CrossMark

GIRL special issue introduction

Justine Jacot[1] · **Philip Pärnamets[1,2]**

1 A historical and methodological overview of the conference series

The Lund conference series on Games, Interaction, Reasoning, Learning and Semantics began in 2012. Two conferences were organized in 2012 and 2013, under the denomination of *Lund Conference on Games, Interactive Rationality and Learning* (GIRL@Lund), before it changed for the third, to become the *Lund Conference on Games, Interaction, Reasoning, Learning and Semantics* (GIRLS@Lund) in 2014, when it was also decided that the series would become bi-annual, and resume in 2016.

A brief history of the conference series will illustrate its purpose and how it evolved over time, as well as serve as a first thematic introduction to the papers in this special issue. The idea of starting the conference series arose from discussions between research groups at the Department of Philosophy at the University of Lund and the Center for Formal Epistemology at Carnegie-Mellon University. The first call almost took the form of a manifesto. It insisted that the methods of formal philosophy, with their increasing reliance on simulations and empirical tests, brought the field closer to computer science, cognitive science and quantitative social sciences, and that the time was ripe for showcasing the cross-disciplinary potential inherent in this development. The first GIRL conference called for contributions in the areas of learning-theoretic models of inquiry, network-theoretic

✉ Justine Jacot
justine.jacot@fil.lu.se

Philip Pärnamets
Philip.Parnamets@lucs.lu.se; philip.parnamets@ki.se

[1] Department of Philosophy and Cognitive Science, LUX - Lund University, Box 192, 221 00 Lund, Sweden

[2] Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, 171 65 Solna, Sweden

approaches to scientific communities, agent-based models of reasoning and deliberation, and game-theoretic evolutionary models of communication. We chose those topics because of the 'in-house' expertise at Carnegie-Mellon and Lund, but also because they could attract contributions from researchers in philosophy, cognitive science, linguistics, economics, and artificial intelligence. Our hope was to promote the investigation of new areas where game- and learning-theoretic models could lead to new results, and suggest innovative approaches to topics both old and new.

The call for the second conference was framed to reflect discussions with the participants in the first conference. It focused on models in which equilibrium behavior emerges from the interaction of non-reflective or severely resource-bounded agents, and in which problem-solving is achieved by resource-bounded reasoners. It also, encouraged approaches that shared a similar outlook on formal models, influenced by ecological approaches to rationality and the evolutionary interpretation of game theory, as presented in the first conference. As such, the conference series stepped even further from the bounds of traditional philosophical inquiry and to the realm of a philosophically minded cognitive science meeting. The second meeting repeated the success of the first, and bred the atmosphere of intense scholarly debate and inter-field discussions that had been its goal. For its third and fourth meetings, the conference series evolved again, becoming GIRLS—adding semantics to its core topics.

In retrospect, one could say that the history of the then-GIRL, now-GIRLS conference series illustrates the type of process of which the conference series promotes the study. Indeed, the GIRL series evolved into the GIRLS series from the continued interaction of bounded-rationality academics. What had begun with a mixed bag—or perhaps more appropriately, a *smorgasbord*—of topics, that *could* be approached with models sharing a family resemblance, evolved to reveal a deeper unity, with common themes, and methods, and influences. The papers selected for this special issue were chosen to reflect the product of this evolution, its underlying themes, and the unity of its naturalistic outlook on formal methods.

2 Overview of the papers

The seven papers featured in this special issue are based on work that has been presented in a GIRL/GIRLS conference, but none of them has been presented as such. Rather, they are all the continuation and evolution of conference presentations, that benefited in one way or another from exchanges that occurred during the conferences, and for many of them, afterward. The papers can be grouped into three themes, and provide a snapshot of the varied fauna of ideas found at the conference. The first theme concerns rationality: bounded, unbounded and ecological (Cowden, Rich, Van Rooij). The second theme addresses the emergence of meaning, linguistics and communication (Barrett, Bruner, Gärdenfors, Spychalska). The third theme is transversal and addresses methodological concerns in terms of signaling games (Barrett, Bruner, and Cowden to a certain extent).

2.1 Rationality: bounded, unbounded and ecological

Van Rooij, Wright, Kwisthout & Warenham: Rational analysis, intractability, and the prospects of 'as if'-explanations

This contribution is a critical discussion of rational analysis models in Bayesian cognitive science. Rational analysis, initially proposed by Anderson (Anderson 1990, 1991), constrains functional theories of cognition by assuming a *principle of rationality*, namely that a cognitive system optimizes the behavior of an organism in response to its environment. One issue of the approach is the computational intractability of Bayesian rationality, which is usually dismissed by arguing that rational analysis rationalizes behavior on the assumption that human agents behave 'as if' they were perfect Bayesian agents, and nothing more. The success of this strategy hinges crucially on what 'as if' can be taken to mean, and this issue has been hotly debated in the recent years (Jones and Love 2011; Vilares and Kording 2011; Bowers and Davis 2012).

Van Rooij, Wright, Kwisthout & Warenham address the question of intractability, by showing that tractable rational analysis is possible, even when the cognitive function analyzed is in general intractable. A function f may indeed be intractable when its inputs are taken in a given domain D, but tractable if the domain is restricted to some $D' \subset D$. Under the assumption that cognitive functions are amenable to rational analysis under 'normal' conditions, and that these conditions coincide with sub-domains where Bayesian rationality is tractable, one can vindicate Bayesian rationality without the threat of intractability. The authors also compare their strategy to others proposed in the literature, and argue that those attempts are ultimately unsuccessful.

Cownden, Eriksson & Strimling: The implication of learning across perceptually and strategically distinct situations

Mathematical game theory is, generally, the theory of strategic interaction, and more narrowly, the theory of equilibrium selection. Initially developed by mathematicians and economists (Neumann and Morgenstern 1944; Nash 1950), it incorporates strong assumptions about players' rationality. These rationality assumptions can be dispensed with altogether, without loss of explanatory power, for instance in the evolutionary interpretation of game theory, developed by biologists (Maynard Smith and Price 1973; Maynard Smith 1982; Hofbauer and Sigmund 2003). Under this interpretation, strategies are viewed as hard-coded behavior, and the distribution of a strategy in a population depends on fitness and environmental pressure, rather than on strategic reasoning.

There are, however, agents that fall in between the hyper-rational agents of classical game theory, and the infra-rational agents of evolutionary game theory, and Cownden, Eriksson and Strimling present a model of such a population. It comprises agents that are initially naive about their environment, which consists of a variety of games. These agents are randomly paired and randomly assigned to play a game from the environment. Furthermore, their ability to discriminate between games with very different strategic structure (e.g., between a Stag Hunt and a Prisoner's Dilemma) depends on their perceptual capabilities. The authors present an implementation of these agents in artificial neural networks, and investigate a variety of scenarios, ranging from those where Stag Hunts and Prisoner's Dilemmas are indistinguishable, to those where they

are fully distinguishable. One of their main theoretical contributions is to show that, under conditions of perceptual uncertainty, all-or-nothing cooperative behavior will emerge depending on the relative proportion of Stag Hunt to Prisoner's Dilemma games. The authors also consider cases where a population trained in a given environment is transported to another, and discuss the relations between their simulation-based approach and previous theoretical results.

Rich: Comparing the axiomatic and ecological approaches to rationality: fundamental agreement theorems in SCOP

In the past century, psychologists and behavioral economists have empirically tested traditional axiomatic theories of rationality, and found some evidence for deviations from predictions based on them (Tversky and Kahneman 1974; Kahneman et al. 1991). Subsequently, some theorists have argued that we should move away from an *a priori* axiomatic conception of rationality, and endorse an ecological view, according to which to be *rational*, for a given organism, is to perform well in a given real-world context. Where axiomatic approaches tackle 'generic' problems (decision-making, belief update and belief revision, and logical inference, among others), ecological approaches focus on computable heuristics for specific problems, and evaluate them in terms of their accuracy, speed, and efficiency.

Rich's paper provides a vantage point, from which it can be determined whether or not there is a trade-off between generality, on the one hand, and context-dependency and adaptability, on the other. She first expresses the normative requirements of axiomatic theories and ecological ones in a common formalism, the *state-context property* (SCOP) formalism (Gabora and Aerts 2002). Then, she shows that compliance with a given axiomatic basis in a given context, and best performance in terms of accuracy, speed and efficiency in that context, are not exclusive. In other words, given the proposed modeling of context, the gulf between axiomatic and ecological approaches to rationality appears to narrow.

2.2 Meaning, language and communication

Gärdenfors: Levels of communication and lexical semantics

Gärdenfors' paper is a contribution to lexical semantics, and suggests a theoretical basis for the distinction between parts of speech, or word classes, such as nouns, verbs, and adjectives. Gärdenfors targets the view of semantics inherited from Frege and Tarski, which dissociates the meaning of words from the word classes, as well as classical cognitive semantics (Rosch 1973; Langacker 1987), in which meaning is assumed to be in the head of the speaker. On the Frege and Tarski view, syntactic (logical) form is distinct from surface form. The analysis of the surface form does not require more than the resources of a first-order language, that includes only quantifiers, Boolean operators, individual terms (constants and variables) and predicates. Non-logical terms are interpreted in a uniform manner, i.e., by mapping them to a domain of objects (for singular terms) or to the n-th Cartesian product of the domain (for predicates). Thus, 'surface' classes such as *verbs* and *adjectives* are reduced to the same 'logical' class, namely predicates, and differ only in arity. The distinction between those word classes in natural language thus appear to be merely accidental.

Gärdenfors' key contribution is thus to argue that not only the meaning of words, but also word classes themselves, arise from the communicative interactions between language users. Language, on this view, is not only *not* in the heads of the speakers, in the sense of meanings, but also its form arises due to similar coordinative mechanisms. To support his claim, Gärdenfors relies on the game-theoretic notions of fixpoints in communication games and on an understanding that at its basics, communication is for solving problems of actions, i.e., coordination (see also Gärdenfors (2014)). A division of communication into a hierarchical set of levels: (1) Instruction; (2) Coordination of common ground; and, (3) Coordination of meanings, is proposed. Gärdenfors uses these levels to show how resolving coordination problems within them gives rise to and necessitates word classes. As a consequence, word classes need not be identified by their syntactic roles, but should rather be described in cognitive and communicative terms. Furthermore, the existence of the same word classes in different languages suggests universal patterns of cognition and communication.

Spychalska: At least not false, at most possible: between truth and assertibility of superlative quantifiers

Spychalska's paper is a contribution to the logical analysis of grammatical differences, that remains more conservative than Gärdenfors' paper relative to the Frege-Tarski semantic tradition. She examines the linguistic differences between the superlative quantifiers 'at most'/'at least' and the comparative quantifiers 'fewer than'/'more than', and argues that the standard theory of generalized quantifiers cannot account for those differences.

Spychalska proposes to apply the distinction between *assertibility* and *truth-conditions* and argues that the assertibility conditions of superlative quantifiers are modal in nature and are different from the assertibility conditions of comparative quantifiers. She uses quantified epistemic logic to formalize those modal assertibility conditions, and offers a discussion of the interaction between the assertibility conditions and the minimal semantic meaning. In particular, she shows that assertibility conditions affect the process of inference evaluation, and that the epistemic aspect is a part of the assertibility conditions of superlative quantifiers, but not of comparative quantifiers, and that it does not enter the quantifiers' truth-conditions. Truth-conditions of superlative and comparative quantifiers are equivalent and determine the minimal meaning, while assertibility conditions constitute the enriched meaning. Both truth conditions are embedded in the assertibility conditions and can be accessed by the cancellation of the epistemic component.

Bruner, O'Connor, Rubin and Huttegger: David Lewis in the lab: experimental results on the emergence of meaning

The paper co-authored by Bruner, O'Connor, Rubin and Huttegger presents an experimental approach to the investigation of linguistic conventions, and in particular their possible emergence in a society without explicit agreement. More precisely, they investigate the following question: *can learning from experience lead players to arrive at a signaling system of a Lewis signaling game?*

They first derive predictions from their theoretical framework, which integrates Lewis' 2-players *Sender-Receiver* signaling games [Lewis (1969) later developed by Skyrms (2010)], a 2×2 variant (with two possible states, and two possible messages)

and a 3×3 variant (with three states and three messages) in an evolutionary setting. The predictions are based on the *replicator dynamics*, a model of evolutionary change pioneered by Taylor and Jonker (1978). The replicator dynamics are chosen because the outcomes they yield are close to the outcomes of various learning models. Hence, models of replicator dynamics can be used to predict how agents will learn in the lab. The empirical set-up adopts the methodology of experimental economics, where participants are offered a reward for playing some strategic games, and where the experiment is designed to be largely context-free.

The empirical results seem to validate the theoretical predictions of signaling games. Specifically, when there is an equal (unbiased) initial probability for a signal being observed, subjects tend to converge towards a signaling system; whereas when the probability is unequal (biased), a sub-optimal pooling equilibrium can emerge, and it increases as Nature is more biased. Since the empirical results were somewhat more difficult to interpret for the 3×3 variant, the authors present a simulation designed to mimic the experimental set-up, that allows them to conclude that learning to signal takes a longer time when the game is more complicated. Their overall conclusion is suggesting that, since players reach coordinating strategies in every case, meaning can emerge naturally through repeated interaction.

2.3 Signaling games and rule following

Barrett: The evolution, appropriation, and composition of rules

Barrett's paper proposes a model for the evolution of a simple rule-following behavior based on sender-predictor games a la Skyrms-Lewis (Lewis 1969; Skyrms 2004, 2010), and further suggests that the evolution of more general complex rule-following might adopt the same pattern. In a simple sender-predictor game, there are two interacting agents. One is the sender who observes a state of nature and the second is a receiver who makes a prediction (along with a corresponding action) about nature based on the signal. Using a simple, bounded, reinforcement rule, near perfect coordination can be achieved, albeit gradually. Importantly, while it is natural to think of the agents as being separated, nothing in the form of the game prohibits an interpretation where the sender and the predictor are interacting parts of the same cognitive system.

Indeed, Barrett encourages such an interpretation and proposes that sender-predictor games form a simple model for how an agent can evolve a representational system for incoming stimuli *together* with a predictive rule relying on those representations. Barrett demonstrates the strength of the proposed framework by showing how interacting agents can learn complex truth-functional rules, such as NAND, as well as how a transitive ordering rule can similarly evolve. The last case is illustrated with reference to a previously reported experiment (Bond et al. 2003) on jays, which suggests that birds make transitive inferences based on prior experience. The author shows that not only the birds did that, but were also constructing a full representation of the environment based on partial information, then appropriated a previously acquired rule to make transitive inferences on the basis of their full representation. Finally, he shows that once the agents have learned one rule, changing the state of nature such that another rule relating nature to signals is more optimal, for example OR instead of NAND,

will remap quicker by appropriating the existing rules compared to a sender-predictor game starting 'from scratch'. In sum, Barrett argues that evolving a complex rule by the composition of basic rules may be more efficient than the direct evolution of the complex rule, at least in a (cognitive) system whose learning dynamics embody signal-predictor games.

3 Afterword and acknowledgments

The GIRL conference series was organized four times from 2012 to 2016, and took place in the department of Philosophy at the University of Lund. In 2012 and 2013, it was organized by Emmanuel Genot, researcher in theoretical philosophy; Justine Jacot, PhD candidate in theoretical philosophy; and Philip Pärnamets, then PhD candidate in cognitive science, now post-doc in psychology at Karolinska Institutet. The organizing committee was adjoined two other members in 2014: Erik Mohlin, then post-doctoral fellow in economics at the University of Oxford, now assistant senior lecturer at the department of Economics in the University of Lund, and Cailin O'Connor, assistant professor of philosophy at the University of California, Irvine.

In 2012, the conference was kindly funded by the department of Philosophy in Lund and the Center for Formal Epistemology in Carnegie Mellon University. In 2013, the conference was funded by Vetenskapsrådet (the Swedish research Council). In 2014, the conference was funded partially by Vetenskapsrådet and partially by the LUIQ (Lund Information Quality) research group led by Erik Olsson, professor of theoretical philosophy in Lund. In 2016, it was organized as part of the research project "Knowledge in a Digital World", a joint venture between the department of Philosophy and the department of Arts and Cultural Sciences funded by a framework grant of Vetenskapsrådet.

The editors of this volume wish to thank all the people who made this conference series possible, especially Björn Petersson and Tomas Persson, respectively former and current heads of the department of Philosophy in Lund, for their constant and kind support of the conferences; Kevin Kelly, for bringing the support and funding of the CFE in Carnegie Mellon University; Ylva von Gerber, for all her help in practical matters and her creative touch; Emmanuel Genot, for his contribution to the organization of the conferences and his help with this special issue; Erik Mohlin and Cailin O'Connor who, besides being invaluable colleagues, allowed to broaden the fields reached by these conferences; the many speakers who brought so different and challenging topics to be discussed; last but not least, the audience who made each talk an occasion for vivid discussion and exciting exchange of ideas. We would also like to say how grateful we are to Wiebe van der Hoek, Editor-in-Chief of Synthese, who invited us to edit this volume, and Esther Rentmeester, Journal Production Manager, for their patient help and continual support.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(03), 471–485.

Bond, A. B., Kamil, A. C., & Balda, R. P. (2003). Social complexity and transitive inference in corvids. *Animal Behaviour*, *65*(3), 479–487.

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414.

Gabora, L., & Aerts, D. (2002). Contextualizing concepts using a mathematical generalization of the quantum formalism. *Journal of Experimental & Theoretical Artificial Intelligence*, *14*(4), 327–358.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge: The MIT Press.

Hofbauer, J., & Sigmund, K. (2003). Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, *40*(4), 479–519.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(04), 169–188.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, *5*(1), 193–206.

Langacker, R. W. (1987). *Foundations of cognitive grammar. Volume I: Theoretical prerequisites*. Stanford: Stanford University Press.

Lewis, D. (1969). *Convention: A philosophical study* (p. 2002). Oxford: Blackwell. (Re-edition Blackwell).

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*, 15–18.

Nash, J. (1950). Non-cooperative games. *The Annals of Mathematics*, *54*(2), 286–295.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 20–350.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge: Cambridge University Press.

Skyrms, B. (2010). *Signals. Evolution, learning and information*. Oxford: Oxford University Press.

Taylor, P. D., & Jonker, L. (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, *40*, 145–156.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Vilares, I., & Kording, K. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Pinceton, NJ: Princeton University Press.