

# Teaching Philosophy with Argumentation Maps

## Review of *Can Computers Think? The Debate* by Robert E. Horn

Thomas Metzinger  
Fachbereich 1 - Philosophie  
Universität GH Essen  
45117 Essen  
GERMANY

thomas.metzinger@uni-essen.de  
<http://www.uni-giessen.de/~gm1001/metz.htm>

Copyright (c) Thomas Metzinger 1999

PSYCHE, 5(30), December 1999  
<http://psyche.cs.monash.edu.au/v5/psyche-5-30-metzinger.html>

KEYWORDS: Artificial Intelligence, intentionality, Physical Symbol Systems, Turing test, Gödel's theorem, functionalism, imagery, connectionism, dynamicist cognitive science, consciousness, artificial consciousness, personhood.

REVIEW OF: Robert E. Horn (Ed.) (1998) *Can Computers Think? The Debate*. MacroVU Press: [www.macrovu.com](http://www.macrovu.com), a set of 7 maps and a handbook. ISBN: 1-892637-08-1. Price: US\$ 125.

Let me confess right at the outset that I am truly enthusiastic about this project: It could constitute the beginning of a new phase in academic teaching. I will use this in my own teaching of philosophy of mind as soon as I have a chance, and I definitely look forward to the experience. This is not to deny the existence of a number of disadvantages and possible dangers associated with these new didactic materials for the mind sciences (see below); however, all in all, what we are here witnessing are the first fruits of an exciting and promising project. It deserves all the support it can get.

In his seminal article in the 1950 issue of *Mind* Alan M. Turing wrote: "I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted." Time has proved him to be wrong on this point, while progressively

demonstrating how essential and far reaching his own contributions have actually been, laying the foundations for research programs like AI and theoretical approaches like philosophical functionalism. The emergence of systematic and rigorous attempts at modelling the mind in artificial and formal systems may, in a couple of centuries, look like one of the central intellectual achievements of humanity in the 20th century. The origin and historical development of these attempts will therefore have to form one of the central elements of academic teaching for many years to come -- not only in philosophy, but in cognitive science and many related new disciplines as well. What the Argumentation Maps developed by Robert Horn and his team present us with is an efficient new tool for navigating this recent phase of our intellectual history. They display the logical structure of the core debates and their wider theoretical context in a non-propositional medium, compressing an enormous amount of information into a pictorial format. If these posters stay on the walls of class-rooms over a period of months while students learn the subject matter step by step, they can be accessed again and again during different phases of the learning experience. Let's face it: the debate on whether machines could think, in the five decades which have passed since the publication of Turing's paper, has gained so much in terms of substance and internal complexity that even for experts in the field it is hard to negotiate the jungle of arguments and counterarguments. We urgently needed something like this.

Human beings are visual creatures and evolution has invested much more neurocomputational resources in our conscious visual model of reality than it has in our recent ability to internally simulate quasi-syntactic operations with discrete symbol-tokens. We are experts at grasping visual environments in a fraction of a second, whereas we are clumsy thinkers, who take much longer to reconstruct logical environments mentally. It is important, then, to supplement learning procedures and the building of a descriptive memory with some additional eye-food. The argumentation maps supplied by Robert Horn and his team do precisely that -- while not ignoring the intellectual content they are transporting. They facilitate the activation and storage of mental models of complex propositional structures in students' brains (this is one way of describing what academic teaching has to achieve in theoretical disciplines like philosophy) by embedding them into perceptual mental models (this is one way of describing what Horn's mapping approach tries to do). In doing so they make use of a visual language, which has its own set of rules, but is more flexible and allows for the quicker grasping of coarse-grained context.

The first set of argumentation maps now available describes more than 800 argumentative "moves", represents 380 authors and took the scientific team more than 7000 work hours to generate it -- drawing on over 1000 sources. They contain 130 interlinked boxes, some 60 photographs of selected authors and come with a handbook, bibliography and, importantly, a "New Claim/Rebuttal Participation Form". In their current paper version, posters are 3 feet high and 4 feet wide. However, as this is an online review and most of you will now be online I encourage you to visit [www.macrovu.com](http://www.macrovu.com) and have a look at some background information -- and even miniature versions of the posters themselves at [www.macrovu.com/CCTGeneralInfo.html](http://www.macrovu.com/CCTGeneralInfo.html). Eventually these maps will possess Hypertext-

links and be projected from screens integrated into the walls of classrooms -- at least this is the distant vision of Bob Horn. A fascinating idea. What nobody seems to have thought about is that they could form floors as well, for instance in science museums constructing an intellectual maze out of the mapped debates, allowing visitors to actually physically walk through the history of AI. Once the principle of partially embedding propositional information into perceptually driven mental models has been grasped, it should eventually be extended to other sensory modalities as well: kinaesthetic and proprioceptive feedback is only one way to go, sound files creating an acoustic landscape are another obvious further extension. Can one develop a taste for the originality or substantiality of certain debates? Would it be possible to learn how to smell fishy arguments?

This project needs and deserves all the support it can get. Let me therefore point out a number of issues that should be thought about carefully in the future, in order to make this an even better didactic tool. The underlying danger is that, as it will likely have no competitors, it may, implicitly, be looked upon as an "authoritative" model of the actual theoretical landscape by generations of students in many different countries. Realistically, many students will not read much further and these maps will stay in their memory for a lifetime, providing a general, overall impression of the problems with constructing intelligent machines. On the other hand, helpful as such maps are, they tend to filter out many relevant aspects of the debates they are mapping, and are in danger of creating a new kind of "infotainment", blurring the border between real science and philosophy and their commercialised, popular and easy-to-digest media versions.

One danger to be avoided is the "Anglo-Saxon Filter Effect" (ASFE). If this excellent didactic tool is to be used globally, the debates mapped must be truly global debates as well -- great care has to be taken to integrate the intellectual progress made in non-English linguistic communities as well. The ASFE is a simple effect constituted by a number of factors, one of them being the fact that English is now the lingua franca of global scientific debate. This is an obvious advantage for all native speakers of English. Authors in non-English-speaking countries like India, Poland or Brazil may, in certain phases of their own career, find it more advantageous to publish valuable contributions in their own language, perhaps not bothering to generate an English translation. In this way important contributions to mankind's global intellectual development may never make it onto centre stage. The ASFE has the potential to create a strange situation for the planetary community, in which a large global audience watches developments on the Anglo-Saxon stage, in an overcrowded theatre as it were, while many brilliant actors sit in the audience, some of them thinking to themselves: "I could have acted that part much better! Anyway -- too late now." Many potential authors of valuable contributions may also be discouraged by the inevitable time lag and general organizational friction involved in actually publishing in English: Who in India, Poland or Brazil will participate in a debate, if they witness Anglo-Saxon authors constantly rushing in making the same points they were just about to make (or did already make in another language)? Science and philosophy are social enterprises as well, and the Anglo-Saxon community will automatically have easier access to Anglo-Saxon media (the "means of production"). In this context, it is important to note that, the ASFE is of course a global social

phenomenon too. To give an example, in Germany there are a number of theoreticians who have repeatedly published about the issues of thinking and conscious machines -- e.g. Ansgar Beckermann, Peter Bieri, Dieter Birnbacher, Holk Cruse, Elmar Holenstein or Martin Kurthen (to name just a few) -- and not all of their papers have appeared in English. Bob Horn and his colleagues have therefore not reflected their arguments in the debate as mapped. In short, if Argumentation Maps are to become a global tool of academic teaching -- as I think they should -- they may actually cause a worldwide strengthening of the ASFE. However, there is an easy solution to this problem: Local editorial committees should be founded, which are responsible for certain linguistic communities like German, or the many different Indian languages, or Portuguese, or Polish. Maps can be continuously updated by such local editorial committees.

Another way in which relevant information could actually be filtered out by this kind of project is of a structural nature: Not all varieties of epistemic progress can be represented as argumentative progress. Call this the "Analytical Scholasticism Effect" (ASE). Being an analytical philosopher myself, I, of course, have a tendency to think that scientific and philosophical work should proceed through rational arguments and by assessing consistent, clearly formulated theses. However, if one steps back and takes a look at it, mankind's intellectual development unfolds on many parallel paths at once. People prominent in the history of ideas can make valuable contributions without actually presenting *any* arguments. They may point out a historical context or the etymological and semantic roots of certain concepts, e.g., "representation" or "consciousness", in which we are highly interested today, thereby preventing current researchers from reinventing wheels or running into the traps that highly intelligent people already encountered a couple of centuries ago. A valuable contribution can also consist in shattering some intuition guiding research, like the ancient and recurring "mind's eye" view of mental representation (modelled after visual attention, as involving distal objects, a homunculus, and an "arrow of intentionality" pointing at a mental, i.e., non-physical, object). Progress can also arise from reviving an old intuition or metaphor in a new context, e.g., the concept of "embodiment" for robotics. Throwing new and playful metaphors into an ongoing debate can be highly relevant, without in any way being argumentative. Just think of Dennett's "intuition pumps", "skyhooks", or semi-theoretical concepts like "Cartesian materialism" or "heterophenomenology". These concepts can hardly count as arguments, but they probably have prevented a large number of bad arguments from ever being published. They certainly are part of the debate and part of the progress being made -- but it would be awfully hard to "map" them. Let me mention another, final example: epistemic progress through systematic facilitation of interdisciplinary cooperation. Obviously, this aspect possesses maximal relevance for the mind sciences, for issues like AI or consciousness. An important contribution can consist in developing low-resolution conceptual tools or thought experiments, which allow empirical researchers to grasp difficult philosophical issues more readily. Hardly any physicist or biologist knows much about the theory of science -- which certainly is a deplorable state of affairs. However, almost all of them have heard about Popper's principle of falsification, i.e., that every rational theory must be able to specify the conditions under which it will count as refuted. This, clearly, is non-argumentative progress. Frank Jackson's Mary thought-experiment or John Searle's Chinese Room

(which is actually mapped on map 4 in the current project, nicely, but arguably too extensively) have done a comparable job for the interdisciplinary community in the mind sciences -- with the argumentative substance behind them not being the major cause of the enormous progress in interdisciplinary communication caused by them. Just think about the progress that was achieved in philosophy of mind by starting to import empirical constraints, e.g., from clinical neuropsychology or from connectionist models of the representational dynamics going on in real-world brains. Simply by informing the philosophical community about the existence of phenomena like blindsight, Anton's syndrome, or the intricacies of colour vision, non-argumentative progress has been made, again probably preventing many futile debates from even taking place. Interdisciplinary information flow generates the growth of knowledge, but it is hard to integrate it into argumentation maps. It must be an object of the best of academic teaching, though. This problem is subtle and multi-faceted, but it should be possible to deal with it. Maybe "virtual arguments" or "data-blocks" could be introduced into the maps, using boxes reporting such things as "During the mid-eighties, in theoretical neuropsychology, it became more and more obvious that X, leading the philosophical debate to shift to Y". One could also imagine extensive "historical context" data blocks, "semantic-history" data-blocks, "Danger! Potential intuitive fallacy lurking!" signs, or thematically focussed bibliographical supplements.

Another potential danger is bias through sponsors. Call this the "Corruption and Epochal Madness Effect" (CEME). The most urgent project I would like to see realized myself would be a set of maps called "The philosophical discussion of the mind-body problem in the second half of the 20th century -- from U.T. Place to Jaegwon Kim". However, I hear that now theoretical biology and consciousness are most likely to follow in terms of future topics to be covered. As such projects develop, and one certainly hopes that they will have a bright future, additional sources of financing will have to be found. Obviously, some of these sponsors will have personal interests and will like to see a certain depiction of the debate, more or less subtly shaped in accordance with their own ideologies or career goals. If one of the topics of the next project actually is consciousness, a whole set of such dangers are lurking in the background. The current scientific and philosophical discussion of consciousness has a particularly low signal-to-noise-ratio: there are many shamelessly self-promoting individuals playing the media, there are politicians, Mafiosi and clowns, and a lot of people who have some sort of an "agenda" to push -- be it some pseudo-spiritual world-view or a highly specific aspect of the problem. For the topic of consciousness, as opposed to intelligent machines, there is the additional problem of "epochal madness": theories of consciousness have existed for a number of millennia, and it would be more than easy to reinvent wheels, be historically blind and, in particular, to overestimate the systematic relevance of contributions made in the last 30 years or so. A mapping project for consciousness would be much more difficult than one for thinking machines. It would have to include early Asian theories of consciousness as well as Greek precursor concepts to the Latin "conscientia", like syneidesis. It would have to discuss Cicero and Seneca as well as Thomas Aquinas, John Locke as well as Franz Brentano -- and could not simply start with the recent question of what happens when Swampman travels to Inverted Earth (e.g., Tye, 1998). German Bewusstseinsphilosophie would take an extra set of at least ten posters -- just think about

the many different concepts of consciousness a single author like Hegel developed (for historical references see Diemer, 1970; Metzinger & Schumacher, 1999). One example of this danger is the box on "Personhood: Historical Background" on Map 1 of the Thinking Machines project, which is clearly useless and absolutely inadequate. Again, this is no fundamental problem. Carefully chosen scientific advisory boards can keep out lobbyists, and epochal madness can be prevented by just the same strategy. Probably, a strict and anonymous reviewing process as practised by leading journals would help a lot in this respect.

Finally, there are some issues of minor importance. On testing these posters out with some German students, I found that they were underwhelmed by the graphic design -- something that would not have occurred to me. However, for a generation grown up with fast-cut music videos, 3D computer games, etc., posters like these may just be a little too "vanilla". It might therefore be a good idea to have a second, more expensive version with really flashy graphics, colour photos, etc., at a higher price. And of course it would be superb if academic institutions could subscribe to constantly updated multimedia versions of these posters, enabling them to download sound files, videos and additional materials from the web. Different users in different countries, with varying access to the Internet, and belonging to different generations, may want to be able to buy different versions of these posters. This certainly is not an urgent problem, but diversification of the product palette and progressive enrichment of multimedia aspects may be future goals worth considering for Bob Horn and his team. What has to happen, I think, is that -- after this highly laudable first initiative has been successfully sustained over a number of years -- the scientific community and students themselves start taking over this project by continuously updating it according to their own needs.

## Acknowledgements

I wish to thank Kevin Korb for help with the English version of this text.

## References

- Diemer, A. (1971). Bewusstsein. In J. Ritter (ed), *Historisches Wörterbuch der Philosophie, Band 1*. Basel und Stuttgart: Schwabe & Co. Verlag.
- Metzinger, T. & Schumacher, R. (1999). Bewusstsein. In H.-J. Sandkuehler (Ed.), *Enzyklopädie der Philosophie*. Hamburg: Meiner.
- Tye, M. (1998). Inverted Earth, Swampman, and representationism. In J. Tomberlin (Ed.), *Philosophical perspectives, Vol. 12: Language, mind, and ontology*. Malden, MA / Oxford, UK: Blackwell.

