

on the true state of the world). Sharing perspectives, then, improves information, which in turn might avoid or mitigate belief polarization—especially in groups that value conformity to neighboring members higher than “information obtained from the world.” Based on computer simulation, such groups are shown to depolarize, provided their learning dynamic registers a net gain through *trading perspectives*.

Sebastian Schwark (Berlin) gave hands-on examples of issues modern societies polarize over (e.g., airport runways, molecular research, carbon reduction), and outlined standard ways for political consultants to deal with them. At policy level, attitudes vary strongly with distance to one’s back yard. Some groups may defy all attitude change measures (financial compensation included) *until* some policy is “put on the ground,” upon which attitudes seem to change.

Presenting joint work in epistemic game theory, Eric Pacuit (Tilburg) and Olivier Roy (Munich) offered *choice rules*—e.g., avoid dominated strategies—as a normative source for action. They outlined conditions under which the rules fail to affect players receiving “differentially good” new information, including strategic information. For instance, weakly dominated strategies may not simply count as “deleted” upon assuming it to be common knowledge that only admissible strategies are played.

Tim Kenyon (Waterloo) discussed empirical studies of groups that reach a state of *false polarization*. Here, because of biases, subgroups tend to overestimate their comparative distance apart, unless group members are instructed to consider the strengths and weaknesses of the other position (aka. ‘counterfactual metacognition’). This being a promising exception, few (if any) reliable depolarization strategies are known, fewer yet remain self-administrable with the right probabilities. Studies seem to support the claim that—rather than improve *self*-discernment—teaching the bias literature (e.g., to undergraduates) merely increases the number of bias ascriptions to third parties.

Undercutting rather than tackling polarization, Mark Colyvan (Sydney) presented a *consensus guaranteed-strategy* for deliberating groups. Here, each agent assigns a weight to the priorities that govern a decision (between several goods or actions). Moreover, they weigh the degree to which they respect each other’s priority-weightings. In the limit—or so a theorem states—, iterated matrix multiplication of updated priority-weights with respect-weights will terminate in consensus. Rather desirably, deliberate initial over-

underbidding is discouraged insofar as it leads to lower respect-weights.

Presentations remain available at the workshop [website](#). Principal investigators are Vincent F. Hendricks (Copenhagen) and Erik J. Olsson (Lund). The next Copenhagen Lund Workshop in Social Epistemology takes place 9–10 December at Lund University.

FRANK ZENKER

Department of Philosophy, Lund University

### Philosophy and Theory of Artificial Intelligence, 3–4 October

The programmatic call for the conference said: “The theory and philosophy of artificial intelligence has come to a crucial point where the agenda for the forthcoming years is in the air—this conference will try to set that agenda and to gather many of the key players.” Gathering many of the key players it certainly did, with most of the ‘who is who’ present (1/3 of participants came to Greece from outside Europe). We had 52 speakers at the event, after a rigorous double blind reviewing procedure—see [here](#) for the program and abstracts. The conference at the leafy Anatolia campus in Thessaloniki and balmy autumn weather provided a pleasant setting with plenty of space for informal interaction, which is often the most productive part of conferences.

As for setting the agenda, at the outset of the conference we had Hubert Dreyfus warning us of the ‘first step fallacy’ for progress in AI (we have made the first step successfully and therefore all the next steps are straight ahead and feasible), while in the last keynote talk Jim Moor warned of a ‘last step fallacy’ in the area of computer ethics (we do not know whether we will ever have the last step of fully responsible artificial agents, therefore we should not make the first step towards robotic ethics). Perhaps one could use these points to characterize the agenda in the field: one camp says that certain mental properties are necessary for intelligence and then discuss how and if these can be achieved in artificial systems, while another camp asks how we can get ahead towards more intelligent behavior without directly aiming for systems that have ‘mental properties’. For the first camp, AI and Cognitive Science are just two sides of the same coin, for the second, the two are only loosely connected.

Much of the discussion at our meeting was in the

first (CogSci) camp, especially on what should come after the demise of cognition as computational symbol manipulation. Aspects of non-classical Cognitive Science proposed included embodiment and morphology (Pfeifer, Gomila, Ziemke), (inter-)active cognition (Bickhard, O'Regan), dynamic systems (Dreyfus) or a new understanding of consciousness (O'Regan, Chrisley) and of meaning (Bishop, Bringsjord). (Of course, I mention only a few names—in any case, organizers see little of their own conferences.) Within that first camp, there was also a considerable current of 'classical CogSci' approaches through computational symbol manipulation or of attempts to combine both in a 'dual-theory' (e.g., Bringsjord, Bach, Gomila, Milkowski). As for the second camp of 'nonCogSci-AI', this can be seen in Pfeifer's (Brooks' style) embodiment, Slovic's 'virtual machines', Scheutz' emphasis on capabilities or in Bostrom's discussion of the consequences of upcoming machine superintelligence—and of course this camp is dominant in technical AI meetings. Finally, there are fundamental issues on the theory of computation that are central for many (C. Smith, Shagrir, Bokulich and others) and there is a move towards respectability for a discussion of 'singularity' (Dreyfus, Bostrom, Sandberg, Yampolskiy etc.). The urgency to discuss the ethics and societal relevance of AI is gaining ground—whether or not AI is ultimately 'possible'.

Overall, the theory and philosophy of AI has set itself free from the single focus on the criticism of computational symbol manipulation; it has moved towards a new Cognitive Science and, in some quarters, a less intimate link with Cognitive Science overall. These developments support a more constructive cooperation with those who do 'the real work'—but also face the real basic problems.

The conference took place at Anatolia College/ACT (Thessaloniki) and was sponsored by EUCogII, Oxford-FutureTech, AAAI, ACM-SIGART, IACAP, ECCAI. The invited papers of PT-AI 2011 will be published in two special volumes of the journal *Minds and Machines* 2012 and the section papers as a book with Springer Publishers in the new 'SAPERE' series in 2012. We have asked the participants for feedback on the event—and if that is positive (which seems likely) we will have more meetings on the "Philosophy and Theory of AI" in the future. You are welcome!

For more information see [here](#).

VINCENT C. MÜLLER

Anatolia College/ACT &  
Faculty of Philosophy, University of Oxford

## **Evolving Knowledge in Theory and Applications, 4 October**

In recent years, intelligent agents in the contexts of open environments and multi-agent systems have become a leading paradigm in AI. Acting successfully in such environments that are uncertain, only partially accessible, and dynamic, requires sophisticated knowledge representation and reasoning techniques for the modelling of the epistemic state of the agent. In particular, in evolving environments, the agent must continuously react to new observations and to any unforeseen changes that occur. Her epistemic state must undergo corresponding changes to provide the agent with a suitable world view at any time. Thus, modern knowledge representation methods have to deal with the evolution of knowledge and belief, due to uncertain or incomplete information, or to changes in the environment.

The workshop *Evolving Knowledge in Theory and Applications* was held on October 4, 2011, in Berlin, Germany, co-located with the 34th Annual German Conference on AI (KI-2011). It was the 3rd Workshop on "Dynamics of Knowledge and Belief" (DKB-2011) organized by the Special Interest Group on Knowledge Representation and Reasoning of the Gesellschaft für Informatik (*GI-Fachgruppe Wissensrepräsentation und Schließen*).

The particular focus of the workshop was on any topics of knowledge representation and reasoning that address the epistemic modelling of agents in open environments, and in particular on processes concerning evolving knowledge and belief, both in theory and in applications. The workshop started with a session on modelling and reasoning in probabilistic approaches.

In his talk *On Prototypical Indifference and Lifted Inference in Relational Probabilistic Conditional Logic*, Matthias Thimm investigated the complexity of probabilistic reasoning in a relational setting. Based on the notion of prototypical indifference he showed that lifted inference is no longer exponential in the number of domain elements when all predicates are unary, but is still infeasible for the general case.

Markov logic is a formalism generalising both first-