Understanding science of the new millennium

Science has become an increasingly complex phenomenon. In one way or another it affects lives of each of us – not only of those who devote their careers to science, but even of those who attack science.

The task of understanding science used to be a challenge for philosophers; in extensive periods of the 20th century one could even have the impression of this being their primary task. But science has changed more rapidly than philosophy, and those philosophers who still try to develop a systematic account of how it works, or should work, are few and it appears that – either for principled reasons or not – the task of giving a systematic account has been commonly abandoned. Thus, when we come to discuss the contemporary conceptions of science in Section 3, it turns out that most of these stem from the first half of the 20th century.

To a large extent the dynamics of science of the late 20th century has been stimulated by ideas and technologies afforded by computer science, itself a characteristic 20th century phenomenon. Indeed, the automation of science goes well beyond handling of the data and can be more and more systematically traced in the process of theory formation. Given the vast amount of data and possible explanations thereof, searching for the true hypothesis appears no longer an exclusively human task. The automation of science bears all the marks of scientific revolution, though a quiet one.[1]

A perspicuous illustration of this revolutionary change is the research on the effects of low-level exposure to lead on children's cognitive abilities as measured by IQ tests. This is one of those epidemiological problems with huge number of possible factors and weak effects which prove particularly difficult to handle with traditional methods. On several computer-assisted analyses of the data H. Needleman discovered a robust weak negative effect of low-level lead exposure on IQ. The upshot of this discovery was the elimination of lead from gasoline in many countries. But Needleman's result was put into doubt by S. Klepper and M. Kamler who questioned the accuracy of the measure of lead concentration in children and IQ tests as a manifestation of their cognitive ability. Given the measurement error they proved that the effect could be zero or even positive. The TETRAD program, described in Subsection 2.1 below, not only confirmed the direction of Needleman's result's, but even helped to demonstrate that the malign effect of lead exposure has been twice as large.[2]

Any serious attempt to give an account of the cognitive aspect of science – as contrasted with e.g. its social or cultural aspects – cannot ignore the automation revolution. In the conception presented in this paper the results of computer science are taken seriously and integrated with many of the ideas concerning what constitutes scientific inquiry that have been proposed at least

[1] For numerous illustrations from physics, astronomy, genetics, medicine, epidemiology and Earth sciences, see (Glymour 2004).

since the early Middle Ages. The central idea is that of *reliable* inquiry. Science makes explicit and elaborates on the methods of acquiring beliefs in our daily life. The central question a scientific method applied to inquiry needs to face is if it is reliable in getting to the truth. On the reliabilist criteria presented here, if a method is logically warranted to get to the right answer – given data and background knowledge available – and to stick to it afterwards, then it is reliable. This is a normative theory, but the standard of reliability is adjusted to the domain of the inquiry at hand.

This paper presents the central tenets of the reliabilist conception of science and briefly outlines the main results underlying it. Roughly, the aim of science is interesting truth about the world around us (scientific realism) and reliabilist epistemology affords us precise answers as to how far science can succeed in this task – given the methods, goals and background assumptions available. The philosophical task to deliver an adequate understanding of science is taken to be continuous with scientific research itself (naturalism), a major part of which is concerned with delivering causal explanations (causality) and can only be carried out with limited resources (computability). Many of the ideas integrated into the reliabilist conception of science and precisely articulated therewith have appeared earlier in both philosophy and science (history of ideas).

I conclude the exposition of the ideas integrated into the reliabilist conception of science with discussions of sample case studies where the reliabilist conception of science has been applied to actual scientific research.

---

[2] A detailed account of this research is in (Scheines 1999; Glymour 2004).

For obvious reasons none of the grand conceptions of science from the distant past is adequate.[3] Towards the end of this paper I give some substance to the claim that among the few contemporary conceptions of science reliabilist epistemology is the the most comprehensive account. For only it systematically elaborates the ideas of reliability and computability accompanying science from its inception. None of the past or contemporary alternatives to a thoroughly revised understanding of science offered by the reliabilist approach can embrace the recent outburst of new sciences such as computer science, Bayesian statistics and cognitive science.

An outline of open problems and directions for future research concludes the paper.

1. Reliabilist epistemology

Knowledge is true belief acquired by application of a reliable procedure. This applies to both our common sense knowledge of the surrounding world and scientific knowledge thereof.[4] I know that I see a white university shuttle approaching me for this belief has been acquired by a mechanism of visual perception which in normal circumstances projects that what I see in fact is what it looks like to me and is reliable in getting me to the truth. Similarly, we

[3] Cf. (Kamiński 1992; Bronk 2001, 147-51).

[4] The principal difference being that of the methods applied in acquiring beliefs. It is commonly recognized that initially epistemological reliabilism was formulated by F. P. Ramsey in (Ramsey 1931). For references to alternative formulations, including R. Nozick's and A. I. Goldman's see (Kawalec 2003, f. 11 on page 102).

know that smoking causes lung cancer for this has been arrived at by a method of inquiry which established that this dependence is not explained away by considering other phenomena and is reliable.

How is scientific knowledge related to truth? When is a method of acquiring beliefs reliable? In articulating the reliabilist conception of science I answer these questions, and then proceed to discuss how various forms of scientific inquiry are related to each other, what kind of explanation prevails in science, what are the essential limitations to science and how it all can be traced back into history of science and philosophy.

1.1 Scientific realism

Science aims at truth. This apparently straightforward claim is interpreted by philosophers of science in two fundamentally divergent ways. According to scientific realists, theories and models tell us what the world is like. The inherent advance in science is to transcend the boundaries of the observable macro world, and tell the true story about the micro and mega worlds which we sometimes cannot observe. The laws of nature or causal dependencies delivered to us by science hold between phenomena regardless of which of them we humans could observe. According to scientific realists our belief in the scientific story about the world follows: we have the same epistemic attitude regardless of which parts of the story cover the realm of the observable.

Scientific anti-realists on the other hand attribute primary importance to the observational-nonobservational distinction. On this view science has to give a true account of what is observable and this we need to believe as true. What science is committed to tell us about the nonobservable world is only instrumental in purpose, and this part of the scientific story about the world we should merely accept, but not believe.

The main argument of the anti-realist viewpoint is from underdetermination of theory by empirical evidence. In a nutshell the argument runs as follows. Suppose you have a given body of empirical evidence, e.g. data from (longitudinal) observational studies of smokers. What the data suggest is that the death rate from lung cancer among smokers compared to non-smokers is almost 9 to 1.[5] Moreover, the rate linearly increases with the number of cigarettes smoked a day. One explanation of the evidence is that smoking causes lung cancer. Admittedly, at this point there are more explanations available. For instance, it could be that the tendency to smoke and to contract lung cancer are causally independent even though they are statistically associated, because they both are an effect of a genetic factor.[6] Anti-realists should now emphasize that these alternative explanations are both empirically adequate (i.e. they equally well fit the data), so how we choose between them turns on some other virtues and has nothing to do with

[5] See (Doll and Hill, 1950; 1952; 1954; 1956; Doll 2002).

[6] This hypothesis was introduced and vehemently defended by one of the classics of statistics R. A. Fisher (1959). A study on twins with discordant smoking habits gave the ultimate response; cf. (Kaprio and Koskenvuo 1989). The rationale for Fisher's approach is offered by R. Doll (2002).

their truth or falsity. What counts is their simplicity, informativeness or some other pragmatic[7] value.

One strategy that realists can use in response is the following.[8] Science reaches the truth by accomplishing the best explanation of the data. We have to articulate the criteria of what constitutes a good explanation in science, and only then ask if what constitutes the best explanation of evidence is confined to the observable. As it turns out the observable-nonobservable dichotomy is inessential in accounting for the best explanation. However the conclusion may be drawn, it may well be that in some cases the best explanation of the evidence is provided in terms of observational entities alone. On this account anti-realism turns out either to impose historically inadequate restrictions on what counts as an explanation in science or to arbitrarily constrain their application to what is observable alone.

Now I proceed to give some substance to the claim that explanation does not turn on the observable-nonobservable distinction. A hypothesis or a theory explains given phenomenon if it demonstrates it to be a manifestation of another phenomenon or regularity already established. If there are two hypotheses explaining a given phenomenon, than scientists prefer one which explains more established regularities in terms of other established regularities. For instance, Copernican theory is preferable to Ptolemaic, because it is only the former which explains the relations between numbers of

---

[7] Pragmatic means in this context a property that is relative to human perspective and independent of semantic, i.e. world-related property, like truth.
[8] Cf. (Glymour 1985, 99-117).

orbits and revolutions in longitude and solar years.[9] Further, if two hypotheses explain the same regularities, than scientists prefer one which explains them not in terms of other regularities, but by means of necessary truth.[10]

For an explanation to demonstrate one regularity as a manifestation of another requires imposing theoretical structure on the phenomena. And whether this structure turns out to be observable or not does not influence the merit of the explanation considered. Belief in the truth of a hypothesis does depend on how well it explains the evidence, but can only be arbitrarily forced to depend on the observable-nonobservable distinction.

Scientific search for the best explanation is often confined to search for causal structure among the phenomena considered. As Clark Glymour et al. (1987, 6) succinctly words it:

> The most common form of explanation in the sciences is to account for why things happen as they do by appealing to the causal relations among events [...].

Scientific realism in this case manifests itself in the search for the true causal structure underlying observed phenomena. This tenet of the reliabilist conception of science is discussed in Subsection 1.4 below.

1.2 The logic of reliable inquiry

[9] (Glymour 1985, 110-11).

[10] An example of such explanation is the explanation of the motion of Mercury by the general relativistic conservation laws (Glymour 1985, 111-12).

Admittedly, there is a plethora of methods available for consideration in arriving at the explanation of the evidence. To choose between these methods scientists need to invoke epistemic norms which would themselves not depend on the subject matter under investigation. Such norms are provided by the logic of reliable inquiry. The essential idea is that a reliable method projecting explanations of the evidence is such that after having changed its mind a finite number of times it gets the true answer and sticks to it afterwards.[11] In other words, the reliable method is logically guaranteed to converge to the truth in the limit. By weakening the sense of convergence (e.g. slower or allowing for more changes of mind) we can obtain reliable solutions to more problems and obtain a corresponding hierarchy of problems facing empirical inquiry.[12]

The application of the norms of reliable inquiry turns on quite general constraints that there be a problem, (countable) alternatives to solve it, and data. These norms afford us domain dependent criteria that equally well apply in natural as well as social sciences, both experimental and nonexperimental.

The most elaborate exposition of the logic of reliability is in (Kelly 1996). I bring out here some of the results in a less formal fashion. The basic notion is that of *evidence item*, i.e. the smallest integral piece of data delivered to us by experience and recorded in a language, e.g. as a number or by means of predicates of first-order logic, for instance $C(a_i)$ denoting "An individual observed contracted cholera". The *set of evidence items* sets up the space of possible observations which are relevant for a given study, e.g. E = {CW,

[11] (Putnam 1965; Kelly 1996; Glymour 1996).
[12] (Kelly 1996, 4; Harrell and Glymour 2002, 260-62).

$\overline{C}W, C\overline{W}, \overline{C}\overline{W}\}$ where CW(a_i) denotes that "An individual was observed who contracted cholera and drank contaminated water" and negation being represented as a bar over a letter.

It is useful to introduce the notion of a sequence of observed evidence items, e.g. three individuals observed contracted cholera and drunk contaminated water (CW, CW, CW). A finite sequence of length n is denoted by $e = (e_1, e_2, \ldots, e_n)$, and an infinite one – called *data stream* – as $\varepsilon = (e_1, e_2, \ldots, e_n, \ldots)$. The first n initial items in ε are denoted by $\varepsilon|n$, e.g. for ε = (CW, CW, …, CW, …) where all evidence items are CW, $\varepsilon|2$ = (CW, CW).

Utilizing these definitions I can proceed to characterize scientific notions such as *hypothesis*. As pointed out in the preceding section an hypothesis aims at an (typically, causal) explanation by generalizing the available evidence. Accordingly a hypothesis admits some possible data streams, but rules out others. Its content then could be represented as a set of admissible data streams. Some hypotheses, like "Everyone who drinks contaminated water contracts cholera", uniquely determine possible data streams, e.g. {ε = (CW, CW, …, CW, …)}.

Background knowledge which is available preceding the study limits the set of logically possible data streams. For instance, researchers investigating the causes of the outbreaks of cholera in the 19th century London ruled out all data streams that would have had only CW and no $\overline{C}W$ up to time t and then, conversely only $\overline{C}W$ and no CW. In other words, irrespectively of the propounded hypothesis the researchers conceded that contaminated water was an important factor in contracting cholera – the question rather was whether it was the only one, or a minor one.[13] Analogously as in the case of hypotheses, background knowledge can be represented in the above introduced framework as a set of possible data streams, i.e. those which are admissible given the available background knowledge.

As there are many hypotheses compatible with background knowledge we need to face a *discovery problem* (H, K) – where H is a set of hypotheses covering data streams compatible with the background knowledge K – of selecting the true hypothesis.

One of the upshots of the idea of scientific realism discussed in Subsection 1.1 is that science ought to be objective and whether the research is carried out by one person or another is immaterial insofar as they both use the same *scientific method* in the proceeding. Thus in characterizing reliability I refer to a method δ whose primarily objective is to project a true hypothesis explaining the observed phenomena. The hypothesis that a method δ projects on the basis of finite evidence sequence e I denote as $\delta(e) = H$. I assume that there are countably many alternative hypotheses and each method can consistently output at most one hypothesis in response to a given evidence sequence. John Snow, one of those searching for the causes of cholera in the 19th century London, used a method $\delta(\varepsilon|n)$ = "All CW" for all n, which for any

[13] In the then used categories the distinction was between an "active agent" and a "predisposing cause". In his introductory text to J. Snow's *On cholera* W. Frost pointed out that the common agreement was that contaminated water is a predisposing cause, but it was the question of its being the active agent that was at issue (1936, xi).

evidence sequence projected the hypothesis “All CW”, i.e. “All individuals contracting cholera drink contaminated water”.[14]

In defining reliability the central notion is that of *convergence* to a true answer. A method δ *converges* to a hypothesis H on a data stream ε *by time n* if and only if for any time n' later than n $\delta(\varepsilon|n') = H$. If there is a time n such that method δ *converges* to a hypothesis H on a data stream ε *by time n* then we can say that δ *converges* to a hypothesis H without further qualification.

In a discovery problem (H, K) a method δ *succeeds* on a data stream ε in K iff δ converges to the true hypothesis on ε. The discovery problem (H, K) is *solved* by a method δ iff δ succeeds on all data streams in K. A method solving a discovery problem is called *reliable*. If δ is a reliable method for a problem (H, K), then δ converges to a true hypothesis on every data stream ε compatible with the background knowledge K.

If at time n a method $\delta(\varepsilon|n)$ outputs a false hypothesis then we say that it *commits an error* at n. A reliable method commits at most a finite number of errors, although with a more restrictive criteria it may be required to commit none. Minimizing errors is one important criterion in choosing among alternative reliable methods.

A method δ can *change its mind*. If at a time $n + 1$ a method outputs a different conjecture then at n $\delta(\varepsilon|n) \neq \delta(\varepsilon|n + 1)$. The number of changes of

[14] To give justice to Snow one would need perhaps to bring in more distinctions. The method projecting that all who drink contaminated water contract cholera was used by Snow not for causal discovery, but causal assessment. He arrived at this hypothesis by examining pathogenesis of cholera and excluded the main alternatives being that cholera is contracted from the air or from inorganic poison; cf. (Snow 1936). Taking into account the onset of the

mind on the way to project the true hypothesis is also an important characterization of method and can be sought to be minimal.

1.3 Naturalism

Naturalism as understood here is the claim that all sciences without exception are subject to the norms of reliable inquiry. More specifically, if the afforded explanation is causal, then there are reliable algorithms for recovering the causal structure as the next subsection makes clear. The naturalist claim propounded on the grounds of the reliabilist understanding of science does not entail however that there is a universal method to carry out all studies come what may. In discussing the notion of reliability I have already noted a variety of ways to set up the criteria of success and I further bring this out in Subsection 3.2 below. The next subsection makes clear the difference in algorithms to handle causal queries differently depending on what kind of assumptions the researcher is ready to commit herself to.

Philosophy has no exemption from the naturalist claim. Therefore, when the challenge is to develop understanding of science, this task has to be accomplished by reliable methods. Elaborating the norms of reliability on the grounds of logic and descriptive set theory is the most perspicuous example of how to meet this demand. And so is the application of the logic of reliability to the study of algorithms recovering causal structure.

---

disease, its symptoms and organs affected he projected the nature of the agent causing the

Such a program for philosophy of science matches what has been demanded of it by the 20th century arch anti-naturalist Max Weber, who stands on the top of the German anti-naturalist 19th century tradition.[15]

Weber as well as other anti-naturalists characteristically emphasized that the subject matter in the social sciences and humanities requires a different set of methods than those of natural sciences. However, the application of the norms of reliable inquiry is not blind-folded, and neither is the application of algorithms searching for causal structure. Both are sensitive to subject-matter concerns and specific understanding thereof. And precisely as envisaged – and to some extent pioneered – by Weber[16] the reliabilist (or "means-ends") framework for causal explanation is in the first place elaborated for general case and can then be applied to account for individual phenomena or agents[17] based on the characteristic data related specifically to them and thus without imposing unrealistic constraints.[18]

1.4 Causality

disease as one of the then unobservable parasitic organisms.

[15] His standpoint invalidates or embraces the points addressed earlier by W. Dilthey, W. Windelband and H. Rickert.

[16] (Eliaeson 2003, 25-30; 50).

[17] How to conduct such an application within the reliabilist framework for causal inference is discussed in some more detail by Pearl (2000, 309-329).

[18] One of Weber's concerns is that scientist should not impose apriori rationality constraints upon the studies agents or phenomena. The reliabilist framework presented here avoids this difficulty in the same way that do so culturally constrained Weber's ideal-types. For reliabilist norms apply to the assumptions conceded by the scientist in question and the data collected by her.

As already noted in Subsection 1.1 above many of the best explanations of the data are causal. Causal explanations and models come in a variety however. Rather than going through a list of all words standing for "cause" it seems more convenient to describe causal inferences and causal models in science by what is characteristic of them. Whenever a given model is intended to afford not merely a prediction of the phenomena to be observed, but also how they will respond to intervention, i.e. the model itself is supposed not to be changed by the intervention, then one can admit that the model is constructed as causal and so is the corresponding inference of the effects of the intervention. Suppose for instance that the propounded model of the relation between smoking and lung cancer holds that there is a genetic factor which is their common cause and which makes these two independent. Even though this model could be used to predict changes in the rate of smoking and the number of people contracting lung cancer, it could be wrong. Why? Because if we were to increase the number of cigarettes smoked for a given group of people, we would observe a corresponding change in a purportedly independent number of people contracting the disease which is not consistent this model.[19] In contrast, a model that has a causal link between the two phenomena, allows to predict correctly the effect of the intervention.

In general, contemporary researchers working on the problem how to model causality and causal inference, seem to commonly accept as canonical a

[19] Intervention is not possible in this case on ethical grounds, but the corresponding data have been presented by the now classic studies on animals (Wynder and Croninger 1953; 1955) and human smoking discordant twins (Kaprio and Koskenvuo 1989).

graphical representation in form of directed acyclic[20] graphs. Such a graph is interpreted as causal if it satisfies so-called the Markov Assumption.[21] Informally, what the assumption amounts to is that for a set of variables V (with C, E and R being mutually exclusive subsets of V) describing given population N it holds that all causal influence variables in R could have on variables in E is without remainder mediated by variables in C – the direct causes of E. Thus, when we elaborate causal dependencies among variables in V we need to spell out (and statistically test for) direct causal dependencies and from these – given the graph – indirect causal dependencies will logically follow. By the Markov Assumption a causal graph precisely determines probabilistic independencies among the variables in V which could then be statistically tested against the data.

The reliabilist norms of scientific research, outlined in Subsection 1.2 above, allow to construct methods of reliable discovery of causal models from data given the background assumptions. This accomplishment of Clark Glymour, Peter Spirtes and Richard Scheines and their collaborators cannot be overestimated. The reliable methods of causal discovery turn on an assumption which turns out to be a converse of Markov Assumption and is called the Faithfulness Assumption.[22]

[20] Informally, an acyclic graph is a graph in which all links between nodes are directed (have pointed arrows at one end) and it is not possible to reach the same node one starts with by following the direction of the arrows.
[21] There is a number of ways how to spell out precisely the Markov Assumption (e.g. pairwise, local, global and factorization) both for directed and undirected graphs; see (Lauritzen 2001, 71-73; Spirtes, Scheines, Glymour, Richardson and Meek 2004, 455-56).
[22] This assumption is known also as stability (Pearl 2000), parsimony (Box and Jenkins 1976) or non-collapsibility of parameters (Lauritzen 2001).

In a nutshell, the Faithfulness Assumption states that statistically significant probabilistic independencies among variables in V measured in the data represent causal dependencies and are not merely coincidental. By this assumption the independencies do not turn on the exact numbers measured, but express a deeper – sometimes called "stable" or "structural" – relation which would not disappear had the variables measured happen to take on different values.[23]

There is a variety of algorithms that utilize the Faithfulness Assumption in recovering causal structure from data corresponding to differences in both background assumptions appropriate for a given domain and statistical techniques (frequentist and Bayesian).[24] The Fast Causal Inference (FCI in short) algorithm, for instance, does not presuppose that the variables V in a model at hand are causally sufficient. For any pair of variables A and B the FCI algorithm will detect if the data allow that there is a latent common cause, i.e. a variable not included in V that would make A and B independent.[25]

Some of the applications of reliable methods of causal discovery I highlight in Section 2 below.

Causal models with latent variables are prevalent in the social sciences. In response to computational problems in searching over directed acyclic graphs with latent variables a search over mixed ancestral graphs has

---

[23] As with the Markov Assumption, Faithfulness can be spelled out differently for different purposes; see (Spirtes, Scheines, Glymour, Richardson and Meek 2004, 457-59).
[24] The web site http://www.phil.cmu.edu/projects/tetrad/ links to the software and sample applications.

been proposed.[26] The latter graphs contain only observed variables and the search never introduces latent variables. At the same time it represents the same conditional independence relations among variables as the directed acyclic graph. The problem with searching over mixed ancestral graphs is that the graph found in search may represent a collection of causal graphs which predict different effects of interventions.

Causal models discussed in econometrics are typically constructed with the assumption of no latent variables. This, however, seems to be a natural extension of these models, but one which is susceptible to numerous problems. One of them is that in the case of systems with feedback, which is represented as a cycle in a causal graph, there is no straightforward extension of the Markov Assumption discussed in Subsection 1.4. The basic question here is how to uniquely determine a data generating process for a given directed cyclic model or an interesting class of such processes. It follows that there is no general theory of how to calculate the effects of interventions in the case of such graphs whose variables can take only a finite set of values.

Another challenge for the reliabilist approach to causal inference in nonexperimental research is brought out in (Robins et al. 2003). It is proved that in this context, when there may be unobserved or unrecorded common causes of recorded variables, there is no method of causal search that would be guaranteed to reliably approximate the correct (asymptotic) result on the basis of a finite sample size. Because of the unknown time order and the

[25] (Spirtes, Glymour and Scheines 2000, 138-47; Kłopotek and Wierzchoń 2002).
[26] Cf. (Spirtes, Glymour and Scheines 2000, section 12.5.7).

possibility of unobserved confounding the search for causal structure may be seriously misguided in observational studies. One suggested solution is to perform sensitivity analysis which demonstrates how the estimates of the causal structure would change with the amount of unobserved confounding. Another proposal is to continue to work with search methods which satisfy some weaker criteria, e.g. Bayesian consistency.

1.5 Computability

Science is a limited enterprise insofar as humans developing it have limited cognitive powers. If these powers or capacities are assumed to be computational, then what the limitations of scientific inquiry essentially amount to are limitations of computation.[27]

The study of reliable methods of inquiry provides a natural framework to study how this limitation affects scientific inquiry, yielding results which sometimes turn out to be quite surprising and almost always surpassing methodological rules of thumb. For instance, the results presented in (Kelly 1996) show that there are problems which cannot be solved by scientific methods which output only hypotheses consistent with the data observed so far and background assumptions (vs verificationism) or change hypotheses

[27] Since in this paper our focus is on understanding science, I refer the reader to the extensive discussion on the subject and in particular to the argument defending the computational theory of mind as presented in (Glymour 1992).

only when they face contradiction with data (vs falsificationism). In Section 3 below I spell out the rationale for these discrepancies somewhat more fully.

1.6 History of ideas

The aim of this subsection is twofold. Surely, I point out the entrenchment of the basic ideas underlying reliabilist understanding of science. By the same token, however, I demonstrate that this conception of science takes seriously the conviction – propounded forcefully since the 1960's – that the historical development of science constitutes the background against which to measure the adequacy of any account of science.

In the history of reliabilism preceding recent developments starting with H. Reichenbach's and H. Putnam's works K. Kelly highlights two key moments, i.e. emergence of Plato's reliabilist conception of knowledge and Peirce's dispensing with the requirement of certainty:

Plato seems to assume that inquiry must be logically guaranteed to terminate with certainty that its answer is correct. Similar requirements were common among ancient skeptics such as Sextus Empiricus. This very strict version of logical reliabilism is still evident in the work of Descartes, Hume, and Kant, over two thousand years later. Subsequently, it was proposed by C. S. Peirce that inquiry might converge to the truth without providing a clear sign that is has done so. (Kelly 1996, 3) Ancient Greeks brought out two fundamentally different ways how to conduct scientific inquiry, each of which can be epitomized by a distinguished inquirer: Euclid and Socrates. The former envisioned science as proceeding top-down by spelling out the most obvious and fundamental assumptions as axioms and then deducing the remaining knowledge from them. The latter conceived of science as a bottom-up enterprise where one collects positive and negative examples and on this basis spells out the underlying intuitive idea in a more and more general form which would include all positive and no negative examples.

Which of these approaches is to be chosen in developing a conception of science in large measure turns on the assumed notion of knowledge. Plato combined Socratic strategy with recollection: at some point examination of a finite amount of evidence prompts the inquirer to recognize a conclusion which she already knew. The account of recollection as the warrant of truth of the conclusion at hand seems, however, susceptible to cogent skeptical objections. No finite evidence warrants a general conclusion (for instance even after enormously large amount of cases observed where smoking causes lung cancer it is remains *logically* possible that there are cases where it does not) and it is impossible to scrutinize all possible cases.

If what one seeks is however the *logical* warranty of arriving at truth, one needs to turn to the top-down strategy instead. Given the background assumptions which constrain the number of possibilities to be considered, one can investigate whether there is a reliable method for solving the problem at hand in the manner described in Subsection 1.2 above. This has been clearly conceived by many philosophers and scientists alike, but most perspicuously so – and perhaps for the first time – in the 13th century by Ramon Lull. This Franciscan monk's idea was

> that Moslems (and others) may fail to convert to Christianity because of a cognitive defect. They simply were unable to appreciate the vast array of the combinations of God's or Christ's virtues. Lull believed that infidels could be converted if they could be brought to see the *combinations* of God's attributes. Further, he thought that a *representation* of those combinations could be effectively presented by means of appropriate machines, and that was the key to his new method. Lull designed and built series of machines meant to be used to present the combinations of God's virtues. (Glymour 1992, 70)

An important step towards elaboration of the reliabilist understanding of science was accomplished by Leibniz in the 17th century. He was convinced that it is possible to extract the complete alphabet of primitive notions and thereupon mechanically develop the rest of human knowledge.[28] What was important, however, was the idea that it is possible to study deductive inference by means of algebraic methods applied to abstract symbols representing propositions.

It is George Boole, however, who turned Leibniz's idea into a theory of inference or "algebra of thought". He claimed that each discourse determines a *domain* which is structured in a way that can be represented as operations on more simple sets of objects. Boole introduced variables to

[28] Having assigned symbols to primitive and complex notions, one can apply algebraic methods to seek for identities among those symbols (Glymour 1992, 86).

represent those simple sets, 1 to represent the whole domain and 0 its complement and a couple of elementary operations on the sets corresponding to algebraic operations on numbers, i.e. addition, multiplication and complement (corresponding to the negative sign "–"). The most important result was a set of ten axioms which constitute the "laws of thought" true in every domain and by the same token in every discourse.[29] What matters for the elaboration of the reliabilist conception of science is that there is a mechanical procedure (algorithm) to decide whether a given set of Boolean sentences (premises) entails another (conclusion).

Gottlob Frege, although unaware of Boole's accomplishments, turned logic into its modern form. Frege's guiding idea that all mathematics can be reduced to logic failed. Nonetheless, his program allowed the extention of the domain of application of logic to cover – for the first time since its inception by Aristotle – mathematical reasoning. In particular, Frege afforded a characterization of the notion of proof adequate for mathematical proofs. That characterizations led to the discovery that in Frege's logic there are true formulas which cannot be mechanically proved from the axioms.

These and some others (most notably probabilistic) historical accomplishments have been integrated into the first contemporary attempt to elaborate a reliabilist conception of science, which is due to H. Reichenbach.[30] He conceived of science as aiming at a reliable discovery of relative frequencies: we first observe how many times a given type of event occurs

[29] A thorough exposition is in (Glymour 1992, 95-114).
[30] (Reichenbach 1949).

among other events and then project this ratio as the limit in the long run (straight rule). If there is a limit of relative frequencies, then the straight rule – as Reichenbach claimed – will lead us to the truth. The straight rule, on this account, is then the reliable method of science.

Reichenbach's student H. Putnam generalized the frequency analysis of reliability in the 1960's.[31] Roughly, the setup is the following. Suppose that we investigate the truth of a given hypothesis (e.g. "Smoking cigarettes causes lung cancer") and after each new piece of evidence we output a claim whether the hypothesis is true or not. To decide it on the available (finite) evidence we employ a rule, call it "R" . Putnam's proposal can then be summarized as a criterion for reliability of R as follows (Glymour 1992, 262):

> In every logically possible world *W* […] and for every possible order of presentation to the investigator of the individual facts in *W*, there exists some finite number of facts after which *R* outputs only T [for true] if the hypothesis is true and outputs only F [for false] if the hypothesis is false.

What is characteristic of a reliable rule or method is that it makes at most a finite number of errors, gets the right answer and does not change its mind afterwards. However, it is not required that the rule or method signals when the right answer has been reached.[32]

---

[31] Cf. (Putnam 1956; 1963; 1965). Simultaneously, the idea of reliability was being developed by M. Gold (1965; 1967), R. Solomonoff (1964a; 1964b) and others (Blum and Blum 1975; Angluin 1980) in the context of language learning. The question there was: how to account for the fact that a child effectively learns a language in a relatively short amount of time.

[32] A comprehensive exposition of the results going beyond Putnam's is in (Kelly 1996) and (Jain et al. 1999).

2. ***Applications***

It is well beyond the limits of the present paper to provide even an outline of applications of the reliabilist conception of science. The illustrations to follow should rather be conceived of as points of reference for further studies. The program TETRAD is itself a tool that already generated voluminous literature and applications. The examples which do not explicitly cite algorithms in TETRAD could well be reformulated to do so.

2.1 TETRAD

The ideas presented in Section 1 are integrated in a computer program called TETRAD. It embodies both the Markov and Faithfulness Assumptions which jointly allow for search of the true causal structure represented as a directed acyclic graph. The input to algorithms in the program consists of data (cell counts or correlation matrix) and background knowledge which rules out some of the logically possible graphs. If there is a unique causal graph that can be reliably identified on the basis of data and background knowledge, TETRAD finds it and estimates it.[33] Otherwise, the program outputs a class of graphs that cannot be distinguished on the basis of the available data.

Here I will outline only one application.[34] P. Blau and O. Duncan (1967) in a now classic study examined the American occupational structure. On the basis of a sample of more than 20 000 they concluded that individual's occupation is directly determined by one's education, first job and father's occupation, and only indirectly by father's education. TETRAD discovered the same dependencies on the basis of the data and the background assumptions concerning the time order of variables. It supplemented the original model of Blau and Duncan with causal dependence between father's education and one's first job. Hence, the obtained graph is almost complete (i.e. except for one edge every variable is linked to all other variables).

The theory underlying the program leads to an intriguing explanation of this model. When the model is complete, one may well suspect that the correlations in the data between variables do not result from genuine causal dependencies between them. Rather, they result from a faulty procedure in collecting the data. Blau and Duncan implicitly assumed that the causal dependencies for all the variables in the model are the same for all individuals considered. The result produced by TETRAD indicates that this is not the case and the sample data do in fact come from different and mixed subpopulations with different causal factors determining various individual's occupations.

---

[33] In (Spirtes, Scheines, Glymour, Richardson and Meek 2004) the estimation of causal models is introduced analogously to the estimation of statistical models.

[34] For numerous other applications of the program the Reader is referred to (Cooper and Glymour 1999; Spirtes, Glymour and Scheines 2000).

2.2 Social psychology

The reliabilist criteria for causal discovery and prediction clearly mark causal dependencies from statistical ones. The distinction, apparent as it seems, is often confused in modeling techniques used by both students and researchers. The confusion inherent in many standard research methods is sometimes remedied by the wit of their users. And thus only occasionally comes out straightforwardly. A perspicuous example is the infamous book by C. Murray and R. Herrnstein *The Bell Curve* (1996). The conclusion of the book, which was unacceptable to public opinion in North America, was that the social stratification of the American society reflects differences in individual cognitive abilities and thus – given the almost equal educational opportunities – is at the bottom biologically grounded. The ethnic differences purportedly follow with Afro-Americans being the most inferior part of the society.[35]

In arriving at the unwelcome conclusion Murray and Herrnstein use the techniques for data analysis which are standard throughout the social sciences: multiple regression, logistic regression and factor analysis. For instance, the same techniques have been used in the study by Blau and Duncan referred to in the previous subsection.

J. Gould in dismissing this result focused on there being one feature – general intelligence or cognitive ability – which is purportedly measured by

IQ tests. However, even if IQ does measure a bundle of hereditary features,[36] the weight of the argument in *The Bell Curve* is on the claim that the feature or features that IQ tests measure, causes directly many of the social outcomes indicated by wealth, occupation, martial status, illegitimate children, parenting style, dependence on welfare, crime and political involvement. And this Murray and Herrnstein established by multiple regression.

To establish whether behavior X of an individual is influenced by her cognitive capacity by multiple regression the following must hold (Glymour 2001, 198):

> *cognitive ability does not (directly) influence* X *if and only if cognitive ability and* X *are independent conditional on the set of all the other regressors*.

For this to hold the researcher would have to know – independently of the data analysis – that no regressor has a common cause with the outcome variable and that the latter is not a cause of any of the former. For regression evaluates influence of a regressor on X conditional on *all* other regressors and not on any subset thereof.

In particular, when the Markov Assumption is applied to the model proposed by Murray and Herrnstein[37] it turns out that without imposing causal assumptions preceding the data analysis it is not possible to rule out the possibility that cognitive ability and X are causally independent conditional on

---

[35] Cf. (Herrnstein and Murray 1996, 269).
[36] For a detailed criticism of how Herrnstein and Murray misuse factor analysis in establishing IQ as a purported measure of single factor, i.e. cognitive capacity, see (Glymour 2001, 177-89).

a common cause. The common cause could well turn out to be an environmental factor, thus undermining the substance of *The Bell Curve*.

2.3 Cognitive and developmental psychology

Recent experimental studies of the learning mechanisms in children have been largely informed by the reliabilist conception of causal discovery. The paper by A. Gopnik and Glymour et al. (2001) has triggered a series of research papers which provide ample evidence to the fact that children use specialized cognitive systems in order to recover an accurate "causal map of the world".[38] The mathematical model of these systems is as outlined in Subsection 1.4 above and as applied in TETRAD algorithms.

In a series of experiments with young children Gopnik and her collaborators[39] obtained results which cannot be accounted for by the non-causal models of learning, especially the Rescorla-Wagner associative model and the trial-and-error model. Many of these experiments use a specially devised machine, called the blicket detector,[40] which lights up and plays sounds when blocks are place upon it. Preschoool children – as young as three years of age – recognized causal patterns from the data (e.g. common effects and common causes) and made correct predictions on how to activate the

[37] See (Glymour 2001, 200) for details.
[38] For a detailed discussion of causal maps see (Gopnik et al. 2004).
[39] Cf. (Gopnik and Sobel 2000; Gopnik et al. 2001; Gopnik et al. 2004; Lagnado and Sloman 2004; Mochon and Sloman 2004; Schultz and Gopnik 2004; Sobel et al. 2004; Sloman and Lagnado 2005).

machine. These predictions included also effects of unobserved interventions, which were not imitations of the previously observed actions. Also, in the case of experiments with the puppet machine they correctly posited an unobserved common cause (Gopnik et al. 2004, 28-29). Similar results have been observed for children's causal inference in different domains, e.g. psychological, biological and physical (Schultz and Gopnik 2004).

Some experiments reveal a difference between three and four year old children. But as Sobel et al. (2004) emphasizes this is not due to different causal mechanisms being employed at different ages. The differences rather stem from the fact that the application of causal maps requires certain information-processing capacities, like keeping track of different hypothesis or using prior knowledge to limit the space of possible causal explanations. The enhancement of these capacities with age enables children to be more successful in accurate causal inference and intervention.

2.4 Econometrics

A spectacular, although somewhat restricted, application of the reliabilist ideas to scientific research as discussed in Section 1 and Subsection 2.1 above, is in (Swanson and Granger 1997) – the latter co-author being the Nobel Prize winner in economics. The major restriction in this work is trading the Faithfulness Assumption for some pieces of background knowledge, which

[40] Cf. (Gopnik and Sobel 2000).

allow to restore causal dependencies among variables considered, especially the assumption of chain ordering of variables in the model. In the ramifications of the present paper I only give intuitive idea of this application and invite the Reader to study the literature.[41]

C. Sims in the 1980's raised influential criticism of the standard structural modeling techniques as developed by the Cowles Commission in the 1940's, especially by T. Haavelmo (1944). The idea underlying structural models was that as the factors affecting a given variable they list on the right hand side of the equation the substantive variables plus a structural error term summarizing the influence of all the factors omitted in the model. The structural character of the model represented the fact that the equations would not change with an intervention on one of the variables. The point of Sims criticism was that to make such models identifiable from the data, economists employed highly dubious prior "knowledge". In particular, the structural errors in these models could be directly measured. The alternative auto-regression (VAR in short) models Sims proposed, contain all the substantive variables[42] and prediction errors, i.e. the measured deviance of the predictions of the models and subsequently observed data. The VAR models, and the prediction errors in particular – although identifiable – do not bear the warranty of invariance to intervention, i.e. in general these are not causal.

[41] (Swanson and Granger 1997; Hoover 2001; Demiralp 2000; Bessler and Lee 2002; Moneta 2003; Demiralp and Hoover 2004; Hoover 2005; Phillips 2005).

[42] What Sims proposed amounts to regressing a given variable on all the substantive variables in the past and the only limitation on how far we need to reach in the past is how the statistically given variance of estimators would go up.

The standard technique to convert the VAR model to a structural model is the Cholesky decomposition.[43] The problem, however, addressed by Sims, returns in a disguise. For typically there is more than one Cholesky decomposition for a given VAR model and in consequence, multiple possibilities how to spell out the true causal (structural model). What Swanson and Granger propose in the cited paper essentially amounts to assuming that there is a chain structure among the variables and this background assumption allows to infer testable constraints on partial correlations observed in the data. Given the assumption the recovery of the causal model from a VAR model is driven by the evidence.

The work by the research team at the CMU affords several natural ways to generalize the result by Swanson and Granger, as they themselves note (Swanson and Granger 1997, 357). One point to note is that the standard TETRAD algorithms can be applied to recover the causal model from the data without arbitrary assumptions on their causal ordering.[44] Further, the reliabilist approach allows one to search for structural models with latent variables, i.e. allows one to discuss the case omitted in the econometric literature when the variables in the model have an unobserved common cause.

[43] For details see (Enders 1989, 307-10).

[44] This strategy has been recently applied to VAR models by numerous authors, cf. (Demiralp 2000; Bessler and Lee 2002; Moneta 2003; Demiralp and Hoover 2004; Hoover 2005; Phillips 2005).

2.5 Chaos

Although in the present paper the focus is on the social sciences, the reliabilist conception of science – as spelled out in Section 1 – applies to natural sciences as well.[45] A detailed illustration I resume here is co-authored by M. Harrell and Glymour (2002), where the principles of reliable discovery are applied to the theory of chaos.

Admittedly, there is more than one definition of chaos, but it is generally assumed that a characteristic of chaos should include "sensitive dependence on initial conditions" (SDIC in short). And SDIC is usually defined by means of the Lyapunov exponent[46] of the chaotic system: the greater the exponent, the greater is SDIC (zero or negative indicate no dependence on initial conditions).

In accordance with the reliabilist criteria that the Lyapunov exponent is positive can only be learned from time series data if there is some point in data collection after which the Lyapunov exponent is projected always as positive. That this can indeed be learned from the data if the exponent is positive is proved in (Harrell and Glymour 2002). If the Lyapunov exponent is

[45] Another detailed example I would recommend for the interested reader is O. Schulte's application of reliabilist criteria to particle physics (Schulte 2000).

not positive, it can also be learned, although in some weaker sense of converging to the right answer.

3. Alternatives to reliabilist epistemology

Although the reliabilist ideas are – and always have been – ubiquitous in science, few contemporary philosophers of science recognize this fact. There are numerous ways to account for this failure. One cogent reason is the historical development of the discipline after Kant. Following Aristotle, philosophers preceding Kant adopted what I characterized in Section 1 as naturalism, i.e. continuity between philosophy and science. For Kant, however, philosophy was not to investigate the phenomena of the world around us, but rather to justify the ultimately objective phenomenon, i.e. science itself. "How science is possible?" became then the question of prime importance and replaced the preceding interest in the world itself. Through the influential school of neo-Kantians in Marburg, this view had been adopted by the Vienna circle[47] which set it up as the standard for the philosophy of science of the 20th century. The subsequent search for "the logical reconstruction" of science or its "logic" epitomize it well.

[46] The Lyapunov exponent is given by: $\lambda_n = \frac{1}{n}\sum_{i=0}^{n-1} \ln\left|f'(x_i, \delta_i(n)\right|$, where x_i is the ith iterated value of the function f, f is the first derivative at x_i, λ_n is the Lyapunov exponent at x_i and n is the number of data points.

[47] Cf. (Jeffrey 1973; Richardson 1998).

In the remainder of this section I shortly outline how the advantages of the reliabilist conception of science with regard to its contemporary alternatives.

3.1. Bayesianism

Bayesianism as the conception of science brings to the fore its dynamics.[48] On this view the strength of scientists' belief in the truth of a hypothesis is represented as a degree of probability of this hypothesis. If two scientists differ with regard to their prior belief in the hypothesis, but they adopt the same – Bayesian[49] – mechanism for updating their believes each time new empirical evidence is obtained, then – it is proved that – they both converge to the true hypothesis.

For the mechanism to work one needs to assume that maximum strength of belief is given exclusively to logical truths (probability 1) and minimal (probability 0) to logical contradictions; all other propositions are believed without being certainly false or true.

---

[48] (Jeffrey 1973; Horwich 1982; Howson and Urbach 1993). For a discussion of Bayesianism as contrasted with reliabilism see (Earman 1992).

[49] The central idea is expressed by the Bayes theorem: $\mathrm{P}(H|E) = \mathrm{P}(H)\cdot \frac{\mathrm{P}(E|H)}{\mathrm{P}(E)}$, where H and E stand for hypothesis and evidence, $\mathrm{P}(H|E)$ is the strength of belief in H updated after new evidence E has been acquired, $\mathrm{P}(H)$ is the strength of belief prior to acquiring the evidence, $\mathrm{P}(E|H)$ is the strength of belief in E given H is true, and $\mathrm{P}(E)$ is the strength of belief in E irrespective of whether what is true is H or some other alternative hypothesis. Bayesians characteristically extend the scope of the Bayes theorem of the standard theory of probability to apply not only to statistical inference, but to all nondeductive inference in science. The standard Dutch book argument to support it is not conclusive (Glymour 1980, 71-73; Maher 1997).

The Bayesian conception of science, however, does not do justice to the history of science where "explicitly, probability is a distinctly minor note" in how the arguments were presented (Glymour 1980, 66). Moreover, it faces - sometimes insurmountable – difficulties in explicating the notions relevant to science, e.g. simplicity as a criterion for preference of one theory to another.[50]

Another problem is identified by C. Glymour (1980, 86-92) as the problem of old evidence.[51] It is a common expectation in science that the new theory not only correctly predicts the upcoming evidence, but is also consistent with the known evidence, collected in order to evaluate the preceding theories. This case apparently distorts the dynamics of the Bayesian account of scientific inference for – as the Bayesian mechanism has it – the strength of the updated belief cannot differ from the initial one.[52]

It turns out that computational limitations of scientists as cognitive agents bear importantly on the viability of Bayesianism. One instance is the class of inductive problems that are solvable by Bayesian learners, but not solvable if we impose upon them the computational limitation.[53] Another instance is the class of inductive problems solvable by a non-Bayesian computable learner, but not solvable by a consistent computable Bayesian learner.[54]

---

[50] Cf. (Forster 1995).
[51] See also (Leamer 1978). For an extensive discussion see (Earman 1983).
[52] From the fact that the evidence is known it follows that its probability is 1, and so is its likelihood on *H*. Hence, from the Bayes theorem we obtain that $P(H|E) = P(H) \cdot 1$.
[53] (Osherson, Stob and Weinstein 1988).
[54] (Juhl 1994).

Finally, for the Bayesian solution to be computable one needs to compromise with the skeptical challenge. As some of the logically possible alternative hypotheses are assigned probability 0 some of the possible worlds are eliminated at the outset. Sure, there can no longer be a logical warranty that the actual world we are in is not one of those excluded. This is perhaps the most perspicuous difference with the reliabilist conception of science which affords us the criteria of logical warranty that a reliable method arrives at the truth irrespective of the skeptical worry.

3.2. Falsificationism

Aware of the skeptical scenario, K. Popper propounded an alternative understanding of the dynamics of science.[55] Scientists are encouraged to come up with bold conjectures, which are then subject to empirical tests. If they fail, they are conclusively falsified – they logically contradict the evidence. If they stand the test, they are entertained only tentatively, in view of a possible falsification by some further empirical test.

In reliablist terms what Popper requires a criterion of success somewhat weaker than those required by the skeptic. For on this account scientific method is required to project a given hypothesis until there is a point in the data stream such that the data contradicts it. Popper requires scientific

[55] (Popper 1935; 1963).

method to be reliable only in the following sense: any hypothesis it has to be refutable with certainty, i.e. after the method outputs 0, it stops.

Falsificationism leaves out what appears to be at least equally important part of scientific enterprise: reliable learning of truth, i.e. the success in verification of a hypothesis. The opposite of Popper's criterion of success is verification of (existential) hypothesis with certainty when a method outputs 1 and stops.

Moreover, as suggested cogently by Reichenbach and elaborated by Putnam, yet weaker criteria of success are possible too. In general, what can be relaxed is the requirement that apart from hitting upon the truth scientific method also *signals* that it has reached it. In this sense one can speak of convergence to the truth in the limit for both verification and refutation.

It follows, as demonstrated in (Kelly 1996, 95), that sticking to Popper's criterion sometimes would even compromise reliability, for instance, in the case when the hypothesis in question has a logical warranty of verification in the limit.

3.3 Constructive Empiricism

Constructive empiricism[56] is a programmatically epistemological version of anti-realism discussed in Subsection 1.1. It sets empirical adequacy as the aim of science and claims further that the truth of what goes beyond the observable is immaterial. It is thus supposed to undermine any existential claim on the part of science that would not be credited by an unaided observation.

In Subsection 1.1 I noted what may be recognized as the fundamental difficulty of constructive empiricism. It appears to introduce an arbitrary epistemic divide between what is observable and nonobservable, trading thus for what is crucial in scientific enterprise, namely providing the best available explanation of the data. For scientific realist whether the best explanation would draw upon observable or nonobservable entities is of secondary importance as compared with getting *the best* explanation of the data.

In the reliabilist framework, especially in the classification of reliabilist success criteria,[57] there is more to be said about constructive empiricism. As other empiricisms, it seems to be fixed at one level of such criteria for all kinds of scientific inquiries but it somewhat relaxes Popper's requirement that the hypothesis in question be decidable (refutable) with certainty. As Kelly points out, the best way to think of constructive empiricist is that only those hypotheses are worthy of belief which are at least decidable (refutable or verifiable) in the limit.[58] For empiricist is concerned only with

[56] (van Fraassen 1980; 1989; 2002).
[57] (Kelly 1996, 115-117).
[58] If we take into account the computational limitations of scientists then this criterion of success has to be weakened further (Kelly 1996, ch. 7).

empirical adequacy, i.e. correctness of a given hypothesis with regard to the data stream alone rather than – as realist – with regard to the underlying reality.[59]

3.4. Relativism

Some philosophers or sociologists of science would have it that scientific research is not about getting to the truth, but about reaching a common consensus on what is acceptable for proponents of alternative theories.[60] The purported rationale for this view is that each alternative theory sets up its own conceptual apparatus and truth is *relative* to it. If a scientist were to switch to some other theory, she would no longer refer to the same set of truths. Therefore, when scientists reach a consensus it cannot be based on a common truth as there is none and what is decisive is an exercise of political power, personal influence etc.

Although relativism is turned down on scientific realist standpoint, nonetheless reliabilist conception of science still applies.[61] Of course, truth cannot be the basis for the consensus among scientists of different persuasions, but within a given conceptual scheme the questions of reliably getting to truth – the truth relative to a given conceptual scheme – still arise.

[59] The though skeptical problem for realist, that empiricist dispensis with, is that the underlying reality may be different and still the data streams obtained the same.
[60] Cf. (Barnes, Bloor and Henry 1996; Kukla 2000).

One important question concerns scientific revolutions: can we find a conceptual scheme in which the truth value of a given hypothesis[62] is reliably decidable? Another question is: can we reach a point for a given hypothesis after which changes of conceptual schemes do not affect correctness of conjectures of its truth value?

To achieve success on any of the above or similar questions the inquirer, as proved by (Kelly and Glymour 1992) has to follow a strategy balancing between changes of one's conceptual scheme and following the scheme in collecting further data. Still, "for each conception of convergence there is a universal learner that will solve any problem solvable by any learner" (Glymour 1996, 284).

4. Open questions and future research

Let me close this outline of the reliabilist conception of science and its implications for scientific research with the following remark. Within relatively short time[63] the reliabilist conception of science stimulated a vast amount of results on causal inference and modeling which set out a methodology for the social sciences that is fundamentally uniform with the methodology of the natural sciences and which has already proved fruitful in

[61] In other words, on the reliabilist conception of science some of the problems posed by T. Kuhn (1977; 1996; 2000) turn out to be perfectly legitimate, especially the two mentioned below in the main text.
[62] On how to construe it with translatability see (Kelly Glymour 1992).

numerous studies. This impact on science has no counterpart in any of the alternative contemporary conceptions of science. In short, the reliabilist conception of science wins on every single criterion offered: integrity, coherence, precision, response to skeptical challenges, compliance with the history of science and continuity with science.

As it appears, both philosophers and scientists are challenged by the reliabilist conception of science. For there is a number of issues that on the reliablist perspective would require a thoroughgoing elucidation. One issue is the prevalence of unreliable methods of inquiry in science. How to go beyond a mere attribution of irrationality in evaluation is well illustrated by A. Woody (1997). She studied – the history of diagrams representing molecular structure for covalent bonding introduced by G. N. Lewis in 1916. These diagrams work only for relatively light atoms and are not approximations of general theories of atomic structure, especially quantum theory. Woody explains their prevalence in chemistry textbooks and professional literature as “a useful shorthand for certain limited types of information”. The fact that they turn out to be similar to graphical representation of discrete data sets and do not require complex skills makes them an easy instructional tool historically connected to “earlier episodes of chemical practice” (1997, 59). There seem to be plenty of cases like Lewis’s diagrams to be elucidated, e.g. the unrestricted use of regression in causal inference. One can go even further and inquire

[63] With the opening date somewhat arbitrarily assigned to the publication of (Kiiveri and Speed 1982).

about the general underlying mechanism in the functioning of science that promotes the sustenance of unreliable methods of research.

Another issue is the link between reliability and explanation. The voluminous philosophical and methodological literature on explanation omits the important question: How the explanatory virtues of a theory contribute to its being more reliable in getting to the truth? A "rational reconstruction" of explanation and discussion of its nature have little to contribute to this question. Instead, one would expect that an investigation into historical examples, e.g. Kepler's preference of Copernican vs. Ptolemaic theory, may yield a clue to answer the question.

Seemingly endless philosophical discussions of token causation[64] are yet another example where reliabilist perspective would provide with criteria to resolve many issues. It is not only that the norms of reliability form a firmer ground than individual's intuitions about particular cases. They also could help philosophers and scientists to bring about methodological standards for causal discovery of token causes. With such a tool at hand the automation revolution would exercise an immense influence on the public life by design of individual-oriented health care system or more reliable judicial recognition of what are an individual's responsibilities in a particular case.

Finally, there is a set of problems that once used to be in the focus of philosophers and now has been almost entirely shifted towards computer

[64] Token causation applies to cases when one is interested in spelling out causes of a particular event rather than – as is with type causation – spelling out causal dependencies between classes of events, e.g. smoking and lung canser (which are standardly represented by random variables).

scientists, namely scientific discovery. Surely, the language for discussing this topic today has been shifted towards sophisticated mathematics, especially with the advent of research on machine learning in the 1970's and 1980's. Nevertheless, there is ample room for philosopher's ideas to contribute in solving such challenging problems as modeling complex systems in terms of their parts, e.g. the Earth ecosystem or models of gene regulation. Out of five prevailing issues P. Langley (2002) identifies as especially important three have always been of major concern to philosophers: how to develop a system of representation that will ensure commensurability between groups of researchers using different conceptual schemes and systems of representation, how to incorporate background knowledge and revision into the process of learning and how to attain explanations of the data rather than a mere description.

As it seems, we are now facing the dawn of the automation revolution in science. The moral of this paper is that how far this revolution will take us on the road to the truth turns on how deeply it will be framed by the ideas of reliability.[65]

[65] The survey on reliabilism is largerly stimulated by writings and teachings of Clark Glymour to whom I express my great debt for numerous discussions and illuminating comments to earlier drafts of this paper. I am grateful to the Foundation for Polish Science for sponsoring my research at the Center for Philosophy of Science at the University of Pittsburgh where the text has been composed.

Paweł Kawalec

University of Pittsburgh /

Catholic University of Lublin

References

Angluin D., 1980, Inductive inference of formal languages from positive data, *Information and Control,* 45, p. 117-35

Astuti R., Solomon G., Carey S., 2004, Constraints on conceptual development: a case study of the acquisition of folkbiological and folksociological knowledge in Madagascar, *Monographs of the Society for Research in Child Development,* 69, p. vii-165

Barnes B., Bloor D., Henry J., 1996, *Scientific knowledge: a sociological analysis,* London: Athlone

Blum L., Blum M., 1975, Toward a mathematical theory of inductive inference, *Information and Control,* 28, p. 125-55

Box G., Jenkins G., 1976, *Time series analysis, forecasting, and control,* San Francisco: Holden Day

Bronk A., 2001, Stanisław Kamiński – a philosopher and historian of science, Krajewski W. ed., Polish philosophers of science and nature in the 20th century, Amsterdam: Rodopi, p. 141-51

Christensen D., 1996, Dutch-Book arguments depragmatized: epistemic consistency for partial believers, *The Journal of Philosophy,* 93, p. 450-79

Doll R., 2002, Proof of causality, *Perspectives in Biology and Medicine,* 45, p. 499-515

Doll R., Hill A. B., 1952, A study of the aetiology of carcinoma of the lung, *British Journal of Medicine,* 4797, p. 1271-86

Doll R., Hill A. B., 1950, Smoking and carcinoma of the lung, *British Medical Journal,* 4682, p. 739-48

Doll R., Hill A. B., 1954, The mortality of doctors in relation to their smoking habits, *British Medical Journal,* 4877, p. 1451-55

Doll R., Hill A. B., 1956, Lung cancer and other causes of death in relation to smoking, *British Medical Journal,* 5001, p. 1071-81

Doll R., Hill A. B., 1956, Lung cancer and tobacco, *British Medical Journal,* 4976, p. 1160-63

Doll R., Peto R., Boreham J., Sutherland I., 2004, Mortality in relation to smoking: 50 years' observations on male British doctors, *British Medical Journal,* 7455, p. 1519

Earman J., 1992 wyd. 2, *Bayes or bust? A critical examination of Bayesian confirmation theory,* Cambridge, MA.: MIT

Eliaeson S., 2003, *M. Weber's methodologies: interpretation and critique,* Cambridge: Polity

Enders W., 1989, *Applied econometric time series,* New York: Wiley

Fisher R. A., 1959, *Smoking: the cancer controversy. Some attempts to assess the evidence,* Edinbrugh: Oliver and Boyd

Forbes H., Tufte E., 1968, A note of caution in causal modelling, *American Political Science Review,* 62, p. 1258-1264

Forster M. R., 1995, Bayes and bust: simplicity as a problem for a probabilist's approach to confirmation, *British Journal for the Philosophy of Science,* 46, p. 399-424

Frost W. H., 1936, An introduction, Snow J., *Snow on cholera,* New York: The Commonwealth Fund, Londyn: Humphrey Milford, Oxford University Press, p. ix-xxi

Gelman S., 2004, Psychological essentialism in children, *Trends in Cognitive Sciences,* 8, s. 404-409

Glymour C., 1980, *Theory and evidence,* Princeton: Princenton University Press

Glymour C., 1985, Explanation and realism, Hooker C. A. (ed.), *Images of science*, Chicago: Chicago University Press, Churchland P., p. 99-117

Glymour C., 1992, *Thinking things through: an introduction to philosophical issues and achievements,* A Bradford Book/Cambridge, MA.: MIT

Glymour C., 1994, On the methods of cognitive neuropsychology, *British Journal for the Philosophy of Science,* 45, p. 815-35

Glymour C., 2001, *The mind's arrows: Bayes nets and graphical causal models in psychology,* A Bradford Book/Cambridge, MA.: MIT

Glymour C., 2004, We believe in freedom of the will so that we can learn, *Behavioral and Brain Sciences,* 27, p. 661-62

Glymour C., Cooper G. F. (ed.), 1999, *Computation, causation, and discovery,* Cambridge: MIT

Glymour C., Kelly K. T., 2003, Why Bayesian confirmation does not capture the logic of scientific justification, Dept. of Philosophy, CMU, Technical Report 138

Glymour C., Kelly K. T., 2004, Why probability does not capture the logic of scientific justification, Hitchcock C. (ed.), *Contemporary debates in the philosophy of science,* Oxford: Blackwell, p. 94-114

Glymour C., Scheines R., Spirtes P., Kelly K., 1987, *Discovering causal structure: artificial intelligence, philosophy of science, and statistical modeling,* New York: Academic Press

Glymour C., Spirtes P., 1988, Latent variables, causal models and overidentifying constraints, *Journal of Econometrics,* 39, p. 175-98

Gold M., 1965, Limiting recursion, *Journal of Symbolic Logic,* 30, p. 27-48

Gold M., 1967, Language identification in the limit, *Information and Control,* 10, p. 447-74

Gopnik A., Glymour C., Sobel D., Schulz L., Kushnir T., Danks D., 2004, A theory of causal learning in children: causal maps and Bayes nets, *Psychological Review,* 111, s. 3-32

Gopnik A., Schulz L., 2004, Mechanisms of theory formation in young children, *Trends in Cognitive Sciences,* 8, p. 371-77

Gopnik A., Sobel D., Schulz L., Glymour C., 2001, Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation, *Developmental Psychology,* 37, p. 620-29

Gould S. J., 1995, Mismeasure by any measure, New York : Times Books, Jacoby R., Glauberman N. (ed.), *The bell curve debate: history, documents, opinions ,* p. 3-13

Haavelmo T., 1944, The probability approach in econometrics, *Econometrica,* 12, suppl. p. 1-115

Harrell M., Glymour C., 2002, Confirmation and chaos, *Philosophy of Science,* 69, p. 256-65

Herrnstein R., Murray C., 1996, *The bell curve: intelligence and class structure in American life,* New York: Simon & Schuster

Hoover K. D., 2001, *Causality in macroeconomics,* Cambridge: Cambridge University Press

Horwich P., 1982, *Probability and evidence,* Cambridge: Cambridge University Press

Howson C., Urbach P., 1993 2nd ed., *Scientific reasoning: the Bayesian approach,* Illinois: Open Court

Jacoby R., Glauberman N. (ed.), 1995, *The bell curve debate: history, documents, opinions,* New York : Times Books

Jain S., Osherson D., Royer J., Sharma A., 1999, *Systems that learn: an introduction to learning theory,* A Bradford Book/Cambridge, MA.: MIT

Jeffrey R. C., 1973, Carnap's inductive logic, *Synthése,* 25, p. 299-306

Juhl C., 1993, Bayesianism and reliable scientific inquiry, *Philosophy of Science,* 60, p. 302-19

Juhl C., 1994, The speed-optimality of Reichenbach's straight rule of induction, *British Journal for the Philosophy of Science,* 45, p. 857-63

Juhl C., Kelly K., 1994, Realism, convergence, and additivity, *Philosophy of Science, Proceedings,* 61, p. 181-89

Kamiński S., 1992 4th ed., *Science and method: the notion of science and classification of sciences* (in Polish), Lublin: TN KUL

Kaprio J., Koskenvuo M., 1989, Twins, smoking and mortality: a 12-year prospective study of smoking-discordant twin pairs, *Social Science and Medicine,* 29, p. 1083-89

Kawalec P., 2003, *Structural reliabilism: inductive logic as a theory of justification,* Dordrecht: Kluwer

Kelly K. T., 1996, *The logic of reliable inquiry,* New York: Oxford University Press

Kelly K., Glymour C., 1992, Inductive inference from theory-laden data, *Journal of Philosophical Logic,* 21, p. 391-444

Kelly K., Schulte O., Juhl C., 1997, Learning theory and the philosophy of science, *Philosophy of Science,* 64, p. 245-67

Kiiveri H., Speed T., 1982, Structural analysis of multivariate data: a review, *Sociological Methodology,* 13, p. 209-89

Kłopotek M. A., Wierzchoń S. T., 2002, *Evidential reasoning. An interpretative investigation*, Siedlce: Akademia Podlaska

Kuhn T., 1977, *The essential tension: selected studies in scientific tradition and change,* Chicago: Chicago University Press

Kuhn T., 1996 3rd ed., *The structure of scientific revolutions,* Chicago: Chicago University Press

Kuhn T., 2000, *The road since structure: philosophical essays, 1970-1993, with an autobiographical interview,* Chicago: Chicago University Press, edited by J. Conant, J. Haugeland

Kukla A., 2000, *Social constructivism and the philosophy of science,* London: Routledge

Lagnado D., Sloman S., 2004, The advantage of timely intervention, *Journal of Experimental Psychology: Learning Memory and Cognition,* 30, p. 856-76

Langley P., 2002, *Lessons for the computational discovery of scientific knowledge,* manuscript

Lauritzen S. L., 2001, Causal inference from graphical models, Barndor-Nielsen O. E., Cox D. R., Klüppelberg C. (ed.), *Complex Stochastic Systems,* London: Chapman and Hall/CRC, p. 63-107

Maher P., 1997, Depragmatized Dutch Book arguments, *Philosophy of Science,* 64, p. 291-305

Martin E., Osherson D., 1997, Scientific Discovery Based on Belief Revision, *Journal of Symbolic Logic,* 62, p. 1352-70

Martin E., Osherson D., 1998, *Elements of scientific inquiry,* A Bradford Book/Cambridge, MA.: MIT

Mochon D., Sloman S., 2004, Causal models frame interpretation of mathematical equations, *Psychonomic Bulletin & Review,* 11, p. 1099-1104

Novick L., Cheng P., 2004, Assessing interactive causal influence, *Psychological Review,* 111, p. 455-85

Ochsner A., 1953, Carcinoma of the lung, *Chicago Medical Society Bulletin,* p. 138-42

Ochsner A., DeBaker M., 1947, Primary cancer of the lung, *The Journal of the American Medical Association,* 135, p. 321-27

Ochsner A., DeCamp P., DeBaker M., 1952, Bronchogenic carcinoma, its frequency, diagnosis and early treatment, *The Journal of the American Medical Association,* 148, p. 691-97

Osherson D., Stob M., Weinstein S., 1988, Mechanical learners pay a price for Bayesianism, *Journal of Symbolic Logic,* 53, p. 1245-51

Osherson D., Stob M., Weinstein S., 1991, A universal inductive inference machine, *Journal of Symbolic Logic,* 56, p. 661-72

Osherson D., Weinstein S., 1989, Identifiable collections of countable structures, *Philosophy of Science,* 56, p. 94-105

Osherson D., Weinstein S., 1990, On advancing simple hypotheses, *Philosophy of Science,* 57, p. 266-77

Pearl J., 2000, *Causality: models, reasoning, and inference,* Cambridge: Cambridge University Press

Popper K., 1935, *Logik der Forschung,* Vienna: Springer

Popper K., 1959, *The logic of scientific discovery,* London: Hutchinson

Popper K., 1963, *Conjectures and refutations: the growth of scientific knowledge,* London: Routledge

Psillos S., 1999, *Scientific realism: how science tracks truth,* London: Routledge

Putnam H., 1956, A definition of degree of confirmation for very rich languages, *Philosophy of Science,* 23, p. 58-62

Putnam H., 1963, Degree of confirmation and inductive logic, Illinois: Open Court, Schilpp P. A. (ed.), *The philosophy of Rudolf Carnap,* p. 761-783

Putnam H., 1965, Trial and error predicates and the solution to a problem of Mostowski, *Journal of Symbolic Logic,* 30, p. 49-57

Putnam H., 1992, *Renewing philosophy,* Cambridge, MA: Harvard University Press

Ramsey F. P., 1931, *The foundations of mathematics and other logical essays*, edited by R. B. Braithwaite, London: Harcourt Brace

Reichenbach H., 1949, *The theory of probability,* Berkeley: University of California Press

Richardson A., 1998, *Carnap's construction of the world: the Aufbau and the emergence of logical empiricism,* Cambridge: Cambridge University Press

Richardson B. W., 1936, John Snow, M.D. A representative of medical science and art of the Victorian area, Snow J., *Snow on cholera,* New York: The Commonwealth Fund, Londyn: Humphrey Milford, Oxford University Press, p. xxiii-xlviii

Robins J. M., Scheines R., Spirtes P., Wasserman L., 2003, Uniform consistency in causal inference, *Biometrika,* 90, p. 491-515

Scheines R., 1999, Estimating latent causal influences, Glymour C., Cooper G. F. ed., *Computation, causation, and discovery*, Cambridge: MIT, p. 425-37

Schulte O., 1999, Means-ends epistemology, *British Journal for the Philosophy of Science,* 50, p. 1-31

Schulte O., 1999, The logic of reliable and efficient inquiry, *Journal of Philosophical Logic,* 28, p. 399-438

Schulte O., 2000, Inferring conservation laws in particle physics: a case study in the problem of induction, *British Journal for the Philosophy of Science,* 51, p. 771-806

Schulz L., Gopnik A., 2004, Causal learning across domains, *Developmental Psychology,* 40, p. 162-76

Sloman S., Lagnado D., 2005, Do we "do"?, *Cognitive Science,* 29, p. 5-39

Snow J., 1936, *Snow on cholera,* New York: The Commonwealth Fund, Londyn: Humphrey Milford, Oxford University Press

Sobel D., Tenenbaum J., Gopnik A., 2004, Children’s causal inferences from indirect evidence: backwards blocking and Bayesian reasoning in preschoolers, *Cognitive Science,* 28, p. 303-33

Solomonoff R., 1964, A formal theory of inductive inference. Part I, *Information and Control,* 7, p. 1-22

Solomonoff R., 1964, A formal theory of inductive inference. Part II, *Information and Control,* 7, p. 224-54

Spirtes P., Glymour C., Scheines R., 2000 2nd ed., *Causation, prediction, and search,* Cambridge, MA.: MIT

Stalnaker R., 1998, Belief revision in games: forward and backward induction, *Mathematical Social Sciences,* 36, p. 31-56

Stalnaker R., 1999, Extensive and strategic forms: games and models for games, *Research in Economics,* 53, p. 293-319

Swanson N. R., Granger C., 1997, Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions, *Journal of the American Statistical Association,* 92, p. 357-67

Timberlake M., Williams K. R., 1984, Dependence, political exclusion, and government repression: some cross-national evidence, *American Sociological Review,* 49, p. 141-46

van Fraassen B., 1980, *The scientific image,* Oxford: Clarendon

van Fraassen B., 1989, *Laws and symmetry,* Oxford: Clarendon

van Fraassen B., 2002, *The empirical stance,* New Haven: Yale University Press

Waldmann M., Hagmayer Y., 2005, Seeing versus doing: two modes of accessing causal knowledge, *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 31, p. 216-27

Wheaton B., Muthen B., Alwin D., Summers G., 1977, Assessing reliability and stability in panel models, *Sociological Methodology,* 8, p. 84-136

Wynder E., 1954, Tobacco as a cause of lung cancer, *The Pennsylvania Medical Journal,* 57, p. 1073-83

Wynder E., Bross I.,Cornfield J., O'Donnell W., 1956, Lung cancer and women, *The New England Journal of Medicine,* 255, p. 1111-21

Wynder E., Graham E., 1950, Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma, *The Journal of the American Medical Association,* 143, p. 329-38

Wynder E., Graham E., 1951, Etiologic factors in bronchogenic carcinoma with special reference to industrial exposures, *Archives of Industrial Hygiene and Occupational Medicine,* 4, p. 221-35

Wynder E., Graham E., Croninger A., 1953, Experimental production of carcinoma with cigarette tar, *Cancer Research,* 13, p. 855-64

Wynder E., Graham E., Croninger A., 1955, Experimental production of carcinoma with cigarette tar. II. Tests with different mouse strains, *Cancer Research,* 15, p. 445-48

Wynder E., Lemon F., Bross I., 1959, Cancer and coronary artery disease among Seventh-Day Adventists, *Cancer,* 12, p. 1016-28