Dissertation Summary and Future Research

Julius Schönherr

Human cooperation is unique among animals in its extent and sophistication, in part because humans can reason about others' mental states; e.g. their intentions, desires, emotions and thoughts. In short, humans can engage in what is often called mindreading. With this in mind, much research about human cooperation is aptly understood as an attempt to resolve the following tension: On the one hand, mindreading is helpful to buttress cooperation; some would say it is indispensable. On the other hand, mindreading is cognitively costly, much too costly, some say, to serve as a broad foundation for human cooperation in the real world. Furthermore, the demand for cognitively sparse theories of cooperation has increased further by incipient attempts to model cooperation among artificial agents for which computational power is inherently limited. In trying to resolve this tension between cognitively sparse and rich theories of cooperation, much current research asks "How much mindreading is *minimally* required to explain cooperation among both human and artificial agents?". This is in stark contrast with earlier research that was concerned primarily with generating models that could *possibly* explain cooperation.

The broad aim of my three-part dissertation is to advance the so-called minimal theory of human cooperation. More particularly, I argue that although cooperation is unthinkable without the attribution of mental states to one's potential partners, the extent to which these attributions have been thought necessary has been overstated. In showing this, I rely (a.) on recent findings from cognitive science and psychology, (b.) on the classical philosophical method of conceptual analysis, and (c.) on tools familiar from philosophical logic to model knowledge about the mental states of cooperative partners. Each of the three parts of my dissertation will focus on one of these distinct sources.

The first, empirical, part consists of two papers. In "**What's so Special About Interaction in Social Cognition?**", I have defended the idea that core cases of cooperative interaction cannot be explained without appealing to mindreading. However, as I further argue in my paper "**Beyond Interaction**", the relevant capacities are not restricted to cooperative interactions but can be found in a wide array of human activities.

Next, in the conceptual part of my dissertation, I criticize the idea that true cooperation requires thorough knowledge of one's partners' intentions. Traditionally, philosophers have thought that, in truly cooperative acts, the intentions

of each partner must be fully transparent to all parties involved. Recently, this picture has been subjected to criticism. My paper **"Lucky Joint Action"** extends this research. In this paper, I claim that thorough knowledge of one's co-participants' intentions is not just not necessary, but can be rather harmful to successful cooperation. In a follow-up paper – **"Joint Action without Joint Intention"** – I aim to show that the same idea holds true of the intention part of joint action: Certain cases of joint action require the co-participants' intentions to be substantially disparate.

In the final paper of my dissertation – **"Coordinating Without Common Inductive Standards"** – I show that in cooperative situations it is often sufficient that an agent *not doubt* that her co-participants' intentions support the cooperative activity; positive knowledge of their intentions is, again, not required. While the philosophical literature has, at times, sporadically acknowledged that the absence of doubt on the hand and positive belief attributions on the other are different, the structural importance of this distinction has previously gone unnoticed.

In the longer term, I plan to integrate this research into a more comprehensive account of the so-called minimal theory of cooperation.

In addition to my work in social philosophy, I have an active research interest in moral philosophy quite generally, and in forgiveness more particularly. In my paper "**When Forgiveness Comes Easy**" I argue that traditional approaches to forgiveness cannot capture the idea that, sometimes, forgiving need neither involve an inner struggle nor any exercise of virtue. Currently, I'm working on a follow-up paper arguing that forgiving should not simply be characterized as a way to overcome one's moral anger, but, rather, as the *endorsement* one's lack of anger.