was not constructed to explain the anomalous shift, and its derivation of the latter was surprising in that it was not expected beforehand (see subsection 2.2.1). Still, we would lack what we want to explain, i.e. $P(H|E) > P(\neg H|E)$, as we would miss the comparison term $P(\neg H|E)$.

2.5 Conclusion

In the foregoing, I have made an explicit connection between Eva and Hartmann's two novel solutions to the dynamic POE and the Inference to the Best Explanation. This has allowed me to evince that Eva and Hartmann have in mind a specific understanding of IBE's core idea, which is very similar, although not equal, to **ABD2**. On this base, I have made a more complete assessment of the two models.

Specifically, I have shown that - taking for granted the open question if such an understanding of IBE is descriptively accurate - their first model is not adequate to solve the dynamic POE in IBE's terms, while the second one is. The reason is that the crucial condition of the former, **HF4***, is not expression of the aforementioned IBE's idea. Conversely, the only condition of the latter, **A1**,is.

This results added to problems encountered in the two models even before having done the aforementioned explicit connection. Such problems are that two of the three results the first model proves - i.e., (1) $P(H|X) > P(H)$; (2) $P(H|\neg Y) > P(H)$ - are not in line with Eva and Hartmann's idea that hypotheses receive strong confirmation because they are the only hypotheses that adequately explain the old evidence. And, **A1** in the second model seems to beg the question.

Finally, I have pointed out that the explicit realization that real cases of confirmation by old evidence are instances of IBE implies that the static dimension of POE has to be, now, expressed in IBE terms. I have attempted to solve the static dimension of POE so expressed by remaining inside the frame of the standard counterfactual approach. This has been possible by using the results of the heuristic conciliatory approaches which show that better explanations have higher prior and/or likelihood, and that, anyhow, they are the ones with the higher product of these two quantities. However, I have stressed that the problems connected with the counterfactual interpretation of probability are still present when the counterfactual approach is used to solve the static POE in IBE terms. For instance, I have pointed out that in the GTR's example, if the old evidence E had not been known, we would have lost what we want to explain.

Chapter 3

Patterns of Abduction in the Covid-19 Pandemic: the Case of the Alpha variant

3.1 Introduction

Generally speaking, the previous chapter aimed at making explicit that real cases of confirmation by old evidence are instances of IBE or abduction. In other words, it focused on a sense of abduction which refers to the use of explanatory power in *justifying* hypotheses, whose proper place is the so called *context of justification*.

However, in the philosophical literature, there is another sense of the term 'abduction' , which refers to the use of explanatory power in *generating* hypotheses. As such, it belongs to the *context of discovery*. The present chapter will concentrate on this meaning of the world.

The two most famous exponents of such a version of abduction are Charles S. Peirce and Norwood R. Hanson (1958, 1960, 1965). They both believe that abduction so intended follows a particular pattern, and, in fact, they propose very similar logics of generation of new theories. However, as we will see (subsections 3.2.1, and 3.2.2), it will turn out that their logics are better understood, not as logics of generation of new theories, but as logics of adoption of new hypotheses considered to be worthy candidates for further investigations.

A view of abduction very much in this spirit was defended by Gerhard Schurz (2008), who, indeed, proposes different patterns of abduction which function as search strategies, in that their structure determines a particularly promising abductive conjecture.

Schurz detects these patterns of abduction in various contexts which range from common sense to philosophy, and sciences such as paleontology, physics, chemistry, evolutionary theory, interpretation theory, statistics. My aim, in this paper, is to do such a descriptive analysis in the context of the ongoing Covid-19 pandemic. As we all tragically know, the latter is the worldwide epidemic of coronavirus disease 2019 (Covid-19), caused by a new type of coronavirus recently identified, called 'severe acute respiratory syndrome coronavirus 2' (SARS-CoV-2). The reason why the Covid-19 pandemic is a fertile ground to carry out this task is that it has confronted us with an entirely new situation, which, by its nature, cries out for search strategies as Schurz describes them. Particularly, I will look for case studies in the circumstances surrounding the emergence of the Alpha variant, the first SARS-CoV-2's variant of concern.

To that purpose, I will follow this structure. In section 3.2, I will make a historical introduction on the concept of abduction in the context of discovery by setting out Peirce and Hanson's philosophy. Moreover, I will expound Schurz's view of abduction and his topology of patterns of abduction. Section 3.3 will be dedicated to provide a brief introduction on the pandemic, the virus behind it, and the variants. In section 3.4, I will provide two case studies which show the use of abduction in Schurz's sense in the Covid-19 pandemic. The first one concerns the higher transmissibility of the Alpha variant with respect to the other variants, while the second one deals with how it is thought the Alpha variant emerged. Interestingly, we will see that the abduction pattern of the first case corresponds to Schurz's idea of abduction, but it is not present in Schurz's 2008 paper. Differently, the pattern of the second case study belongs to Schurz's classification. Finally, section 3.5 concludes.

3.2 Abduction in the context of discovery

As mentioned in the introduction, abduction in the context of discovery denotes the use of explanatory considerations in *generating* hypotheses. The two most famous exponents of such a version of abduction are Charles S. Peirce, and Norwood R. Hanson, whose philosophy I will briefly present in what follows.

3.2.1 Charles S. Peirce

It was Peirce who coined the term 'abduction' in his work on the logic of science, to denote a type of non-deductive inference different from induction[1].

Even if no coherent picture emerges from Peirce's writings on abduction, it is clear that Peirce's understanding of the word belongs to the context of discovery (Campos 2011, McAuliffe 2015). Indeed, as Peirce says, "[a]bduction is the process of forming explanatory hypotheses. It is the only logical operation which introduces any new idea " (CP 5.172)[2] or "[Abduction encompasses] all the operations by which theories and conceptions are engendered" (CP 5.590).

After that abduction helped us to conceive explanatory hypotheses, deduction and induction come into play. Deduction derives testable consequences from the explanatory hypotheses at hand, while induction helps us to reach a verdict on such hypotheses, which depends on how many testable consequences have been verified.

Now, according to Peirce, abduction as the process of generating new hypotheses belongs to the the logic of science, as it has the following schema (CP 5.189):

The surprising fact, C, is observed.
But if A were true, C would be a matter of course.
Hence, there is reason to suspect that A is true.

However, as Frankfurt (1958, p. 594) rightly observes, this inference does not lead to any new idea. As a matter of fact, the new idea - the explanatory hypothesis A - occurs *before* one infers that there is reason to suspect that A is true, as A already figures in the second premise.

That said, Frankfurt (*ibid.*, p. 595) points out a way to conciliate abduction as belonging to the context of discovery, and its nature as

---

[1] My treatment of Peirce, in this section, closely follows Douven 2017, Supplement: Peirce on Abduction.

[2] When quoting from Peirce's Collected Papers (CP, Hartshorne, Weiss, and Burks 1931–1958), I follow the convention of citing the number of the volume followed by the number of the relevant paragraph.

logical operation. In fact, he highlights that Peirce does not always talk about abduction as a way of devising hypotheses or new ideas, but also as a process of *adopting* hypotheses. Namely, the hypotheses are not true or verified or confirmed, but they are worthy candidates for further investigations. So conceived, abduction works as a sort of *selection function*, determining which of the hypotheses conceived in the stage of discovery are to pass to the next stage and be subject to empirical testing. The selection criterion is that there must be a reason to suspect that the hypothesis is true, and such a reason is that the hypothesis makes the fact a matter of course. In this sense, abduction can still belong to the context of discovery, and, at the same time, it can have the aforementioned logical form.

However, Frankfurt (*ibid.*, pp. 595-596) rejects this proposal by arguing that there may be infinite hypotheses that account for a given fact, as Peirce himself acknowledged. Thus, it cannot be a sufficient condition for the adoption of the hypothesis that the latter, if true, would make a fact a matter of course.

In this regard, Douven (2017) highlights that Frankfurt's objection may not be valid. Indeed, Frankfurt equates 'accounting for a fact' with 'making the fact a matter of course', by interpreting 'accounting for a fact' as entailment.

But, for Peirce, abduction is the process of finding explanatory hypotheses (see above, CP 5.172), and no philosopher of science, nowadays, would say that entailment is sufficient for explanation.

So, it would be reasonable to read 'making a fact a matter of course' as 'giving a satisfactory explanation of that fact'.

In this sense, there could be infinite hypotheses that account for a given fact, in Frankfurt's sense, but just a handful that give a satisfactory explanation. It remains to see if this could be plausible in light of Peirce's further writings.

Thus, wrapping up, Peirce highlights two functions of abduction:

- Abduction as a procedure that generates new hypotheses[3].
- Abduction as selection function in the sense explained before.

---

[3] Notice that two kinds of abductive novelty can be highlighted in Peirce's writings (Anderson 2013, p. 47): *rearrangement*, which consists of a combination of ideas, different from past views, but grounded on ideas or perceptions we already have; *concept creation*, that is, the creation of a new concept, i.e. idea, which we did not have previously.

3.2.2 Norwood R. Hanson

Similar to Peirce, for Hanson (1958), the act of discovery - i.e. the act of suggesting a new hypothesis - belongs to the logic of abductive inference, a logic of science different both from inductive logic and hypothetico-deductive reasoning. Such a logic, again, takes place *before* a new hypothesis is ultimately justified[4]. .

More specifically, according to Hanson (1960, p. 104), abduction has the following logical form:

1. some surprising, astonishing phenomena p_1, p_2, p_3 ... are encountered.
2. But p_1, p_2, p_3 ... would not be surprising, were an hypothesis of H's type to obtain. They would follow as a matter of course from something like H and would be explained by it.
3. Therefore there is is good reason for elaborating a hypothesis of H's type - namely, for proposing it as a possible hypothesis from whose assumption, p_1, p_2, p_3 ... might be explained.

Thus, we see that, Hanson's discovery is, first of all, a process of explaining anomalies or surprising phenomena. This is what triggers the search of an explanatory hypothesis in light of which the phenomena would no longer be surprising or anomalous. The outcome of such a process is not one single specific hypothesis, but the delineation of a type of hypotheses worth of further consideration (Hanson 1965, p. 64).

Schickore (2018, section 6.1) reports a series of objections that can be made to Hanson's abduction, some of which echo Frankfurt's objections to Peirce's abduction. Firstly, Schickore points out that Hanson's schema is too permissive in that there are several hypotheses that explain the surprising phenomena (Harman 1965, Blackwell 1969). Thus, in absence of additional criteria to evaluate the hypothesis yielded by the abductive inference, the fact that a hypothesis explains the phenomena can hardly be a decisive criterion to develop that hypothesis. Moreover, it is highlighted that Hanson's schema is silent about the process of inventing or discovering a hypothesis, which remains unanalyzed. Instead, the schema focuses on the

---

[4] My treatment of Hanson, in this section, closely follows Schickore 2018, section 6.1.

processes by which an exploratory hypothesis is identified as being worthy of pursuit (Laudan 1980, Schaffner 1993). Indeed, as Paavola (2004) highlights, the aim of Hanson's abduction is to provide plausible candidate hypotheses which then have to be tested by other means.

3.2.3 Gerhard Schurz

In subsection 3.2.1 and 3.2.2, we have seen, respectively, that both Peirce and Hanson propose a very similar logic of generation of new theories. However, we have also noticed that both logics are better understood, not as logics of generation of new theories, but as logics of adopting new hypotheses that turn out to be worthy candidates for further investigations.

A view of abduction very much in this spirit was proposed by Gerhard Schurz in his 2008 paper, *Patterns of Abduction*. Indeed, Schurz (p. 205, emphasis in the original) conceives abduction as "a *search strategy* which leads us, for a given kind of *scenario*, in a reasonable time to a most promising explanatory conjecture which is then subject to further test". Such a general definition encompasses different patterns which are classified along three dimensions (*ibid.*):

1. The kind of abduced hypothesis, i.e. the conjecture worthy of further investigation.
2. The kind of evidence the abduction aims to explain.
3. What drives the abduction.

On this base then, we obtain four kinds of abduction patterns. The first one is *Factual Abduction* (*ibid.*, pp. 206-211). Here, both the evidence to be explained and the abduced hypothesis are singular facts. This kind of abduction is driven by known implicational laws, going from causes to effects. And the abduced hypothesis is found by backward reasoning, inverse to the direction of the lawlike implication. So, factual abduction has the following structure (the double line === indicates that the inference is uncertain and preliminary):

(FA) *Known Law:* if Cx, then Ex
Known Evidence: Ea has occurred

================================

Abduced Conjecture: Ca could be the reason.

Depending on the epistemological nature of the abduced fact, factual abduction can be divided in three subpatterns:

1. *Observable Fact Abduction* (*ibid.*, pp. 207-208), in which Ca is an *observable* cause. Thus, the follow up procedure will consist of gaining direct evidence for the abduced conjecture. In this case, the weak support that the abductive inference provides to the conjecture is *replaced* by the strong support provided by direct evidence[5].

2. *First Order Factual Abduction* (*ibid.*, pp. 208-209), whose safest abduced conclusion is an existential conjecture of the kind ∃yCya. However, while in some cases we may be satisfied with the existential conjecture, in others we need to find out the entity whose existence we conjecture. We will see an example of first order factual abduction in subsection 3.4.2.

3. *Unobservable Fact Abduction* (*ibid.*, pp. 209-210), in which Ca is in principle observable, but pragmatically it is not, because, for instance, it is located in the past, as it is happens in the *Historical Fact Abduction*. So, the only way to confirm Ca is by ascertaining whether further empirical consequences, which follow from the abduced hypothesis and the background knowledge *K*, turn out to be true. If this is the case, then both the evidence which triggered the abduction and the one deduced from the hypothesis and *K* provide epistemic support to Ca. Thus, the initial abductive inference *keeps* its justificatory value.

The second kind of abduction pattern is *Law Abduction* (*ibid.*, pp. 211-212). It is driven by known implicational laws, and both the evidence to be explained and the abduced hypothesis are implicational laws. The conclusion is in strong need of *further* empirical support. An example is the following:

---

[5] Hence, we see that for Schurz the distinction between the use of explanatory power in the context of discovery and its use in the context of justification is blurred. Indeed, he believes that, for many (even if not all) patterns of abduction, the justificational role, although minor, is not absent (Schurz 2008, p. 204). On this point, he agrees with Niiniluoto (1999), who claims that "abduction as a motive for pursuit cannot always be sharply distinguished from considerations of justification" (p. S442).

Known law:	$\forall x(Cx \rightarrow Ex)$	Whatever contains sugar tastes sweet
Evidence to be explained:	$\forall x(Fx \rightarrow Ex)$	All pineapples taste sweet

===

Abduced Conjecture:	$\forall x(Fx \rightarrow Cx)$	All pineapples contain sugar.

The difficulty that the two types of abduction so far considered have is that there are lots of possible hypotheses to choose among - as many as there are known laws - and one has to select the most plausible[6]. In other words, following Magnani's terminology (Magnani 2011, p.20), these kinds of abduction are mainly *selective* in the sense that the most promising candidate is drawn from a given multitude of possible explanations. Conversely, the remaining two types of abduction, i.e., *Theoretical Model Abduction* (Schurz 2008, pp. 213-216) and *Second Order Existential Abduction* (*ibid.*, pp. 216-232) are mainly *creative* (Magnani 2011, *ibid.*). That is, they construct something new, e.g. a new theoretical model or a new concept. Differently from before, the difficulty, here, is not a large search space of possible conjectures, but finding just one plausible conjecture which allows for the derivation of the phenomena to be explained. So, let us see of what these two creative patterns consist.

Theoretical model abduction is driven by known theories. The evidence to be explained is a general empirical phenomenon expressed by an empirical law. The abduced product, instead, consists of new theoretical models of these phenomena. The latter is preliminary confirmed. An example is this:

Evidence to be explained:	certain substances (like stones or metals) sink in water, while some other (like wood or ice) swim in water.
Abduction is driven by:	a given theory which presupposes that the ultimate causes are only contact forces and gravitational forces.

---

[6] Consider, however, that in the case of factual abduction, we are aided by a *probabilistic elimination technique*, according to which, our mind, usually unconsciously, quickly scans through our large set of memorized possible scenarios, and only the ones with minimal plausibility pop up in our consciousness (Schurz 2008, p. 207)

Abduction hypothesis:	the body will sink if its density (mass per volume) is greater than the density of water. Otherwise, it will swim.

Since theoretical model abduction is driven by known theories, it works in a given conceptual space, and, as such, it cannot introduce new concepts. Second order existential abduction, on the contrary, can do this. Indeed, the abduced hypothesis postulates the existence of a new kind of property or relation. Like in theoretical model abduction, the explanandum is a general empirical phenomenon expressed by an empirical law, and, depending on what drives the abduction, we have different subtypes of second order existential abduction. If the abduction is driven by *extrapolation*, we have *Micro Part Abduction* (*ibid.*, pp. 216-217). A case in point is the abduction of the atoms which obey the same laws of the macroscopic objects. Thus, here, one extrapolates from macroscopic concepts and laws to the microscopic domain. If abduction, instead, is driven by analogy, then we are in the presence of the *Analogical Abduction* (*ibid.*, pp. 217-218). For instance, the conjecture according to which the sound consists of atmospheric waves was obtained by noticing an analogy between the propagation and reflection of water waves and the propagation and reflection of sound. If abduction is driven by the search for unification in terms of hidden or common causes, then we have the *Hypothetical (Common) Cause Abduction* (*ibid.*, pp. 218-232). The latter, in turn, is divided into *Speculative Abduction* (*ibid.*, pp. 219-223) and *Common Cause Abduction*. In the first one, we explain one phenomenon as effect of one hypothetical (unobservable) cause, introduced merely for the purpose of explaining that phenomenon. Thus, as the name suggests, speculative abduction is not a scientific worthwhile abduction. Conversely, common cause abduction is a legitimate scientific abduction. It is driven by *proper* unification, and it is divided in three sub-patterns. The first one is the *Strict Common Cause Abduction* (*ibid.*, pp. 223-227), namely, the common cause abduction in deductive settings, appropriate when the domain is ruled by (almost) strict causal laws. An example of the latter was the postulation of a common intrinsic property, called 'metallic character', to explain dispositions like high conductivity of electricity or high conductivity of heat, and so on, which substances like iron, copper, etc. have in common. The second one is a statistical procedure called *Statistical Factor Analysis* (*ibid.*, pp. 228-231). It is a probabilistic version of common cause abduction in that it aims at explaining mutually interconnected variables with more than one cause, by following the principle that the less the number of causes compared to the variables, the higher the success of the explanation. The last one is *Abduction to Reality* (*ibid.*, pp. 231-232), and, basically, it is the reasoning from introspective sense data to an external reality which causes such perception.

As mentioned in the introduction, the aim of the paper is to see whether some of the above patterns, which Schurz identifies in various contexts, are used in the ongoing Covid-19 pandemic. But first, let us briefly introduce the latter.

3.3 The Covid-19 pandemic

The Covid-19 pandemic is a worldwide epidemic of coronavirus disease 2019 (Covid-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The latter is a new type of coronavirus discovered by the Chinese authorities and isolated on 7 January 2020. Such a discovery came as a result of an outbreak of pneumonia cases of unknown etiology in Wuhan (China), then traced back to the new virus (World Health Organization 2020a).

Attempts to contain the virus in Wuhan miserably failed, and, in a relatively short period of time, it spread worldwide. Indeed, on 30 January 2020, the World Health Organization (WHO) declared the outbreak a 'public health emergency of international concern' (World Health Organization 2020b), and a 'pandemic' on 11 March 2020 (World Health Organization 2020c).

In the meantime, attempts were made to stop the spread through preventive measures and government interventions. The main preventive measure was surely the mass vaccination program, started in early December 2020 (World Health Organization 2022c). Other measures included social distancing, masks wearing, ventilation, hands washing, quarantine (World Health Organization 2022d, Coronavirus disease (COVID-19) - Prevention).

Regarding government interventions, they included travel restrictions, business restrictions and/or closure, suspension of teaching activities, contact tracing, and so on (for example, in Italy: Il Post 2022).

As mentioned before, SARS-CoV-2 gives rise to the disease called 'Covid-19', whose symptoms vary. The most common ones are fever, dry cough, fatigue, and loss of taste or smell. Less common symptoms involve sore throat, headache, aches and pains, diarrhea, a rash on skin, discoloration of fingers or toes, red or irritated eyes. Severe Covid-19's symptoms, which require immediate medical attention, are respiratory difficulty or shortness of breath, loss of speech or mobility, confusion, chest pain (World Health Organization 2022e). The people at higher risk of developing serious symptoms are the ones aged 60 or over, pregnant people, people with underlying medical problems like high blood pressure, heart or lung problems, diabetes, obesity or cancer (World Health Organization 2022a).

As far as it concerns SARS-CoV-2's transmission (World Health Organization 2021a), it seems that the virus spreads when aerosols - i.e. infectious particles that pass through the air - are inhaled; or via droplet transmission, that is, when infectious particles come into direct contact with eyes, nose, or month. The transmission so described appears to take place mainly between people who are in close contact with each other, such as when they are at a conversational distance. However, when aerosol is involved, the virus can transmit also in poorly ventilated and/or crowded indoor settings, where people tend to spend longer periods of time. This is because aerosols can remain suspended in the air or travel farther than conversational distance. Another way people may become infected is when they touch their eyes, nose or mouth after touching surfaces or objects that have been contaminated by the virus.

During the pandemic, different variants of SARS-CoV-2 developed with varying degrees of infectivity and virulence. In fact, SARS-CoV-2, like all viruses, change over time as it spreads among people. When such changes, or mutations, become significantly different from the original virus, they are called 'variants' (World Health Organization 2021b).

Now, most changes are of little or no interest, but some changes are possible that affect the virus' features, like how easily it spreads, the severity of the disease it causes, or the effectiveness of vaccines, therapeutic medicines, diagnostic tools, or other public health and social measures (World Health Organization 2022b). Since late 2020, WHO began to classify the variants in Variants of Interest (VOIs) and

Variants of Concern (VOCs) (*ibid.*). VOIs are variants that have mutations suspected or known to affect virus' features such as transmissibility, disease severity, immune escape, diagnostic or therapeutic escape; *and* that are circulating widely (e.g., known to cause many clusters of infected people, or found in many countries). A VOI becomes a variant of concern if it is known to be more transmissible; *or* to cause more severe disease; *or* to lead to a decrease in effectiveness of public health and social measures or available diagnostics, vaccines, therapeutics. The variants of concern so far identified by WHO are Alpha, Beta, Gamma, Delta, Omicron[7] (see table 3.1 for details).

WHO labels	Pango lineage	Earliest documented samples
Alpha	B.1.1.7	United Kingdom Sep-2020
Beta	B.1.351	South Africa May-2020
Gamma	P.1	Brazil Nov-2020
Delta	B.1.617.2	India Oct-2020
Omicron	B.1.1.529	Multiple Countries Nov-2021

Table 3.1: Classification of VOCs

The next section will be dedicated to the first variant of concern identified, namely, the Alpha variant. Particularly, I will present two case studies related to it. Both of them are about the use of abduction in Schurz's sense - i.e., patterns whose structure determines a particularly promising conjecture that needs to be subject to further testing. In the first one, the conjecture is the higher transmissibility of the Alpha variant, in the second one it concerns how the Alpha variant emerged.

---

[7] On 31 May 2021, WHO announced to have assigned labels for key variants of SARS-CoV-2, using letters from the Greek alphabet. The reason behind this was not to replace existing scientific names (e.g. those assigned by Pango), which are still in use. Rather, it was to provide easy to say and remember names to avoid that people call the variants by the places where they were detected, which is stigmatizing and discriminatory (World Health Organization 2021c).

3.4 The Alpha variant: two case studies

As reported by The Guardian (Boseley 2021), in late November 2020, something anomalous was observed in the south east of England. UK had been in lockdown for more than three weeks, and in most places Covid-19 cases were dropping. But, in a few Kent boroughs, they were still going up. Indeed, Swale had the highest infection rate in the country, followed by the nearby Thanet.

People at Public Health England (PHE), who monitor the national pandemic, were puzzled. In fact, when outbreaks occur, it could be that workplaces are hit, or that the surge is located in low-income communities, living in overcrowded houses where social distancing is difficult. However, even if to some degree Swale and Thanet had these problems, these could not explain why new cases continued to rise.

Thus, PHE sent Dr Christina Atchinson to understand what was going on. She was a consultant epidemiologist, and head of the rapid investigation unit, which supports overwhelmed regional teams. She talked to Kent's regional director of public health, and to the deputy director of health protection, namely, the people who evaluate and warn of health hazard. They had different hypotheses about what was going on: (i) Kent commuters were bringing the infection back from London, although case numbers in the capital were not rising overall; (ii) workplaces were not Covid safe; (iii) parents were not keeping their distances during school drop-offs. But, despite the warnings of the local authority, numbers kept rising.

Since Atchinson could not find any obvious reason for the surge, as health protection was doing everything right without getting results, she suggested to sequence samples from people fallen ill. Sample sequencing consists of analyzing the features of the coronavirus' genetic material, and it allows to identify the changes viruses undergo.

At this point, Atchinson shared her idea with Meera Chad, incident director of PHE, microbiologist, and infectious disease consultant. She was also part of the *Tuesday group*, a small group of genomics scientists from around the country that meet online every Tuesday to discuss anything unusual. She agreed with Atchinson that they had to have a look at the genomics, which led to two surprising discoveries. The first one was that more than half of the genomes available for Kent belonged to one huge cluster, then labeled 'Alpha variant' (B.1.1.7). And this was striking because no one expected that such a large number of people would fall ill with the same variant. The second one was that the variant had 23 mutations, which are a lot, considering that Sars-CoV-2 genome accumulates around one or two mutations every month as it circulates. Eight of these mutations were particularly concerning, since they were located on the spike protein, which enables the virus to attack human cells. Thus, they could change the way Sars-CoV-2 interacts with the human host.

The exponential growth of the new variant recorded in the following days, along with its unusual genetics, and the fact that the areas in which it was concentrated were also the places with the highest rates of Covid, were enough to raise the alarm. Indeed, on 19 December 2020, Boris Johson announced that London, the south east, and east would enter tier 4[8]. On 4 January 2021, the prime minister announced a full national lockdown. Different countries closed their borders. However, such measures were not enough to keep the variant at bay, which then became a big factor in Europe's third wave.

The case of the discovery of the Alpha variant was the stage for some patterns of abductions in Schurz's sense, as reported by numerous reports and newspapers' articles. Let us see some examples.

3.4.1 Case study 1: The Alpha variant is more transmissible

In an article of The Guardian (Sample 2020), dated 19 December 2020, we read that cases of Covid-19 were raising in parts of the south of England.

One cause advanced to explain such a phenomenon was the conjecture that a new SARS-CoV-2 variant (the Alpha variant) is more transmissible. This conjecture was based on three pieces of evidence: the new variant is detected in parts of the south of England where Covid-19 cases are raising faster ($\mathbf{E_1}$); the new variant is identified in more than 1,000 cases of Covid-19 in the south of England ($\mathbf{E_2}$); the variant is detected in a rapidly growing number of cases ($\mathbf{E_3}$).

---

[8] Before 19 December 2020, in the United Kingdom there was a system that divided the territories into three risk areas, i.e. tier 1, 2, 3. On 19 December, a new tier (tier 4) was introduced, which imposes stricter lockdown measures.

However, it is specified several times that it is still unsure whether the new variant is more transmissible and that further research is needed.

As for the reason why the Alpha variant may be more transmissible, it said that one possibility could be the mutations acquired by the variant, as it is a known fact that there could be mutations that make the virus transmit faster. Indeed, the article points out that, as seen before, this new variant has multiple mutations in the spike protein, the most troubling of which seems to be a deletion (in this case, the loss of two amino acids from the spike protein) called '**H69/V70**', which may make the virus spread faster.

The same deletion was spotted in SARS-CoV-2's samples collected from a Cambridge patient with a weakened immune system, who, however, was not ill with the Alpha variant. The patient was treated with convalescent plasma, i.e. blood plasma containing antibodies from a recovered patient, but, unfortunately, they died of the infection. During the treatment, the virus acquired the mutation and may have become more resistant to the antibodies.

Similarly, an article of The New York Times, (Landler and Castle 2020), of the same day, says: "But the government's medical experts expressed alarm about its [of the variant] *apparent* infectiousness, *noting that* it now accounts for more than 60 percent of the new infections reported in London" (emphasis is mine). And, a little further on: "But some [scientists] said there was *good reason* for *concern* that this variant is more infectious. Preliminary findings suggest that it is spreading so fast in Britain that it is quickly displacing dozens of competing versions of the coronavirus that have been circulating for longer" (emphasis is mine).

So, we see that the conjecture of the higher transmissibility of the new variant is based on two pieces of evidence: the variant now accounts for more than 60 percent of the new infections reported in London ($\mathbf{E_4}$); the variant is spreading so fast in Britain that it is quickly displacing dozens of competing versions of the coronavirus that have been circulating for longer ($\mathbf{E_5}$).

Again, the cause of such a higher infectivity is thought to be, even if it is not sure, the mutations the variant carries: "But Dr. Vallance, a physician and medical researcher, said scientists had identified 23 changes in the new variant, an unusually large number, including several in the "spike protein" that the virus uses to attach itself to host cells, which could increase its transmissibility".

Thus, we see that patterns are underway here, whose structure determines a particularly promising conjecture that needs to be subject to further testing. To see whether these are patterns of abduction in Schurz's sense, and, if so, which ones, it could be useful to formalize the examples.

The Guardian's example

Known law:

$\forall x(Vx \land Sx \land Tx \to \mathbf{E_1}x \land \mathbf{E_2}x \land \mathbf{E_3}x)$, where V = SARS-CoV-2's variant, S = to be especially present in south of England, T = to be more transmissible, $\mathbf{E_1}$ = to be detected in parts of the south of England where Covid-19 cases are raising faster, $\mathbf{E_2}$ = to be identified in more than 1,000 Covid-19 cases in the south of England, $\mathbf{E_3}$ = to be detected in a rapidly growing number of cases.

Evidence to be explained:

$\mathbf{E_1}\mathrm{a} \land \mathbf{E_2}\mathrm{a} \land \mathbf{E_3}\mathrm{a}$, where a = the Alpha variant.

Conjecture:

$V\mathrm{a} \land S\mathrm{a} \land T\mathrm{a}$.

The New York Times' example

Known laws:

$\forall x(Vx \land S_E x \land Tx \to \mathbf{E_4}x \land \mathbf{E_5}x)$, where V = SARS-CoV-2's variant, S_E = to be detected for the first time in the south of England, T = to be more transmissible, $\mathbf{E_4}$ = to account for more than 60 percent of the new infections reported in London; $\mathbf{E_5}$ = to be spreading so fast in Britain that it is quickly displacing dozens of competing versions of the coronavirus that have been circulating for longer.

Evidence to be explained:

$\mathbf{E_4}\mathrm{a} \land \mathbf{E_5}\mathrm{a}$, where a = the Alpha variant.

Conjecture:

$V\mathrm{a} \land S_E\mathrm{a} \land T\mathrm{a}$.

Now, *prima facie*, these examples seem to belong to Schurz's factual abduction (see subsection 3.2.3). In fact, the pieces of evidence to be explained are singular facts, and the abduced hypothesis is driven by known implicational laws going from causes to effects. More specifically, the abduced hypothesis is obtained by backwards reasoning, inverse to the direction of the lawlike implication.

However, this is not so simple, as the abduced hypothesis is not a singular fact in the sense intended by Schurz, namely, a single entity which in principle is directly observable, such as a particular human. Rather, it is a feature (the higher transmissibility) of a single entity (the new variant), which is not directly observable, but testable.

This means that the follow-up procedure does not consist of gaining direct evidence for the abduced conjecture, as it happens for the observable fact abduction, or how it could be for the first order existential abduction when we are not satisfied with the existential conjecture (see subsection 3.2.3). Rather, as it is the case for the unobservable fact abduction, there is need for further empirical consequences that comes from the abduced hypothesis and the background knowledge *K*. However, we have to bear in mind that there is a crucial difference between our example and the unobservable fact abduction. Namely, in the latter, the abduced hypothesis is a singular fact which cannot be observed because, for instance, it is located in the past. Instead, in our case, the abduced hypothesis is not directly observable because it is a feature and not a singular fact.

Indeed, the procedures following the abduction consisted of verifying further empirical consequences deduced from the abducted hypothesis and *K*, as told in what follows.

In an article by The Guardian of 11 January 2021 (Geddes 2021), we read that the higher transmissibility of B.1.1.7 was investigated by seeing whether the new variant replaced the older variants in the UK population. The latter consequence was, in turn, tested by using different methods. One of these involved comparing the growth in cases of the new variant and old variants within the same population. Another consisted of following the pattern of signals in Covid tests across the UK, looking at the increase in cases of 'S gene dropout' . The reason for this was that, while normally molecular positive Covid tests detect parts of three genes of SARS-CoV-2, such tests, when in contact with the Alpha variant, show positivity just for two genes, but not the 'S gene'.

As for how much more transmissible B.1.1.7 is, the researchers tested various mathematical models to see which one best fitted the observed increase in the new variant. And, they discovered that the best model is the one consistent with an increase in the reproduction number of 0.5 to 1 (Volz, Mishra, et al. 2021)[9]. From this value, it was inferred that the Alpha variant is 50-100 percent more transmissible.

At the same time, researchers tried to investigate if the cause of such a higher transmissibility is indeed the mutations acquired by the variant, in particular the several ones in the spike protein.

As showed by a preliminary analysis (Rambaut 2021), two mutations of particular interest are **N501Y** and the aforementioned deletion **H69/V70**. The former has been shown to increase how tightly the spike protein binds to a receptor on the surface of human cells. The hypothesis is that **N501Y** helps the virus to bind better to the cells. The latter has emerged several times before: as said above, it was identified in viruses that evolved to evade the natural immune response in some immunocompromised patients. It was also found in minks in Denmark, which were, consequently, all executed (McKie 2020).

As for mutation **N501Y**, an article from Il Post (2021), dated 13 January 2021, says that a research group from Texas University (Galveston, USA) was starting some laboratory experiments on hamsters, which have long been used to evaluate the transmission ability of the coronavirus. The researchers' aim was to understand whether the mutation leads to a higher concentration of viral particles in the upper airways of hamsters, compared to other variants of the coronavirus. The idea came from a previous study on another mutation that led to a similar outcome, providing clues regarding the causes of the greater transmissibility of some variants than others. In fact, a higher concentration in the upper airways makes it more likely that larger quantities of coronavirus will spread while breathing, or speaking in the case of humans.

---

[9] The reproduction number or R value is the number of people each person passes the virus to. The higher it is, the more widely the virus spreads. To understand what this figures means, consider the following. Let us say that the old variant was spreading at an R value of 1.5, meaning 10 Covid-positive people transmitted the infection to 15 new people. Since the new variant has an increase in the reproduction number of 0.5 to 1, it would have an R value of 2 to 2.5, so 10 people with Covid-19 would transmit infection to 20 to 25 new people.

The article adds that there were already some clues that lead in this direction: a preprint study of late December had indicated a greater quantity of coronavirus's genetic material in samples, taken from the nose and mouth, of some individuals with B.1.1.7 compared to those with other variants of the virus (Golubchik 2021).

Regarding deletion **H69/V70**, scientists at the Cambridge University had suggested that this mutation increases infectivity two-fold in lab experiments, and that it makes antibodies from the blood of survivors less effective at attacking the virus (Kemp 2021).

3.4.2 Case study 2: How the Alpha variant emerged

As said previously, when Atchinson and Chad decided to have a look at the genomics, to understand what was going on in Kent, they came across anomalous phenomena. One of these was that the Alpha variant showed way more mutations than one would have expected.

It is on this anomalous piece of evidence that the most promising conjecture about how the Alpha variant emerged was formulated, as reported by a BBC's article of 20 December 2020 (Gallagher 2020), and by an article of Il Post, dated 21 December 2020 (2020). In fact, here we read that the most likely explanation of the fact that the Alpha variant is unusually highly mutated is that the variant may have arisen in an immunocompromised person who was chronically infected with SARS-CoV-2.

Moreover, from the section *What evolutionary processes or selective pressures might have given rise to lineage B.1.1.7?* of the online report cited in the former case study (Rambaut 2021), it is explained the reasoning that leads from the anomalous piece of evidence to its most likely explanation. That is, previous studies report high rates of mutation accumulation, over a short period of time, in immunocompromised patients who are chronically infected with SARS-CoV-2 (Avanzato, Matson, et al. 2020; Choi 2020; Kemp 2021). The patients are treated with convalescent plasma (sometimes more than once). Under such circumstances, the evolution of the virus is expected to be very different to the one experienced in typical infection. Indeed: (i) in a patient chronically ill with SARS-CoV-2 - that is, a patient that has had the infection for a long time - the virus has time to mutate into many variants within one body, so that the virus population may be unusually large and genetically diverse; (ii) when antibodies are applied in this situation, they create suitable circumstances for the rapid fixation of multiple virus genetic changes, by wiping out some variants of the virus, and leaving others at least partially resistant to the treatment.

These considerations lead the scientists to hypothesize that the unusual genetic divergence of lineage B.1.1.7 may have resulted, at least in part, from virus evolution within a chronically-infected individual. However, they point out: "this remains a hypothesis and we cannot yet infer the precise nature of this event".

Again, some pattern is underway here, whose conclusion is a most promising conjecture then set out for further empirical test operations. As above, let us formalize the aforementioned example, to see if it is a schema of abduction in Schurz's sense, and, if so, which one. But, before engaging in the formalization, let us see whether it is possible to detect the three dimensions along which Schurz classifies the patterns of abduction (see subsection 3.2.3).

Now, the evidence to be explained is expressed by the proposition: 'the Alpha variant is unusually highly mutated', while the abduced hypothesis by the one: 'the Alpha variant arose in an immunocompromised person who was chronically infected with SARS-CoV-2'.

As for what drives the abduction, we have to refer to the online report's section just mentioned, i.e. *What evolutionary processes or selective pressures might have given rise to lineage B.1.1.7?*. In fact, as pointed out earlier, it expounds the reasoning that leads from the anomalous piece of evidence to its most likely explanation. So, let us reconstruct and make explicit the passages of this reasoning. From the previous studies reporting high rates of mutation accumulation, over a short period of time, in immunodeficient patients chronically infected with SARS-CoV-2 and treated with convalescent plasma, a general law is *inducted*. The law is the following: If a SARS-CoV-2's variant arises in an immunocompromised person, who is chronically ill with SARS-CoV-2 and treated with convalescent plasma, then the variant is unusually highly mutated. The mechanism behind such a law is specified by points (i) and (ii). Relying on this (now, known) law, scientists explain the unusual genetic divergence of lineage B.1.7.7 by hypothesizing that it resulted from virus evolution with a chronically-infected individual, even if they are still unsure. Thus, we can conclude that the abduction is driven by the known law stated just now.

At this point, we are ready to formalize the example, object of this case study.

Known law:
$\forall x \forall y((Vx \wedge P_I y \wedge xAy) \rightarrow \mathbf{E_6}x)$, where V = to be a SARS-CoV-2's variant, P_I = to be a person who is immunocompromised, chronically infected with SARS-CoV-2, and treated with convalescent plasma, A = to arise in, $\mathbf{E_6}$ = to be unusually highly mutated.

Evidence to be explained:
$\mathbf{E_6}\mathrm{a}$, where a = the Alpha variant.

Conjecture:
$\exists y(P_I y \wedge V\mathrm{a} \wedge \mathrm{a}Ay)$.

Now, we see that the abduction is driven by a known implicational law, whose antecedent contains the anonymous variable y, i.e. a variable not contained in the consequent. The evidence to be explained and the abduced hypothesis are singular facts. Specifically, the former instantiates with 'a' the consequent of the known law. By backward chaining, we obtain '$\mathrm{a}Ay$'. Surely, the safest abduced conjecture is the one in which we existentially quantify over the variable y, i.e. $\exists y(P_I y \wedge V\mathrm{a} \wedge \mathrm{a}Ay)$.

That is, here, we have a case of first order existential abduction (see subsection 3.2.3). However, remember that only in same cases we will be satisfied with the existential conjecture, and that in others we need to find out the entity whose existence we conjecture. The example at hand is one of these cases, and the follow up procedure will be to find out the patient in which the Alpha variant originated.

However, such a task seems to be hopeless. In fact, we are told by The Guardian (Boseley 2021) that Atchinson and her team went looking for patient zero, i.e., the first person who contracted the virus, by contacting the 20 first cases they knew about. The first sequences in the database were from two people tested, respectively, on 20 September in Kent and on 21 September in London. But, none of the two could be the first person to have it, because, among other things, they had not had links to immunocompromised people. Moreover, it is specified that, with any certainty, it is not possible to track back beyond the sequenced cases, as virus' samples were destroyed after a few days (the discovery of the Kent variant changed this: labs are now asked to keep the samples for a month).

Jeffrey Barrett, director of the institute's Covid-19 Genomic Initiative, says that his best guess is that the first patient might have been somewhere in London, since it is a big and populous city. Instead, the option that the variant might have originated from another country seems unlikely to him, given that it spread rapidly in the UK before taking hold anywhere else.

3.5 Conclusion

The aim of the paper was to ascertain whether abduction patterns in Schurz's sense are used in the context of the ongoing Covid-19 pandemic. To this end, I proposed two case studies regarding the Alpha variant, the first SARS-CoV-2's variant of concern. The first case study is about the conjecture of the higher infectivity of the Alpha variant, while the second one tackles the conjecture regarding how the Alpha variant originates. The methodology behind both case studies was to bring to light abduction patterns *a là* Schurz used in newspaper's articles and reports on those issues. Interestingly, it turned out that the first pattern identified was not present in Schurz's classification. Indeed, *prima facie* it seemed to belong to factual abduction, in that the evidence to be explained is a singular fact, and the abduction is driven by a known implicational law. However, at a closer look, it could not be the case because the abduced hypothesis is not a singular fact, in principle directly observable, but it is a feature, not directly observable but testable. Instead, the abduction pattern detected in the second case study belongs to Schurz's typology, and it carries the name of 'first order existential abduction'.

Conclusion

The three chapters here presented dealt with the relationship between epistemic values and scientific theories. In **chapter 1**, I proposed two lines of response to save from Arrovian impossibility large scale theory choices based on Kuhn's epistemic criteria. In the first one, I showed that a particular meaning of fruitfulness can be measured on an absolute scale, which led me to suggest that there may be choices in which all the kuhnian criteria employed are measured on such a scale. If this is the case, then a technically meaningful form of inter-criterion comparability is possible, and impossibility can be escaped. In the second one, instead, I proposed to use Kuhn's criteria to evaluate the prior probability and the likelihood, which are, in turn, criteria of the Bayesian theory choice functional. The latter, in fact, is not subject to Arrovian impossibility thanks to Sen's escape route. Such a second line of response consisted of different steps. First, I considered some meanings of Kuhn's criteria, and showed how they relate to the prior probability and likelihood. It turned out that accuracy is reflected by the likelihood, while the four criteria of scope, simplicity, internal and external consistency are reflected by the prior. Since these can pull in different directions, the same moral Okasha drew from Kuhn's argument applied for the evaluation of the prior probability. That is, there are many acceptable algorithms to evaluate the prior probability. At this point, I considered an objection to the second response. Namely, Stegenga argues that an algorithm that translates the information provided by Kuhn's criteria in prior probabilities satisfies Arrow's conditions, and so it is destined to the same impossibility. Now, if this is the case, the second response no longer stands. In fact, Arrovian impossibility would obtain also if we use Kuhn's criteria to determine the criteria of the Bayesian theory choice functional. However, I pointed out that Stegenga's claim is flawed, as the algorithm he has in mind has a different mathematical form from Arrow's social choice rule. As a matter of fact, Stegenga's algorithm outputs a list of real numbers, instead of a ranking of theories. So, we cannot ask whether conditions defined for the Arrovian social choice rule apply to Stegenga's algorithm. Finally, I considered an interesting issue, triggered by Stegenga's observation, namely whether the prior probability rule (PPR), associated with Stegenga's algorithm, suffers from Arrovian impossibility. I showed that an observation Morreau (2015) made against the applicability of condition **U** to Okasha's theory choice rule applies also to condition **U** of PPR, given that some of the Kuhn's criteria that can bear on the prior are rigid in Morreau's sense. And, since it is still an open question whether Arrovian impossibility obtains with a domain restriction caused by rigid criteria, the judgement about the impossibility of our algorithm was suspended.

In **chapter 2** I kept exploring the relationship between epistemic values and Bayesianism, continuing what I started in the second line of response proposed in chapter 1. Here, I focused on the value of 'explanatory power', which refers to the goodness of an explanation. Specifically, I considered an inference in which the explanatory power of a hypothesis acquires a special status, namely, the inference to the best explanation (IBE) or abduction. And, I explored how IBE relates to a well known problem of Bayesian confirmation theory, that is, the problem of old evidence (POE). More precisely, I took into consideration two novel solutions to the dynamic POE, recently proposed by Eva and Hartmann, and I pointed out a shortcoming of the second one. Then, I made explicit that these solutions can be read in terms of inference to the best explanation. Such a move allowed me to underline that Eva and Hartmann have in mind an understanding of IBE very similar, even if not equal, to **ABD2**. According to their understanding, if H is the best explanation of E, in the sense that it is the only available hypothesis that adequately explains E, than H's probability assumes a high value. On this base, I further assessed the two models. Namely, I showed that - pending the question if Eva and Hartmann's idea of IBE is descriptively accurate - their first model is not adequate to solve the dynamic POE in IBE's terms, while the second one is. In fact, condition **HF4*** of the first model is not expression of their IBE's idea. Conversely, the only condition of their second model, i.e. **A1**, is. Finally, I pointed out that the explicit realization that real cases of confirmation by old evidence are instances of IBE implies that the static dimension of POE has to be, now, expressed in IBE terms. I tried to solve the static dimension of POE so expressed by relying on the counterfactual approach - i.e. the standard approach to static POE - and the Bayesian IBE. The latter is a probabilistic version of IBE which use explanatory considerations as an aid to evaluate the terms in Bayes' theorem. Indeed, the advocates of the latter show that better explanations have higher prior and/or likelihood, and, anyhow, that they are the ones with the higher product of these two quantities. But, I highlighted that the problems of the counterfactual approach are still present when it is used to solve the static POE in IBE terms. For instance, in the GTR's example, if the old evidence E had not been known, we would not have what we want to explain.

By interpreting 'abduction' as 'inference to the best explanation', in chapter 2 I focused on the use of explanatory power in *justifying* hypotheses. Conversely, in **chapter 3**, I considered another sense of abduction which refers to the use of explanatory power in *generating* hypotheses, especially endorsed by Peirce and Hanson. However, it was seen that their very similar logics are better understood, not as logics of generation, but as logic of adoption of new hypotheses as worthy candidates for further investigations. Then, a view of abduction very much in this spirit was introduced, namely the one defended by Gerhard Schurz (2008). Schurz, indeed, proposes different abduction patterns which are intended as search strategies which lead to a most promising explanatory conjecture, then subject to further tests. In chapter 3, my aim was to see whether abduction patterns in Schurz's sense are used in the context of the ongoing Covid-19 pandemic. To this end, I proposed two case studies concerning the Alpha variant, the first SARS-CoV-2's variant of concern. The first case study dealt with the conjecture of the higher infectivity of the Alpha variant, while the second one tackled the conjecture regarding how the Alpha variant originated. In both case studies, I relied on the same methodology, which consisted of highlighting the abduction patterns *a là* in Schurz used in newspaper's articles and reports on those issues. Surprisingly, the abduction pattern identified in the first case study was not present in Schurz's classification. In fact, despite the similarities, such a pattern can not belong to the factual abduction, since the abduced hypothesis is not a singular fact, in principle directly observable, but a feature, not directly observable but testable. In the second case study, instead, the abduction pattern detected belonged to Schurz's typology, and it is called 'first order existential abduction'.

As a final remark, I would like to add that perhaps, besides these more particular results, something broader can be drawn from this thesis. Chapter one is based on a literature that truly shows how formal methods can provide important insights on popular and established philosophical claims: now we know that the acceptability of different trade-offs of the values is not so easy to defend, as Kuhn and its successors thought and some of the latter still think. On the other hand, I hope to have shown that even the philosophical reflection concerning large scale and somehow older issues can have a place in these more recent developments. Chapter two, instead, wants to deep on what already started by Eva and Hartmann, and to ascertain what happens when we integrate formal methods with longstanding philosophical concepts, like the one of inference to the best explanation. Finally, chapter three, at the opposite of chapter one, aimed to point out how philosophy is underway even in the most recent (and tragic) times.

Bibliography

Amartya, Sen (1970), *Collective Choice and Social Welfare*, Holden-Day, San Francisco.

— (1977), "On Weights and Measures: Informational Constraints in Social Welfare Analysis", *Econometrica*, 45, pp. 1539-1572.

— (1986), "Social Choice Theory", in *Handbook of Mathematical Economics*, ed. by K. Arrow & M. Instiligator, Elsevier, North-Holland, pp. 1073-1181.

Anderson, Douglas R (2013), *Creativity and the Philosophy of CS Peirce*, Springer Science & Business Media, vol. 27.

Arrow, Kenneth (1951), *Social choice and Individual Values*, John Wiley, New York.

Avanzato, Victoria A, M Jeremiah Matson, Stephanie N Seifert, Rhys Pryce, Brandi N Williamson, Sarah L Anzick, Kent Barbian, Seth D Judson, Elizabeth R Fischer, Craig Martens, et al. (2020), "Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer", *Cell*, 183, pp. 1901-1912.

Bird, Alexander (2018), "Thomas Kuhn", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Winter 2018 edition, Metaphysics Research Lab, Stanford University.

Blackwell, Richard Joseph (1969), *Discovery in the Physical Sciences*, Notre Dame [Ind.]University of Notre Dame Press.

Boseley, Sarah (2021), "'Has everyone in Kent gone to an illegal rave?': on the variant trail with the Covid detectives", *The Guardian*, `https://www.theguardian.com/world/2021/apr/03/has-everyone-in-kent-gone-to-illegal-rave-on-variant-trail-with-covid-detectives` (visited on 04/07/2022).

Bovens, Luc and Stephan Hartmann (2003), "Solving the riddle of coherence", *Mind*, 112, pp. 601-633.

Brush Stephen, G. (1989), "Prediction and Theory Evaluation: The Case of Light Bending", *Science*, 246, pp. 1124-1129.

Campos, Daniel G (2011), "On the distinction between Peirce's abduction and Lipton's inference to the best explanation", *Synthese*, 180, pp. 419-442.

Choi, Bina et all (2020), "Persistence and evolution of SARS-CoV-2 in an immunocompromised host", 383, pp. 2291-2293.

Crelinsten, Jeffrey (2013), *Einstein's jury*, Princeton University Press, Princeton.

Dawid, Richard, Stephan Hartmann, and Jan Sprenger (2015), "The No Alternatives Argument", *British Journal for the Philosophy of Science*, 66, pp. 213-234.

Douven, Igor (2017), "Abduction", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Summer 2017 edition, Metaphysics Research Lab, Stanford University.

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

Eells, Ellery (1985), "Problems of Old Evidence", *Pacific Philosophical Quarterly*, 66, pp. 283-302.

— (1990), "Bayesian Problems of Old Evidence", in *Scientific Theories*, ed. by C. W. Savage, University of Minnesota Press, Minneapolis, pp. 205-223.

Eva, Benjamin and Stephan Hartmann (2020), "On the origins of old evidence", *Australasian Journal of Philosophy*, 98, pp. 481-494.

Frankfurt, Harry G (1958), "Peirce's notion of abduction", *The Journal of Philosophy*, 55, pp. 593-597.

Gaertner, Wulf (2001), *Domain Conditions in Social Choice Theory*, Cambridge University Press.

Gallagher, James (2020), "New coronavirus variant: What do we know?", *BBC*, `https://www.bbc.com/news/health-55388846` (visited on 04/08/2022).

Garber, Daniel (1983), "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory", in *Testing Scientific Theories*, ed. by John Earman, University of Minnesota Press, Minneapolis, pp. 99-132.

Geddes, Linda (2021), "The new UK Covid variant: your questions answered", *The Guardian*, `https://www.theguardian.com/uk-news/2021/jan/11/the-new-uk-covid-variant-your-questions-answered` (visited on 04/06/2022).

Glymour, Clark (1980), *Theory and Evidence*, Princeton University Press, Oxford.

Golubchik, Tanya et all (2021), *Early analysis of a potential link between viral load and the N501Y mutation in the SARS-COV-2 spike protein*, `https://www.medrxiv.org/content/10.1101/2021.01.12.20249080v1` (visited on 05/20/2022).

Hanson, Norwood Russell (1958), "The logic of discovery", *The Journal of Philosophy*, 55, pp. 1073-1089.

— (1960), "Is there a logic of scientific discovery?", *Australasian Journal of Philosophy*, 38, pp. 91-106.

— (1965), *Patterns of discovery: An inquiry into the conceptual foundations of science*, CUP Archive.

Harman, Gilbert H (1965), "The inference to the best explanation", *The philosophical review*, 74, pp. 88-95.

Hartmann, Stephan and Branden Fitelson (2015), "A New Garber-Style Solution to the Problem of Old Evidence", *Philosophy of Science*, 82, pp. 712-717.

Hartmann, Stephan and Jan Sprenger (2019), *Bayesian Philosophy of Science*, Oxford University Press, Oxford.

Hartshorne, C., P. Weiss, and A. W. Burks (1931–1958), *The Collected Papers of Charles Sanders Peirce*, Harvard University Press.

Howson, Colin (1984), "Bayesianism and Support by Novel Facts", *British Journal for the Philosophy of Science*, 35, pp. 245-251.

— (1985), "Some Recent Objections to the Bayesian Theory of Support", *British Journal for the Philosophy of Science*, 36, pp. 305-309.

— (1991), "The 'Old Evidence' Problem", *British Journal for the Philosophy of Science*, 42, pp. 547-555.

— (2017), "Putting on the Garber Style? Better Not", *Philosophy of Science*, 84, pp. 659-676.

Il Post (2020), "Cosa sappiamo della nuova variante del coronavirus", *Il Post*, `https://www.ilpost.it/2020/12/21/variante-coronavirus-b117/` (visited on 04/07/2022).

— (2021), "Cosa fare contro le varianti del coronavirus", *Il Post*, `https://www.ilpost.it/2021/01/13/variante-inglese-sudafricana-coronavirus/` (visited on 04/07/2022).

— (2022), "Il governo estende le restrizioni a tutta Italia", *Il Post*, `https://www.ilpost.it/2020/03/09/isolamento-coronavirus-italia`

/?fbclid=IwAR0-8XBJqprcc-dFVCv59rLx6muqegmP_PUwf7zHAdNrtu-uGKse8Y8bsic (visited on 06/01/2022).

Jeffrey Richard, C. (1983), "Bayesianism with a Human Face", in *Testing Scientific Theories*, ed. by John Earman, University of Minnesota Press, Minneapolis, pp. 133-156.

Kelsey, David (1987), "The Role of Information in Social Welfare Judgements", *Oxford Economic Papers*, 39, pp. 301-317.

Kemp, Steven et all (2021), "SARS-CoV-2 evolution during treatment of chronic infection", *Nature*, 592, pp. 277-282.

Kuhn, Thomas S. (1970), *The Structure of Scientific Revolutions*, Second Edition, University of Chicago Press, Chicago.

— (1977), "Objectivity, Value Judgment, and Theory Choice", in *The Essential Tension: Selected Studies in Scientific Tradition and Change*, University of Chicago Press, pp. 320-39.

Kuipers Theo, A. F. (2000), *From Instrumentalism to Constructive Realism: On Some Relations Between Confirmation, Empirical Progress, and Truth Approximation*, Springer.

Landler, Mark & and Stephen Castle (2020), "Boris Johnson Tightens U.K. Lockdown, Citing Fast-Spreading Version of Virus", *The New York times*, https://www.nytimes.com/2020/12/19/world/europe/coronavirus-uk-new-variant.html (visited on 04/07/2022).

Laudan, Larry (1980), "Why Was the Logic of Discovery Abandoned?", in *Scientific Discovery*, ed. by T. Nickles, D. Reidel, Dordrecht, vol. 1, pp. 173-83.

— (1990), "Demystifying Underdetermination", in *Scientific Theories*, ed. by C. Wade Savage, University of Minnesota Press, pp. 267-97.

Lipton, Peter (1993), "Is the Best Good Enough?", *Proceedings of the Aristotelian Society*, 93, pp. 89-104.

— (2001), "Is explanation a guide to inference? A reply to Wesley C. Salmon", in *Explanation*, ed. by G. Hon and S. S. Rakover, Kluwer Academic, Dordrecht, pp. 93-120.

— (2004), *Inference to the best explanation*, 2nd ed., Routledge, London.

List, Christian (2013), "Social Choice Theory", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Winter 2013 edition, Metaphysics Research Lab, Stanford University.

Longino, Helen E. (1990), *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*, Princeton University Press.

— (1996), "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy", in *Feminism, Science, and the Philosophy of Science*, ed. by Lynn Hankinson Nelson and Jack Nelson, Kluwer Academic Publishers, pp. 39-58.

Magnani, Lorenzo (2011), *Abduction, reason and science: Processes of discovery and explanation*, Springer Science & Business Media.

McAuliffe, William HB (2015), "How did abduction get confused with inference to the best explanation?", *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy*, 51, pp. 300-319.

McKie, Robin (2020), "Veterinary workers cull 17 million Danish mink to halt new Covid", *The Guardian*, `https://www.theguardian.com/world/2020/nov/08/veterinary-workers-cull-17-million-danish-mink-to-halt-new-covid` (visited on 04/07/2022).

McMullin, Ernan (1982), "Values in Science", *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1982, pp. 3-28.

— (2008), "The Virtues of a Good Theory", in *The Routledge Companion to Philosophy of Science*, ed. by Martin Curd and Stathis Psillos, Routledge.

Morreau, Michael (2015), "Theory Choice and Social Choice: Kuhn Vindicated", *Mind*, 124, pp. 239-262.

— (2019), "Arrow's Theorem", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Winter 2019 edition, Metaphysics Research Lab, Stanford University.

Musgrave, Alan (1988), "The ultimate argument for scientific realism", in *Relativism and realism in science*, ed. by R. Nola, Kluwer, Dordrecht, pp. 229-252.

Niiniluoto, Ilkka (1983), "Novel Facts and Bayesianism", *British Journal for the Philosophy of Science*, 34, pp. 375-379.

— (1998), "Verisimilitude: The third period", *The British Journal for the Philosophy of Science*, 49, pp. 1-29.

— (1999), "Defending abduction", *Philosophy of science*, 66, S436-S451.

Okasha, Samir (2000), "Van Fraassen's critique of inference to the best explanation", *Studies in History and Philosophy of Science Part A*, 31, pp. 691-710.

— (2011), "Theory Choice and Social Choice: Kuhn Versus Arrow", *Mind*, 120, pp. 83-115.

— (2015), "On Arrow?s Theorem and Scientific Rationality: Reply to Morreau and Stegenga", *Mind*, 124, pp. 279-294.

Paavola, Sami (2004), "Abduction as a logic and methodology of discovery: The importance of strategies", *Foundations of Science*, 9, pp. 267-283.

Popper Karl, R. (2002), *The Logic of Scientific Discovery*, Reprint of the revised English 1959 edition. Originally published in German in 1934 as "Logik der Forschung", Routledge, London.

Quine, W. V. O. and J. S. Ullian (1978), *The Web of Belief*, Second Edition, New York: Random House.

Rambaut, Andrew et all (2021), *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations*, `https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563` (visited on 04/08/2022).

Reiss, Julian and Jan Sprenger (2020), "Scientific Objectivity", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Winter 2020 edition, Metaphysics Research Lab, Stanford University.

Salmon Wesley, C. (2001), "Explanation and confirmation: A Bayesian critique of inference to the best explanation", in *Explanation*, ed. by G. Hon and S. S. Rakover, Kluwer Academic, Dordrecht, pp. 61-91.

Salmon, Wesley (1990), "Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes", in *Scientific Theories*, ed. by C. Wade Savage, University of Minnesota Press, pp. 14-175.

Sample, Ian (2020), "How a new Covid strain may have spread virus in south of England", *The Guardian*, `https://www.theguardian.com/world/2020/dec/14/how-a-new-covid-strain-may-have-spread-virus-in-south-of-england` (visited on 04/08/2022).

Schaffner, Kenneth F (1993), *Discovery and explanation in biology and medicine*, University of Chicago press.

Schickore, Jutta (2018), "Scientific Discovery", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Summer 2018 edition, Metaphysics Research Lab, Stanford University.

Schurz, Gerhard (2008), "Patterns of Abduction", *Synthese*, 164, pp. 201-234.

Sprenger, Jan (2015), "A Novel Solution of the Problem of Old Evidence", *Philosophy of Science*, 82, pp. 383-401.

Stegenga, Jacob (2015), "Theory Choice and Social Choice: Okasha Versus Sen", *Mind*, 124, pp. 263-277.

Teller, Paul (1973), "Conditionalization and observation", *Synthese*, 26, pp. 218-258.

Van Fraassen Bas, C. (1989), *Laws and Symmetry*, Oxford University Press, Oxford.

Volz, Erik, Swapnil Mishra, Meera Chand, Jeffrey C Barrett, Robert Johnson, Lily Geidelberg, Wes R Hinsley, Daniel J Laydon, Gavin Dabrera, Áine O'Toole, et al. (2021), "Assessing transmissibility of SARS-CoV-2 lineage B. 1.1. 7 in England", *Nature*, 593, pp. 266-269.

World Health Organization (2020a), "Novel Coronavirus (2019-nCoV): situation report, 1", *World Health Organization*, `https://apps.who.int/iris/handle/10665/330760`.

— (2020b), *Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)*, `https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)` (visited on 06/01/2022).

— (2020c), *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*, `https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020` (visited on 06/01/2022).

— (2021a), *How does COVID-19 spread between people?*, `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted` (visited on 06/01/2022).

— (2021b), *What are variants of SARS-COV-2, the virus that causes COVID-19?*, `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-(covid-19)-variants-of-sars-cov-2` (visited on 06/03/2022).

— (2021c), *who announces simple easy to say labels for sars cov 2 variants of interest and concern*.

— (2022a), *COVID-19: symptoms and severity*, `https://www.who.int/westernpacific/emergencies/covid-19/information/asymptomatic-covid-19` (visited on 06/01/2022).

— (2022b), *Tracking SARS-CoV-2 variants*, `https://www.who.int/activities/tracking-SARS-CoV-2-variants` (visited on 06/03/2022).

World Health Organization (2022c), *What vaccines are there against COVID-19?*, `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-(covid-19)-vaccines` (visited on 06/01/2022).

— (2022d), *Coronavirus disease (COVID-19) - Prevention*, `https://www.who.int/health-topics/coronavirus#tab=tab_2` (visited on 06/01/2022).

— (2022e), *Coronavirus disease (COVID-19) - Symptoms*, `https://www.who.int/health-topics/coronavirus#tab=tab_3` (visited on 06/01/2022).

List of Tables