Bioinformatics and Biological Reality

Ingvar Johansson
The Institute for Formal Ontology and Medical Information Science (IFOMIS)
Saarland University
Germany

(Published in *Journal of Biomedical Informatics* 39: 274-287; issue 3, 2006)

Ingvar Johansson
IFOMIS, Saarland University
Postbox 151150
D-66041 Saarbrücken
Germany
ingvar.johansson@ifomis.uni-saarland.de
Fax: +49 (0)681 302 64772

Abstract

Many bioinformaticians seem to shy away from believing that we can have knowledge about a mind-independent biological reality. This paper attempts to show that this tendency is neither well-founded nor harmless. Even though most bioinformaticians work only with terms and concepts, they cannot altogether disregard the question whether these terms and concepts have any real referents. The paper consists of three parts. Part I clarifies three different positions in the philosophy of science with which it would be good for the philosophical outlook of bioinformaticians to become familiar, and it defends one of them, Karl Popper's epistemological realism. Part II discusses a distinction which is necessary for epistemological realism and is of practical importance for bioinformatics, the distinction between the *use* and *mention* of terms and concepts. Part III, finally, contains some brief concluding words about realism, both in general and in relation to bioinformatics.

Keywords: Gene Ontology, CRISP, HL7 RIM, bias, fictionalism, realism, truthlikeness, use-mention distinction, Karl Popper, Thomas Nagel.

Many bioinformaticians seem to shy away, if not from believing that there is a mind-independent biological reality, at least from believing that we can have knowledge about such a reality. The aim of this paper is to try to eliminate this tendency towards epistemological anti-realism. The paper consists of two main parts and a brief concluding part. Part I clarifies three different positions in the philosophy of science with which it would be good for the philosophical outlook of bioinformaticians to become familiar. When they are spelled out in some detail it becomes evident that these positions are mutually exclusive, but when seen only vaguely, the false impression may arise that one can sometimes rely philosophically on one position and sometimes on another. I will label them "Myrdal's Biasism", "Popper's Epistemological Realism", and "Vaihinger's Fictionalism", respectively.[1] I will defend Popper's position. Part II infuses new blood into the common semantic distinction between the *use* and *mention* of terms and concepts;[2] both the red and the white blood corpuscles in this new fluid come from the philosophy of intentionality. The view here defended both underpins the epistemological realism defended in part I, and shows that this realism is not only important for bioinformaticians' philosophical self-understanding, but is also sometimes of direct relevance for their ordinary work.

Part I

IA. Myrdal's Biasism

Now and then we think of, and even perceive, the world in a way that is closer to how we would like it to be than how it really is. In such situations, we are biased. But how often does this occur? And what are the consequences of such bias for scientific research? One position in the philosophy of science can be captured by the following thesis and proposal:

- Thesis: Every conceptualization and theory is biased.
- Proposal: Admit that you are biased and make the causes of this bias (valuations, social positions and backgrounds, etc.) explicit, both to yourself and to your readers.

This position, nowadays widespread, was first put forward in the fifties by the economist Gunnar Myrdal (who shared the Nobel Prize in economics with Friedrich Hayek in 1974), but only as a thesis about conceptualizations in the social sciences [1,2, 3 (chapter 7)]. Myrdal's views quickly reached the general philosophical audience thanks to Ernest Nagel's criticism of them in his classic *The Structure of Science* [4].

At the time Myrdal was writing, it was commonly assumed that scientists in their research activities ought to be, and mostly were, neutral with respect to valuations and values that are not purely scientific in the way some methodological norms are.[3] In criticism of this assumption, Myrdal claimed (a) that it is impossible for social scientists to free themselves

[1] The content of this paper has gradually come to fruition over the course of many conferences and workshops concerned with philosophy and informatics. The conference "Ontology and Biomedical Informatics" in Rome, 29 April – 2 May 2005 finally triggered me to make these thoughts as clear as possible – even to myself. Both biasism (but not Myrdal's) and Vaihinger's fictionalism were, quite independently of me, put on the agenda in Rome by Alexa McCray's talk "Conceptualizing the World: Lessons from History".

[2] I will deny my own preferences and use "term" and "concept" instead of "word" and "meaning", respectively, in order to conform to the usage of bioinformaticians. To a non-Platonist philosopher such as myself, the term "concept" suggests too many allusions to entities that exist in some extratemporal realm of their own, independently of human beings. "Meaning", on the other hand, has no such associations. Meanings exist directly only in people.

[3] Myrdal prefers the term "valuations", since he thinks that the term "values" gives rise to misleading associations of being something objective; see [2], p. 8.

from all such valuations, and (b) that such valuations necessarily distort research. According to Myrdal, since the value-neutral social scientist is a myth, social science is always more or less biased and more haunted by conflicts than the natural sciences are, and the only thing that scientists can do to become more objective is to find out and clarify, both for themselves and their readers, what kind of valuations they have. Looking at the historical development of the natural sciences, one might then add that even though there is much scientific consensus among natural scientists at most points in time, there is nonetheless a great divide between natural scientists belonging to different epochs (contrast Europe, for example, at the times of Newton and Einstein). Such differences, it has been argued, are due not to the discovery of new facts but to the different cultural values of the centuries and scientists in question. In this way, many people have moved from Myrdal's own biasism, which is restricted to the social sciences, to the generalized version, which applies to all sciences that are not purely formal. Logic and mathematics are mostly regarded as being outside the scope of biasism, but I have seen no claim that bioinformatics should be so regarded.

As I will show, biasism (in whatever version) contains at least three serious philosophical flaws, each of which is sufficient reason to reject it.

1.

It makes no sense to speak of something being to the right if there is nothing that can be said to be to the left; similarly, it makes no sense to speak of bias if it cannot be contrasted with truth. Biasism does not in Myrdal's writings, and cannot without losing its sense, take the concept of truth wholly away. What it does do is to claim that we cannot *know* truths and that we should therefore speak of research results as being true-for-certain-valuations instead of being just true.

Biasism does not say that *sometimes* scientists are biased and put forward research results that are distorted and therefore false. The claim of biasism is that this is *always* the case; either only in the social sciences (the restricted thesis) or in all the non-formal sciences (the general thesis). Let me compare biasism in science with issues of legal jurisdiction. Judicial procedure seeks to find non-biased judges and jury members. If biasism were applied to such procedures, it would amount to the claim that there are no non-challengeable persons at all. Because of this generality, the thesis of biasism has to be applied to itself. It then implies the following disjunction: Either biasism is false or it is true, but in the latter case it says of itself that it is biased and therefore false. That is, it is necessarily false. Therefore, of course, it should not be adhered to.

However, the self-referential paradox of biasism can be taken away. The defenders of biasism merely have to claim that their thesis lies outside the harmful influences of valuations and that they, therefore, are in a position to state a known truth: "All theories are biased, except the theory of biasism." But now another problem rears its head. They have to explain why their thesis – a thesis which belongs to the sociology of knowledge – is, in contradistinction to all other scientific and philosophical hypotheses, not influenced by valuations. If their thesis really is true, then it seems to be a mystery why not also scientific assertions of other sorts can be true. As far as I know, no one has solved this problem; I think it is unsolvable.

There are at least two reasons why many otherwise good researchers do not notice the paradoxical character of biasism. First, it seems to be natural for people who make assertions such as "Humans are always fools" and "Humans are always liars" to place themselves outside the scope of what they say; if not altogether, at least at the moment of making the assertion. Those who have asserted "Humans are always biased" might have followed this habit without noticing it. Second, in the case in hand, it is easy to make a so-called fallacy of composition. That is, from the fact that something is possible in *each* case, one falsely thinks that one can draw the conclusion that this something is possible in *all* cases taken

collectively. Obviously, from the fact that in a marathon race *each* starting runner may win, one cannot validly draw the conclusion that *all* runners may win. Similarly, but less obviously, from the fact that *each* scientific hypothesis may be biased, one cannot by means of mere logic draw the conclusion that *all* scientific hypotheses may be biased.

2.

The biasist proposal says that scientists should make the causes of their biases explicit, but according to the biasist thesis, even such a presentation of one's bias must itself be biased and therefore false. Why? Because to state what has caused one's bias is as much to put forward a hypothesis as other empirical assertions are. According to the thesis, it must be impossible to know what the true causes of one's bias are. If biasism is true, researchers do not only automatically get a distorted view of what they study, they also get a distorted view of what has caused their distorted research results. There are, so to speak, distortions all the way down. Therefore, it would be vain to follow the biasist proposal.

However, as in relation to the first flaw, the defenders of biasism may attempt to bypass this self-referential oddity by qualifying their position. Confronted by this second curiosity of their position, they may claim that their proposal makes good sense since researchers are *less* biased when they try to find truths about the causes of their bias, than they are when they try to find other scientific truths. For instance, it may rhetorically be asked: Isn't it easier for an economist to find out what his sex, ethnicity, social background, and social valuations are than to find out how, in some respect, the market works?

I have two counter-remarks. The relevant problem is much harder than merely discovering facts about one's social position and background. The real problem is to discover what causes distortions in one's own research. In such an undertaking, one has also to take into account the fact that sometimes people with the same social position and background have different opinions. Furthermore, and more decisive, is the fact that this qualification breaks the biasist frame. If there are degrees in the way researchers are biased, there are degrees of distortion in research results; and if there are degrees of distortion, there are degrees of being true or false, i.e., degrees of being "truthlike" (to anticipate the section on Popper's realism below). But if there are such differences of degree of distortion between ordinary hypotheses and hypotheses about factors that cause bias, then there seems to be no reason why one should not be able also to detect such differences in truthlikeness between ordinary hypotheses.

The critique that I have presented so far takes seriously the fact that biasism puts forward an all-embracing thesis, which, in effect, replaces the notion of "true" with the notion of "true-for-certain-valuations", which, in turn, can ground notions such as "true-for-us" and "true-for-them". Such a replacement leads, as I have pointed out, to inconsistencies. This criticism, it has to be noted, by no means implies that we are never justified in talking about bias in science. In local cases, and having recourse to the notion of truth, we seem now and then to be justified in asserting that some scientists have been biased. Local accusations of bias must be kept distinct from biasism, which contains a universal thesis.

For several decades now, biasism has come rather spontaneously to many researchers and people with academic backgrounds. One causal factor behind this fact might be the following. Nowadays, a large number of people in Western societies earn their living performing research or research-like activities in which the final research report takes the form of a consensus statement written by a group. This is true of public commissions of inquiry, be they initiated by the state or some regional or local authority; it is true of research departments in big firms; and of the managements of many research institutes. In such groups, after the research is performed, there comes a phase in which the final results are negotiated. This process can easily convey the false impression that there are no truths at all, only negotiations about truths and, therefore, only truths-for-certain-valuations, truths-for-us, and truths-for-them. As far as I can see, many bioinformaticians have a similar kind of

experience when they try to do justice to the advice and opinions of experts in various domains of knowledge.

3.

Let us imagine for a while that biasism has no self-referential problems. Nonetheless, another curiosity appears. All research needs a regulative idea, something that tells the researchers what to look for. Traditionally, the overarching regulative idea has been truth. This does not mean that truth has to be in the center in every phase and corner of research. For example, physicists now and then play with possible solutions to some equations without, for the moment, bothering about *truth* at all. Similarly, some biologists may play with simulations of various biological processes; and researchers in the humanities may play with certain possible interpretations of texts. This means only that there is even in these play-situations an indirect connection with the discovery of real natural laws and physical facts, the discovery of real biological processes, and the discovery of true interpretations, respectively. Today's science relies on a division of labor in which many parts have a very indirect connection to experiments and observations. This being noted, the third flaw of biasism can be stated as follows:

- Biasism wants science to get rid of the regulative idea of truth, but it has no adequate alternative to offer.

According to the biasist proposal, researchers should admit that they are biased and make the causes of their bias explicit. But if, as biasism asserts, we do not have access to the truth, what is the purpose of this proposal? Since a rational person should not seek truth if he or she firmly believes that one cannot come any closer to it, the proposal in effect implies that researchers must exchange the regulative idea of truth for the idea of truth-for-the-researcher's-valuations. The latter could be specified either as the idea that researchers should try only to promote their long-term interests or that they should in the course of their research simply try to have as much fun as possible. Although such a substitution has no logical flaws, it amounts to a substitution of science with something else. It implies, contrary to Myrdal's intentions, that researchers should be allowed consciously to deviate from data and consciously to ignore data that that they suspect are problematic *given their valuations*, as long as they are as frank as possible about what these valuations are. In secret, individual researchers may well have such goals, but these goals cannot possibly be made the public goal of science. Who, for instance, would fund researchers who say that they will use their research money only in order to promote their own egoistical interests or only in order to have fun?

Note that the remarks just made are not at all concerned with the question of how *research problems* within one's discipline *are chosen*. Such choices can of course easily be related to valuations and interests. This reminds us that several factors may account for the popularity of biasism, one of which is a conflation of the choice of a regulative idea for one's research (truth, long-term subjective interests, short-term fun) with the choice of a research problem. Another factor may be a neglect of the fact that all researchers are confronted with what might be called an existentialist situation and choice: Shall I primarily try to find the true solution to my problem, or shall I primarily try to "find" a result that promotes my interests, or shall I primarily try merely to have fun? I call the choice "existentialist", since it is inevitably a personal choice that every researcher has to make; it can be made consciously, half consciously, or subconsciously, but it has to be made. I write "primarily", since in the best of all possible research worlds seeking truth, promoting one's interests, and having fun can be realized by means of the same activity.

Yet another factor behind the popularity of the biasist proposal, perhaps the most influential, is the fact that there really is something to the proposal when it is viewed from the perspective of the *readers* of research reports. However, this requires that the proposal be put within a traditional framework in which both the researchers and the readers are seeking the truth – a fact which is not noted by the adherents of biasism. In what way can someone who reads a research report, and who is interested in the truth, be helped by coming to know the valuations of the researchers in question? On non-biasist premises, the answer is simple. As soon as there is a division of labor in the knowledge enterprise of a community, the sources of knowledge traditionally discussed in epistemology, namely reason and observation, are complemented by trust (in people providing information) [5, 6]. In order for laymen to accept knowledge or information from researchers, and in order for researchers to accept knowledge or information from other researchers, the one without knowledge has to trust the one with knowledge; in information science, knowledge engineers normally trust the domain experts. Readers may be helped in this trust issue if each researcher states: "Trust me or not; I have done my best to find the truth with ordinary methodologies, but if you suspect that I have distorted facts in order to further my interests, then I can tell you that my sex, ethnicity, social backgrounds, and social valuations are as follows: …" An imagined example, for simplicity's sake not taken from biology, will make the point more concrete.

Think of the following situation. Two different investigations have been made about the income distribution for a certain kind of job. According to report A, the average income is 15% higher for men than that for women, but according to report B it is 25%. The researcher behind report A states that he is a male income statistician who thinks that men ought to have higher salaries than women, and that, in particular, a 15% difference is too little, whereas the person behind report B states that she is a female statistician who thinks that men and women ought to have the same salaries for the same kind of job, and that a 25% difference is far too much. Whose report should be trusted? Both reports cannot be true, although both can be false. In my opinion, if it is impossible to perform further investigations of one's own, it is somewhat rational to trust the person whose values one shares. There is thus a kernel of truth in the biasist proposal that researchers should make their valuations, social positions and backgrounds, and so forth, visible, but this kernel has here been placed within a context where traditional truth seeking is taken for granted. That is, the researchers in the example have asked themselves "What is the truth, and what are the facts?", and their readers have asked themselves if the researchers have really found the true income distribution. A researcher who suspects that he (or she) unconsciously distorts facts ought to make his (or her) investigation twice. He should first make it, so to speak, spontaneously, and he should then work through it once more with the conscious intention of trying to detect hitherto unconscious distortions.

Conclusion: neither researchers in biomedical informatics nor in any other field should try to understand their work philosophically in terms of biasism.

IB. Popper's Epistemological Realism

Outside the philosophy of science, Karl Popper is most well known for his defense of democracy in *The Open Society and Its Enemies* [7]. Within the philosophy of science he is best known for his *falsifiability criterion* and for his advocacy of *fallibilism*.

Fallibilism is the view that no presumed knowledge, not even scientific knowledge, is absolutely certain. Popper's falsifiability criterion consists in the thesis that scientific hypotheses, but not metaphysical assertions, are falsifiable, and that, therefore, scientists but not metaphysicians are able to state in advance what empirical findings could make them regard their hypotheses as false. In order to try to put his falsifiability criterion to real work,

Popper connects this criterion with several other general methodological rules. Here, however, I will present only his general epistemological realism.[4]

As will become clear in what follows, I wholeheartedly accept the central features of Popper's epistemological realism, but I believe, and have argued, that his falisifiability criterion and its concomitant rules have to be rejected [11]. Thus Popper's general realism can be dissociated from his methodological rules and from his views that there is a gap between science and metaphysics as well as a criterion for detecting this gap. In particular, I will highlight a notion that is crucial to Popper's realism and which he verbalizes using three different expressions: truthlikeness, verisimilitude, and approximation to truth [9 (chapter 9)]. This extremely important notion is unfortunately neglected outside circles of Popper experts. In my opinion, the core of Popper's epistemological realism can be captured by the following thesis and proposal:[5]

- Thesis: Every conceptualization and theory almost certainly contains some mismatch between theory and reality.
  (Compare Myrdal: Every conceptualization and theory is biased.)
- Proposal: Seek truth but expect to find *truthlikeness*.
  (Compare Myrdal: Make your valuations, social positions and backgrounds, etc. visible.)

Popper's epistemological realism combines fallibilism with the traditional idea that truth seeking has to be the regulative idea of science. The key to this mix is the notion of truthlikeness (verisimilitude, approximation to truth), roughly that a statement can be more or less true. The intuition behind this notion is easily captured. Look at the three assertions (1) "The sun is shining from a completely blue sky", (2) "It is somewhat cloudy", and (3) "It is raining"; or at the assertions (1) "There are four blood groups plus the Rh factor", (2) "There are four blood groups", and (3) "All blood has the same chemical composition". In either case, if the first assertion is true, then the second assertion has a higher degree of truthlikeness and approximates truth better than the third one. This is not to say that the second is *more likely to be wholly true* than the third. The sentences "X is *probably* true" and "X *probably* has a high degree of truthlikeness" express relations between an assertion X and its evidence, whereas the sentences "X is true" and "X has a high degree of truthlikeness" express relations between the assertion X and facts (truthmakers) in the world. The former sentences express evidential relations, the latter express semantic-ontological relations;[6] the idea of truthlikeness belongs to a correspondence theory of truth.[7]

At the end of a line of all possible progressively better and better approximations to truth, there is of course truth. To introduce degrees of truthlikeness as a complement to the simple opposition between true and false is a bit – but only a bit – like switching from talking only

[4] This realism is best spelled out in [8], in particular in chapters 1-4 and 10, and in [9], in particular in chapters 2, 5, and 7-9. His falsifiability criterion and most of his methodological rules are put forward in [10].

[5] Of course, an epistemological realism presupposes a philosophical-ontological realism. With respect to the spatiotemporal world, Popper has a kind of level ontology (with which I wholly agree), according to which neither biological reality nor mental reality can be ontologically reduced to lower levels. Also, he thinks that thought contents have a kind of objective existence in what he calls the "Third World"; material reality is the First world and mental reality the Second World.

[6] The possible conflation between being "truthlike" and being "probably true" comes more easily in some other languages. In German, for instance, the corresponding terms are "*wahrheitsähnlich*" (truthlike), "*wahrscheinlich*" (probable), "*Wahrheit*" (truth), and "*wahr*" (true).

[7] The correspondence theory of truth says that the truth of an assertion (truthbearer) consists in a relation to reality or in a correspondence with facts (truthmakers). Note that there can be no degrees of "falsitylikeness"; there are no non-existent facts with which an assertion can be compared. But, of course, one may use "being falsitylike" as a metaphor for having a low degree of truthlikeness.

about tall and short people to talking about the numerical or relative lengths of the same people. The difference is this. Length corresponds both to real comparative and numerical concepts, but there are no such concepts for verisimilitudes. All lengths can be linearly ordered (and thus give rise to a real comparative concept), and a general numerical distance measure can be constructed for them (which gives us a quantitative concept). Popper thought that such concepts and measures of degrees of truthlikeness could be constructed, but like many others I think that the ensuing discussion shows that this is impossible [12 (chapter 7)]. That is, we have only a qualitative or semi-comparative concept of truthlikeness. Some philosophers think that such a concept of truthlikeness can be of no use [12 (p. 198-9)], but this is too rash a conclusion.

To demonstrate that even a semi-comparative concept of truthlikeness can be useful and important, I will use an analogy. We have no real comparative concept for geometrical shapes, to say nothing of a quantitative concept and measure. Nonetheless, we continue to use our qualitative concept of shape; we talk about shapes, point to shapes, and speak informally about similarities with respect to shape. Sometimes we make crude estimates of similarity with respect to shapes and are able on this basis to order a small number of shapes linearly (shape A is more like B than C, and A is more like shape C than D, etc.); we might be said to have a semi-comparative concept. In my opinion, such estimates and orderings of a small number of cases are also sufficient to ground talk of degrees of truthlikeness.

In the same way that a meter scale cannot be used before it has been connected to something external to it, a standard meter, so the concept of truthlikeness of theories cannot be used until one has judged, for each domain in which one is working, some theory to be the most truthlike one. In this judgment, evidential relations, left out of account in the definition of truthlikeness, stage a comeback. As I have said, truthlikeness informally measures the degree of a theory's correspondence with facts, not the degree of its correspondence with evidence. Nonetheless, in order to judge how close a theory comes to the facts, degrees of evidence must somewhere come into play. Note that such judgments are commonplace decisions even for biasists and social constructivists. They are made every time some course book in some discipline is chosen to tell students some facts.

The notion of truthlikeness is epistemologically very important. The history of science tells us that it is no longer possible to believe that science progresses by adding one bit of truth to another in the way brick houses are built by laying bricks on top of each other. Whole theory edifices have often had to be revised, and new conceptualizations introduced; this sort of development will probably continue for a long time, perhaps forever. If, in this predicament, one has recourse only to the polar opposition between true and false, and is asked whether one believes that there are any true theories, be it in the history of science, in today's science, or in the science of tomorrow, then one has to answer "There are none". If, however, one has recourse to the notion of truthlikeness, then one can answer as follows:

There are so far no absolutely true empiricial theories, but, on the other hand, there are not many absolutely false theories either. Most theories in the history of science contain some degree of truthlikeness, even if only a very low degree. Today, however, some theories have what is probably a very high degree of truthlikeness. Why? Many modern inventions and modern standardized therapies which are based on scientific theories have proven extremely effective. It seems highly unlikely that all such inventions in technology and medicine are based on theories with very low degrees of truthlikeness, to say nothing of the thought that these theories are mere fictions (see next section) or figments of the imagination. Think, for instance, of travel to the moon, images from Pluto, computers, the internet, the GPS system, physiologic contraception, artificial insemination, and organ transplantations.

It is now time to add a quotation from Popper in order to show how he himself summarizes his views on truthlikeness:

> I have in these last sections merely sketched a programme of combining Tarski's theory of truth with his Calculus of Systems so as to obtain a concept of *verisimilitude* which allows us to speak, without fear of talking nonsense, of *theories which are better or worse approximations to truth*. I do not, of course, suggest that there can be a criterion for the applicability of this notion, any more than there is one for the notion of truth. But some of us (for example Einstein himself) sometimes wish to say such things as that we have reason to conjecture that Einstein's theory of gravity is *not true*, but that it is a *better approximation to truth* than Newton's. To be able to say such things with a good conscience seems to me a major desideratum of the methodology of the natural sciences [9 (p. 335)].

Just as in ethics there are people who only think in terms of white or black and who always want to avoid nuance and complication, so in science there are people who simply like to think only in terms of true or false. Not many decades ago scientists thought of their research only in terms of being certainly true; today, having familiarized themselves with the history of science, many – and especially in domains like informatics – think of it only in terms of being certainly false or certainly fictional (see next section). In neither of these positions – being certain that one has truth on one's side, or laying no claims to truth at all – must researchers fear criticism; but on fallibilist premises researchers must.

Applying the notion of truthlikeness to the history and future of science allows us to think of scientific achievements the way engineers think of technological achievements. If a machine functions badly, engineers should try to improve it or invent a new and better machine; if a scientific theory has many theoretical problems and empirical anomalies, scientists should try to modify it or create a new and more truthlike theory. As in engineering it is natural and common to invent imperfect devices, it is in science natural and common to create theories that turn out not to be true. In both cases, however, there is an obligation to try to improve on things, i.e., improve on problematic machines and problematic theories, respectively. Also, and for everybody, it is of course better to use existing technological devices than to wait for tomorrow's, and it is better to trust existing truthlike theories than to wait for the science of tomorrow.

Most rules have exceptions. Perhaps bioinformaticians, unlike scientists in other disciplines, need not bother about the history of science or think through the conflict between Popperian fallibilism and biasism? Isn't it enough for bioinformatics simply to systematize what the present-day experts in biology tell them? No, it is not. Biological knowledge grows rapidly, and even a young discipline like bioinformatics will no doubt soon have to revise some of its achievements in light of new biological knowledge. In the Gene Ontology (GO), this is taking place before our eyes. The constructors of GO list for example what they call "obsolete molecular functions". In some cases, what is listed would better have been called only "obsolete *terms* for molecular functions", but in some cases the "obsolete molecular function" is a biological structure that was earlier wrongly ascribed a molecular function (e.g., amyloid protein, azurin, and cell surface antigen). Neither biasism nor fictionalism (see next section) can make sense of such a straightforward use of the term "obsolete", but Popper's epistemological realism can.

IC. Vaihinger's Fictionalism

In the 1920s and 30s Hans Vaihinger's book *The Philosophy of As-If* [13] saw much success. In one way it is related to the general positivist trend of those times, according to which only sensations or sense-data exist; and in another way it is related to the social constructivist

trend of the previous decades, according to which everything in the world is a social construction. The essence of his position is:

- Thesis: Absolute truth, if such there is, is not attainable.
  (Compare Popper: There is absolute truth, but it is probably not attainable.)
- Proposal: Regard your theories as referring to fictions; don't concern yourself with truth and falsehood.
  (Compare Popper: Regard your empirical theories as referring to the world; try to find out if they are false.)

Vaihinger holds that there is only one kind of real entity, the contents of our sensations (this is his positivist side). Things and persons in the ordinary sense, matter and energy as spoken of in physics, and things in themselves as postulated by some philosophers, are all merely fictions. Nonetheless, there are reasons for us to live *as if* many of these latter kinds of entities are real; the expression "live as-if X exists" which is at the heart of Vaihinger's philosophy should be understood as follows:

- If there were Xs and we knew it, then we would have to expect some specific things to happen, and we would have to act in some specific ways. In fact, however, we know that there are no Xs. Nonetheless we ought to create expectations and act *as if* there are Xs.

In some parts of his book, Vaihinger makes clear distinctions between (i) "hypotheses" (which are directed towards reality and demand verification), (ii) "semi-fictions" (which abstract away some known features of an entity, as for example the irrationality of humans is abstracted away in the concept of "homo oeconomicus"), and (iii) "pure fictions" (which are based on no abstraction of this sort); but in the end he turns everything (except the contents of sensations) into pure fictions and says:

> we are able ultimately to demonstrate that what we generally call truth, namely a conceptual world coinciding with the external world, *is merely the most expedient error*. […] So-called agreement with reality must finally be abandoned as a criterion [13 (p. 108)].

He stresses the importance and necessity of postulating fictions in all areas of life, practical, scientific, as well as ethical. Like most modern Anglo-American social constructivists, he implicitly takes it for granted that we can communicate with each other about such fictions, in other words, he implicitly regards communication as real.[8] Since sensational contents play a subordinate role in his philosophy, it is no accident that his ideas can be summarized in such a way that they become, as here, lumped together with those of present day social constructivists.

It is interesting to note how similar Vaihinger's and Popper's theses are and, despite this, how dissimilar their proposals are. In my opinion, the small difference between their theses is of no importance at all. Even if Vaihinger had subscribed to the view that there is some low probability that absolute truth is attainable, I am sure that he would have put forward the same fictionalist proposal. Conversely for Popper, even if he had thought that absolute truth is in principle unattainable, he would still have put forward the same falsificationist proposal.

[8] It should be noted that some French post-structuralists, e.g., Derrida, even regard the idea of communication as a fictional idea; nonetheless, they communicate this thesis in many books.

What, then, makes Vaihinger and Popper differ so radically in their proposals? In short, my answer is: Vaihinger's lack of the notion of truthlikeness.

False and fictional assertions are in one respect different and in another similar. They are different in that it is possible to tell a lie using a false assertion but not using a fictional one. When we lie we present as true an assertion that is false, but fictional assertions are beyond the ordinary true-false dimension. The two are similar in that neither refers to anything in reality that corresponds exactly to the assertion in question. A false empirical assertion lacks as a matter of fact a truthmaker, and a fictional assertion cannot possibly have one. Therefore, it is easy to confuse the view that all theories are false with the view that all theories are about fictions. Nonetheless, it is astonishing how easily Vaihinger moves from speaking about theories as being false to speaking about theories as being about complete fictions. Why does he not believe that there can be degrees of fictionality? There is a gulf neither between his semi-fictions (or "idealizations") and pure fictions nor between his semi-fictions and hypotheses; and Vaihinger never tries to prove that there is. Obviously, but not noted by Vaihinger, the less that has been abstracted away in a semi-fiction, the closer an assertion about it is to a hypothesis, and the more that has been abstracted away, the closer an assertion about it is to a purely fictional assertion. Assertions about semi-fictions might be said to be semi-true, and since being semi-true takes degrees, we have only created another name for truthlikeness.

If one assertion is more truthlike than another, then it is by definition also less false. However, this "falsity content" (to take an expression from Popper) can easily be turned into a "fictionality content", whereupon the more truthlike assertion can also be said to be a less fictional assertion. When we are reading about, say, Sherlock Holmes, we have no difficulty in placing this fictional character in a real setting, London between 1881 and 1904. In many fictional discourses not everything is fictional, and we often have no difficulty in apprehending such mixtures of real and fictional reference. Something similar is true when one reads about the history of science. For example, when I read about the false hypothesis that there is a planet Vulcan between Mercury and the Sun, which might explain some seeming falsifications of Newtonian mechanics, then I had no problem in taking Vulcan to be a fictional entity postulated as existing in the real solar system in about the same way as I take Holmes to be a fictional character in a real London. When I read about the false hypothesis that there is a chemical substance, phlogiston, which *exits* burning material (where in truth, as we now know, oxygen *enters* burning material), then I have no problem in taking phlogiston to be a fictional substance in the world of real burnings. When I read about Galen's view that (what we call) the arterial system contains pneuma or spiritus, then I have no problem in taking this pneuma to be fictional, but the arterial system to be real. Those who write about the history of science often make the reader look upon statements which were once false assertions as being assertions about fictions. In retrospect, we should look upon superseded theories as containing unintentionally a mix of reality and fiction in the way reality and fiction can be intentionally mixed in novels. This is to give fictions their due place in science, but such local uses of fictions must be kept distinct from fictionalism, which contains a universal thesis.

Fictionalism has, it should also be noted, one major flaw in common with biasism. Apart from all other curiosities, fictionalism is self-reflectively inconsistent. Fictions are created, but if everything apart from the contents of our sensations is a fiction, then there is nothing except such content that can create the fictions. However, contents of sensation do not have such a capacity. Unfortunately, Vaihinger and most fictionalists do not see the need for this kind of self-reflection.

The fact that Popper's epistemological realism is far more reasonable than biasism and fictionalism does not imply that it is completely free from problems and philosophical lacunae. Popper was not, for instance, interested in the philosophy of language. In the next

part of the paper, I will put forward a semantic idea that not only underpins epistemological realism, it is also of direct relevance for some seemingly non-philosophical work in bioinformatics.

Part II

IIA. Use and mention – in the light of an optical metaphor

When we look at things such as stones, trees, and walls, we cannot see what is on the other side. But things like water and glass are such that we can look through them to the other side. In the case of glasses, microscopes, and telescopes, this feature is extremely useful. By *looking through* such lenses, we are able better to *look at* something else. This phenomenon of "being-aware-of-*x*-through-*y*" is not restricted to the visual sense. It can be found in the tactile realm as well. You can grip a tool and feel the tool against your palm, but when you are very good at using such a tool, this feeling disappears. You are instead primarily aware of whatever it is that the tool is affecting. For instance, when you are painting a wall with a brush, you are only (if at all) indirectly aware of your grip of the brush, and are instead aware only of what you are painting. You are *feeling through* the brush and *feeling (at)* the wall. What glasses are for people with bad sight, the white cane is for blind people.

Speech acts, listening acts, writing acts, and reading acts – in short, language acts – are, just like glasses and the white cane, tools for improving everyday life. They can be used to convey and receive information, to give and take orders, to express emotions, and to do many other things. Even though language acts do not have the same robust material character that tools have, they nonetheless display the same feature of being able to be both "looked at" and "looked through". When you look at linguistic entities, you are directly aware of them as linguistic entities, but when you look through them you are at most indirectly aware of them. When, for example, you are conveying or receiving information in a language in which you are able to make and understand language acts spontaneously, you are *not* looking *at* the terms, concepts, statements, and propositions in question; nor are you looking at grammar and dialects. Rather, you look through these linguistic entities in order to see the information (facts, reality, objects) in question. When, then, are we looking *at* linguistic entities? We look at them, for example, when we are reading dictionaries and are examining terminologies. If I say "Look, the cat has fallen asleep", I want someone to look through the term "cat" and my assertion in order to get information about a state of affairs in the world. But if I say "In WordNet, the noun 'cat' has 8 senses" [14], then I want someone to look at the term "cat".

My distinction between looking *at* and looking *through* is similar to the traditional distinction between the *use* and *mention* of linguistic entities,[9] and it applies both to factual talk and to reading novels. In fictional discourse, terms are *used* as much as they are in talk about real things, but they are used in a very special way. Fictional discourse is *about* fictional characters; it is not about terms and concepts. In fact, we are standardly using the same terms and concepts both in fictional and factual discourse.

When you are not using lenses, you can look at them and investigate them as material objects of their own in the world. For instance, you can try to find out what their physical properties and internal structures are like. In the world of practice, we investigate tools this way only when they are not functioning properly and are in need of repairing. Something similar holds true of terms and concepts. Only when our language acts are not functioning

[9] I do not regard the distinction between use and mention as the same distinction as that between object language and meta-language. The use-mention distinction does not split ordinary language into distinct levels.

well – think for instance of learning a new language – do we normally bother to look *at* terms and concepts in dictionaries.

Furthermore, we are able to switch quickly between looking through and looking at things. Car drivers should look through, not at, the windshield, but when driving they should also have the ability to take a very fast look *at* it in order to see whether, for instance, it has been damaged by a stone. Something similar is true of people using a foreign-language dictionary. They should be able to take a look at a certain foreign term and then immediately start to look through it by using it.

Let me summarize:

1. In the same way that we can both look at and look through many material things, we can both look at and look through many linguistic entities.
2. In the same way that we can quickly switch between looking at and looking through glass, we can quickly switch between looking at and looking through linguistic entities.

And let me then continue the analogy by adding still another similarity:

3. In the same way that consciously invented material devices for "being-aware-of-*x*-through-*y*", such as microscopes and telescopes, have provided new information about the world, consciously invented linguistic devices for "being-aware-of-*x*-through-*y*", such as scientific conceptual systems, have provided new information about the world.

By means of the *invention* of new concepts, we can sometimes *discover* hitherto completely unnoticed facts. Often, we (rightly) regard discoveries and inventions as wholly distinct affairs. Some things, such as stones, can only be discovered, not invented; others, such as bicycles, seem only to be inventions. One person might invent and build a new kind of house, and other persons may later discover it; but the first person cannot both invent and discover it. These differences between inventing and discovering notwithstanding, devices for "being-aware-of-*x*-through-*y*" present an intimate connection between invention and discovery. By means of new "being-aware-of-*x*-through-*y*" inventions, we can discover *x*. There are many *x*'s that we can discover only in this way.

The third point above should partly be understood in terms of the notion of truthlikeness: If an existing conceptual system is confronted by a conflicting conceptual system which has a higher degree of truthlikeness, the latter should supersede the former. But the notion of truthlikeness should also be understood by means of the distinction between looking at and looking through. I introduced the idea of truthlikeness with the three assertions "The sun is shining from a completely blue sky", "It is somewhat cloudy", "It is raining", and I said that, *given that the first assertion is true*, the second one seems intuitively to be more truthlike than the third. A standard objection to such a thesis is that this sort of comparison can show us nothing relevant for a correspondence theory of truth,[10] since what we are comparing are merely linguistic entities, namely assertions. However, this objection overlooks the distinction between looking at and looking through. Looking at the assertions allows us to see only similarity relations between the assertions as such, but when we have learned to switch from looking at them to looking through them – at reality – then we can coherently claim that the second is more truthlike than the third.

In the same way that our choice of kind of lens may determine what we are able to see, so our choice of concepts determines what we can grasp. Such a determination is compatible

[10] For a rough description of the correspondence theory of truth, see footnote 7.

with the view that we can get knowledge about the world: it does not render truth a wholly social construction. When, through a concept, we look at and grasp something in the world, this concept often (i) *selects* an aspect of the world, (ii) *selects* a granularity level (for instance, microscopic or macroscopic), and (iii) *creates* boundaries where there are no pre-given natural boundaries. Nonetheless, the concept (iv) *does not create* this aspect, this granularity level, or what is bounded. Think of the concept "heart". It selects a biological aspect of the human body, it selects a macroscopic granularity level, and it creates a boundary line between the heart and its surroundings which does not track physical discontinuities at all points, as for example where the heart meets the aorta and the veins [15]. But, nonetheless, our invention of the concept "heart" does not *create* our hearts, and there were hearts many millions of years before there were concepts.

IIB. The fallacy of mixing use and mention

All ontologies in information science contain terms. The builders of such ontologies look mainly *at* the terms in question, whereas the users of ontologies look mainly *through* them. Like the users, the experts of the various domains of knowledge generally look through the terms. However, an ontology such as the online lexical resource WordNet presents a special case, for (if it is to be called an ontology at all) it is an ontology of terms and their meanings not about what the terms are used to refer to; it is like a dictionary, not like a taxonomical textbook. With respect to the term "cat", WordNet (2.0) starts as follows:

The **noun** "cat" has 8 senses in WordNet.

1. **cat**, true cat -- (feline mammal usually having thick soft fur and being unable to roar; domestic cats; wildcats)
2. guy, **cat**, hombre, bozo -- (an informal term for a youth or man; "a nice guy"; "the guy's only doing it for some doll") [14]

It is doubly clear the term "cat" is mentioned and not used in WordNet; it is looked at and not through. Both the scare quotes around the term "cat" and the fact that it is preceded by the term "noun" make it clear that WordNet contains no talk of real cats; both scare quotes and context can disambiguate between use and mention. Here, therefore, matters are clear. In many biomedical ontologies, however, use and mention are systematically confused.

The group that constructed the Gene Ontology, the GO Consortium, says that "[t]he Goal of the Consortium is to produce a structured, precisely defined, common, controlled vocabulary *for describing* the roles of genes and gene products in any organism [16 (p. 26), italics added]." That is, it is not an ontology for looking at terms but for looking through terms. GO consists of three distinct ontologies, one for cellular components, one for molecular functions, and one for biological processes. One graph in the latter ontology (as it looks when this is being written in June 2005) can be reproduced as in the figure below; it is to be read from bottom to top. The graph contains arrows that represent both subsumption relations (*is_a*) and part-whole relations (*part_of*).

Gene_Ontology
part_of
biological process
is_a
physiological process
is_a
metabolism
is_a
nucleobase, nucleoside, nucleotide, and nucleic acid metabolism
is_a
transcription
is_a
transcription, DNA-dependent
part_of
transcription initiation (GO: 0006352)

When a user of GO reads this he is, I am sure, looking through the terms. That is, he reads it from the bottom up as signifying something like: "Each transcription initiation is part of a DNA-dependent transcription, which is a kind of transcription, which is a special kind (nucleobase, etc.) of metabolism, which, like all metabolisms, is a physiological and biological process." So far so good, but I have stopped at "biological process". What about the last step? Reading it in the same way would yield: "Each biological process is part of the Gene Ontology"? But this is obviously false. It should instead be read: "The *term* 'biological process' is part of the Gene Ontology's hierarchy of *terms*". Thus use and mention of "biological process" are here mixed. When one reads the ontology from the bottom up and arrives at "biological process", this term should be regarded as *used*, but when one continues reading upwards, it should be regarded as *mentioned*.

Since most people are able to switch unproblematically between looking through and looking at terms, the human users of the GO may perhaps do so here without noticing, and no harm will have been done. However, automated information-extracting systems are not able to make such switches. Obviously, GO would be a better construction without this mixture of use and mention. As the graph stands, it allows a fallacious inference to the effect that if something is a biological process then it is part of a certain human artifact called the Gene Ontology. This might be called the fallacy of mixing use and mention.

The same kind of fallacy appears as well (at least in June 2005) in Computer Retrieval of Information on Scientific Projects (CRISP). There, one finds subsumption relations which can be represented as in the hierarchy below (to be read from the bottom up).

immunology
is_a
antigen
is_a
allergen
is_a
airborne allergen
is_a
pollen

Here, "antigen" should be *used* in relation to "allergen" ("Each allergen is an antigen"), but *mentioned* in relation to "immunology" ("The *term* antigen is an immunological *term*").

"Allergen" is a term among other terms in the field of immunology, whereas allergens themselves are among the entities that immunology studies.

The Health Level 7 Reference Information Model (HL7 RIM) conflates use and mention, i.e., use and mention of clinical data, with the unfortunate result that the users are told that HL7 RIM cuts them off from the world:

> Act as statements or speech-acts are the only representation of real world facts or processes in the HL7 RIM. *The truth about the real world is constructed through a combination (and arbitration) of such attributed statements only, and there is no class in the RIM whose objects represent "objective state of affairs" or "real processes" independent from attributed statements. As such, there is no distinction between an activity and its documentation*. Every Act includes both to varying degrees. For example, a factual statement made about recent (but past) activities, authored (and signed) by the performer of such activities, is commonly known as a procedure report or original documentations (e.g., surgical procedure report, clinic note etc.) [17, italics added].[11]

Let me now return to the confusion of use and mention in GO. GO's three ontologies state at the root "biological process *part_of* Gene_Ontology", "molecular function *part_of* Gene_Ontology", and "cellular component *part_of* Gene_Ontology", respectively. What to do in order to remove the inferences to the effect that biological processes, molecular functions, and cellular components belong to the artifact Gene Ontology instead of to biological reality? One solution – "the multiple ontology solution" – is for the GO Consortium to take its talk of three ontologies literally and to construe GO as having three top nodes instead of one. For those cases where a single topmost node is needed, the consortium could then employ "Gene-Ontological Entity" and write: "biological process *is_a* gene-ontological entity", "molecular function *is_a* gene-ontological entity", and "cellular component *is_a* gene-ontological entity", respectively. This way out I will call, after the author of The Foundational Model of Anatomy (FMA), "the Cornelius Rosse solution". At the top of FMA, the four nodes "anatomical structure", "body substance", "body space", and "anatomical boundary" are subsumed by the single top node "Anatomical Entity" [20].

However, just as the multiple ontology solution has its price, lack of a single top node, so also does the Rosse solution. For when we look through the terms "biological process", "molecular function", "cellular component", and all the terms they subsume, we find entities in the world that share a certain abstract feature (being a biological process, etc.) quite independently of GO. This is not unproblematically the case when we look through the term "gene-ontological entity". To describe something as being a gene-ontological entity is merely to say, either tautologically that it is an entity classified by the GO, or, more informatively, that it is an entity of interest for gene research. That is, what makes a cellular process a biological process, an organelle a cellular component, and a catalytic activity a molecular function is of another character than what makes cellular processes, organelles, and catalytic activities into gene-ontological entities. But there is here no confounding of use and mention of terms – and most things in life do have a price.

IIC. Use and mention – in the light of a good philosophy of intentionality

As I pointed out in section IIA, both terms (words) and concepts (meanings), as they are most commonly used, are invisible in the sense that we most often look through them at the entities

[11] The quoted statement (as well as other ones in HL7 RIM) is criticized in [18]. The terminology of the National Cancer Institute Thesaurus is discussed in [19], and some of the critical remarks made here are concerned also with the use-mention distinction.

to which they refer. The need for a distinction between term and concept arises as soon as we discover a synonymy, be it between two terms in the same language or in different languages. For we then have to specify what makes the terms different and what makes them similar, i.e., synonymous. The terms differ because they are constituted by different syntactic unities such as letters or words conceived of as purely graphical or acoustic patterns, and they are synonymous (as we say) because they express the same concept. A term is a fusion of a syntactic unity and a concept.[12] One looks *through* the concept, not through the syntactic unity, i.e., concepts are to terms what lenses are to glasses, microscopes, and telescopes.

The metaphor of "looking through" concepts is sustained by a certain approach in the philosophy of intentionality. The term "intentionality" was introduced to contemporary philosophy by Franz Brentano in the nineteenth century. It refers to phenomena such as perceiving, thinking, reading, and desiring. Intentional phenomena contain a *directedness* towards something. Mostly, it is a directedness that originates in a person who is in a so-called "intentional state" (e.g. in a state of perceiving a certain object), or who performs an "intentional act" (such as asserting something). There are, however, different opinions on how to analyze intentional phenomena. In my opinion, Edmund Husserl [21] and John Searle [22] have come the closest to the truth.[13] Let me quote Searle:

> it is at least misleading, if not simply a mistake, to say that a belief, for example, is a two-term *relation* between a believer and a proposition. An analogous mistake would be to say that a statement is a two-term relation between a speaker and a proposition. One should say rather that a proposition is not the *object* of a statement or belief but rather its *content*. The content of the statement or belief that de Gaulle was French is the proposition that the de Gaulle was French. But that proposition is not what the statement or belief is about or is directed at. No, the statement or belief is about de Gaulle [22 (p. 18-19)]

Intentional phenomena are marked by a tripartition between (intentional) act, (intentional) content, and (intentional) object. Assume that you are reading a physician's report about your heart, which tells you that your heart has some specific features. At a particular moment, there is then your reading *act* along with what you are reading about, the intentional *object*, i.e., your heart and its properties. But since your heart exists outside of your reading act, there must be something within the act itself in virtue of which you are directed towards your heart and its properties. This something is called the *content*; in assertions, it consists of propositions. Even false assertions contain a proposition.

According to many non-Husserlian and non-Searlean analyses of intentionality, you are in your act of reading directed only towards the proposition, but then there is outside your awareness also a relation of representation between the proposition (the content) and the object (your heart). According to Husserl and Searle, on the other hand, you are in your reading directed towards your heart (object) by means of the proposition (content). The first kind of analysis leaves no room for any sensible talk of "looking through" concepts and propositions, but Husserl's and Searle's analyses do. Though Husserl's and Searle's philosophical frameworks differ in other respects,[14] the overlap highlighted makes it

[12] Those who are amenable to Ferdinand de Saussure's linguistics can read the last sentence as follows: A sign is a fusion of a signifier and what is signified. Let me add that Saussure consciously abstracted all "looking-throughs" and referents away from his studies. Some of his present-day followers, however, seem to take the position (criticized in this paper) that there simply are no referents.

[13] In this respect see Searle [22], p. 18-9, 57-61, 97, and Husserl [21], the fifth investigation, §11 and the appendix to §21 ("Critique of the 'image-theory' and of the doctrine of the 'immanent' objects of acts").

[14] Searle is a naturalist in the sense that he thinks that everything that exists exists in the spatiotemporal world in which we live. Husserl seems at first to have been a naturalist, but later he came to propound so-called transcendentalist views.

reasonable to believe that the distinction between looking at and looking through concepts can be embedded within a very plausible theory of intentionality. This gives further credence both to epistemological realism in general and to epistemological realism in bioinformatics.

Part III

IIIA. The last word and the last word but one

In the first part of this paper I advocated Popper's realism, in particular his notion of truthlikeness. In the second part I advocated a Husserl-Searlean analysis of intentionality, in particular the view that in assertions one is directed towards the world by "looking through" terms and concepts. Social constructivists often ask: "From what position are you talking?" In order to answer this question, I will bring in still another great thinker, Thomas Nagel. I regard myself as speaking from the kind of naturalist rationalist position that he has tried to work out in *The View From Nowhere* [23] and *The Last Word* [24]. Below are two quotations. The first is from the introduction to the latter book, and the second is its ending paragraph.

> The relativistic qualifier—"for me" or "for us"—has become almost a reflex, and with some vaguely philosophical support, it is often generalized into an interpretation of most deep disagreements of belief or method as due to different frames of reference, forms of thought or practice, or forms of life, between which there is no objective way of judging but only a contest for power. (The idea that everything is "constructed" belongs to the same family.) Since all justifications come to an end with what the people who accept them find acceptable and not in need of further justification, no conclusion, it is thought, can claim validity beyond the community whose acceptance validates it.
>
> The idea of reason, by contrast, refers to nonlocal and nonrelative methods of justification—methods that distinguish universally legitimate from illegitimate inferences and that aim at reaching the truth in a nonrelative sense. Those methods may fail, but that is their aim, and rational justification, even if they come to an end somewhere, cannot end with the qualifier "for me" if they are to make that claim [24 (p. 4-5)].

> Once we enter the world for our temporary stay in it, there is no alternative but to try to decide what to believe and how to live, and the only way to do that is by trying to decide what is the case and what is right. Even if we distance ourselves from some of our thoughts and impulses, and regard them from the outside, the process of trying to place ourselves in the world leads eventually to thoughts that we cannot think of as merely "ours." If we think at all, we must think of ourselves, individually and collectively, as submitting to the order of reasons rather than creating it [24 (p. 143)].

Reason, Nagel says, has to have the last word. However, this statement needs to be qualified. As the logician Per Lindström notes with regard to Nagel's book: "reason has the last word – or perhaps only the last but one, since reality, reason tells us, has always the absolutely last word" [25 (p. 3-6)].

IIIB. Bioinformaticians and biologists

Not only for biologists, but even for bioinformaticians, biological reality should have the last word. That is, bioinformaticians should ideally consult, and only consult, biologists who try to "consult" nature. What then does the actual situation today look like? According to my experience, researchers in different disciplines *tend* to adopt different philosophical positions. Bioinformaticians, anthropologists, sociologists, cultural-studies-people, and other researchers who work mainly with data assembled in texts or interviews, tend towards various forms of anti-realism; most biologists and natural scientists tend towards both ontological and epistemological realism. According to some indications, biologists who work only with computer simulations tend towards anti-realism. Therefore, I will end with some comments on some different epistemological situations; situations which are doubly idealized. It is assumed (a) that bioinformaticians and biologists not only tend to adopt a certain epistemological position but in fact wholeheartedly adopt one, and (b) that all bioinformaticians are either realists or anti-realists, and that the same goes for all biologists. This gives rise to the following four situations:

(i) bioinformaticians are anti-realists and biologists are realists
(ii) both groups are realists
(iii) bioinformaticians are realists and biologists are anti-realists
(iv) both groups are anti-realists.

In my opinion, situation (i) comes closest to the actual situation today. Anti-realist bioinformaticians work with data handed over to them by realist biologists. This does not, happily enough, have the consequence that the information systems so constructed are useless for realists who want to look through the terms in question and at biological reality. This is because, even though the bioinformaticians may have been looking only at the terms or at the concepts and not at reality,[15] the terms come originally from domain experts who mainly looked through them at biological reality. Therefore, the anti-realist bioinformaticians become intermediaries, which the realist user can mostly forget. However, as I have pointed out in this paper, some features of some information systems can be improved if bioinformaticians became realists as well. Still another thing makes it desirable that bioinformaticians become realists: they could then possibly provide feedback to biologists. Let me explain.

Since bioinformatics is more concerned with large-scale systematic classifications than present-day biology normally is, bioinformaticians may well discover in classifications inconsistencies, indeterminacies and other curiosities that have escaped the notice of biologists. What to do in such situations? If the bioinformaticians are realists, the natural thing to do is to ask the biologists to "consult" nature again (cf. [27]), but if they are anti-realists they can in principle accept inconsistencies and indeterminacies, which may after all reflect inconsistencies and indeterminacies in the ways people "use language", and so forth .

Even though it seems highly unlikely that situation (iii) – bioinformaticians are realists and biologists are anti-realists – should ever become actual, I will use its mere possibility to make two points clear. First, there might be a gap between the public philosophical position of a researcher and the way in which he or she actually conducts research. A declared epistemological realist may actually treat data as frivolously as if there were no theory-independent reality to serve as standard for correctness, and, conversely, a declared anti-realist may in fact be just as careful as a realist in trying to do justice to all the relevant data. Second, if realist bioinformaticians meet anti-realist biologists of this kind, then they can look

[15] With respect to the term "concept", I remind the reader of footnote 2 and ask her or him also to read [26].

upon the latter as if they were realists. They can do the opposite of what, in a famous case, Osiander did.

Osiander is the man who wrote the preface to the realist Copernicus' treatise *On the revolutions of the heavenly spheres* (1543). Here, Copernicus claimed, contrary to what the Church at the time said, that the planets do not revolve around the earth but around the sun. In his preface, however, Osiander said that the book should be read as saying *merely* that the calculations of the orbits of "the heavenly spheres" could be simplified if one regarded the world *as if* it were such that the planets move around the sun. Copernicus' realist theory was reinterpreted by Osiander as being a fictionalist theory. But the opposite interpretational move is also possible. Something that is said to describe something merely fictional, or to be a mere social construction, can be reinterpreted as in fact describing something real. I have done such a reinterpretation when reading some books in the humanities and the social sciences in which the author pronounces him- or herself to be a social constructivist. Realist bioinformaticians can make it in relation to anti-realist biologists.

And finally, what would the world be like under (iv), the philosophically worst of all possible scenarios, in which both bioinformaticians and biologists are anti-realists? Well, if their anti-realism were merely idle talk, then, by definition, the world would continue very much as if they were realists. But if their anti-realism deeply influenced their practical work, this would be a true disaster. Since bioinformatics includes medical informatics and, as I have used the term "biologist", biologists include physicians and medical researchers, I can make my point clear by means of a rhetorical question: Would you like to be treated for a physiological illness by a physician who is not sure that there are human bodies, and who uses information systems created by medical information scientists who believe they are working only with terms that lack real referents?

There are even those who proclaim death to be a social construction. They are, thereby, proclaiming the death of their own anti-realism.

References

[1] Myrdal G. Value in Social Theory. A Selection of Essays in Methodology. London: Routledge & Kegan Paul; 1956.

[2] Myrdal G. Objektivitetsproblemet i samhällsforskningen ("Objectivity in Social Research"). Stockholm: Rabén & Sjögren; 1968.

[3] Myrdal G. Critical Essays on Economics. New York: Pantheon Books; 1973.

[4] Nagel E. The Structure of Science. London: Routledge & Kegan Paul; 1961.

[5] Coady C.A.J. Testimony: A philosophical study. New York: Oxford University Press; 1992.

[6] Kalman H. The Structure of Knowing. Existential Trust as an Epistemological Category. Uppsala: Swedish Science Press; 1999.

[7] Popper K. The Open Society and Its Enemies. London: Routledge & Kegan Paul (first edition in two volumes); London 1945.

[8] Popper K. Conjectures and Refutations. The Growth of Scientific Knowledge. London: Routledge and Kegan Paul; 1963

[9] Popper K. Objective Knowledge. Oxford: Clarendon Press; 1972.

[10] Popper K. The Logic of Scientific Discovery. London: Hutchinson; 1959.

[11] Johansson I. A Critique of Karl Popper's Methodology. Gothenburg: Scandinavian University Books; 1975.

[12] Keuth H. Die Philosophie Karl Poppers. Tübingen: Mohr Siebeck; 2000 (in English, The Philosophy of Karl Popper. Cambridge: Cambridge University Press; 2004).

[13] Vaihinger H. The Philosophy of As-If. London: Routledge and Kegan Paul; 1924 (first German edition 1911).
[14] WordNet: <http://www.cogsci.princeton.edu/cgi-bin/webwn2.0?stage=1&word=cat>.
[15] Smith B. Fiat Objects. Topoi 2001; 20: 131- 48.
[16] The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nature Genetics 2000; 25: 25-9.
[17] HL7 Reference Information Model, version: V 02-04 (7/28/2004) <http://www.hl7.org/> chapter 3.1.1.
[18] Vizenor L. Actions in Health Care Organizations: An Ontological Analysis. In: Fieschi M, Coiera E, Li Y-C.J., editors. MedInfo 2004. Amsterdam: *IOS* Press; 2004, p. 1403-7
[19] Ceusters W, Smith B, Goldberg L. A Terminological and Ontological Analysis of the NCI Theasurus. Forthcoming in Methods of Information in Medicine.
[20] Smith B, Mejino Jr. J L V, Schulz S, Kumar A, Rosse C. Anatomical Information Science. In: Cohn A G, Mark D M, editors. Spatial Information Theory. Proceedings of COSIT 2005 (Lecture Notes in Computer Science). Berlin/Heidelberg/New York: Springer; 2005, p. 149-64.
[21] Husserl E. Logical Investigations vol. 2. London: Routledge; 1970 (in German 1921; this is a revised version of the first ed. of 1901).
[22] Searle J. Intentionality. Cambridge: Cambridge University Press; 1983.
[23] Nagel T. The View From Nowhere. Oxford: Oxford University Press; 1986.
[24] Nagel T. The Last Word. Oxford: Oxford University Press; 1997.
[25] Lindström P. Näst sista ordet ("The Last Word But One"), Filosofisk tidskrift 1/2001: 3-6.
[26] Smith B. Reasoning with biomedical terminologies. Journal of Biomedical Informatics, under review.
[27] Johansson I. Ontologies and Concepts. Two Proposals. In: Proceedings of the KI2003 Workshop on Reference Ontologies and Application Ontologies. <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/> (vol-94).

Acknowledgment

This paper was written under the auspices of the Wolfgang Paul Program of the Alexander von Humboldt Foundation, the Network of Excellence in Semantic Interoperability and Data Mining in Biomedicine of the European Union, and the project Forms of Life sponsored by the Volkswagen Foundation.

Without my (so far) three years with Barry Smith and IFOMIS, my views would probably never have been able to distil in the way they have in this paper. Thanks Barry! Thanks also to Katherine Munn, IFOMIS, for improving my English and making good pedagogical proposals. Some nice improvements are, I would also like to say, caused by criticism from two anonymous referees.