Self-awareness, willpower, and long-term intentional action

Fabio Paglieri

Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche
fabio.paglieri@istc.cnr.it

In this talk I will discuss the impact of having a representation of one's own internal states (metacognitive self-awareness) on the capacity to control one's behavior. In particular, I will discuss how self-awareness enables new forms of self-control, broadly defined as the capacity to deliberately refrain from following a dominant impulse (occurrent or potential) in the name of what one considers to be a superior interest (e.g. long-term goals, moral considerations, the benefit of someone else, etc.). Specific attention will be given to intertemporal self-control, that is, the capacity to persist in one's conduct over prolonged periods of time, without giving in to temptations or distractions (Paglieri & Castelfranchi 2008).

There are two main ways in which self-awareness impacts on intertemporal self-control (Castelfranchi 2012):

1. **Self-image**: by developing a more or less coherent image of who we are and/or want to be, we acquire a standard against which all our behaviours are measured, which in turn pressures our conduct for coherence.

2. **Internal negotiation**: by remembering and/or foreseeing shifts in our preferences, we acquire the capacity (and the need) to bargain with ourselves, which results in frequent attempts to self-influence our future selves into doing what the present self considers best.

The first phenomenon has been mostly studies in social psychology (Laughlin 1970; Higgins 1987; Sheldon & Kasser 1994; Bandura 1997; Swann 1999), whereas the second is a standard topic of inquiry in philosophy (Elster 1979, 2000; Parfit 1984; van Fraassen 1984, 1995; Bratman 1999) and social science (Schelling 1984; Akerlof 1991; Ainslie 2001; Read 2001; Bénabou & Tirole 2004). Both strategies raise a host of fascinating questions, some of which will be briefly discussed in this talk:

- How do self-image and coherence of conduct exactly influence each other (Baumeister et al. 1993; Miceli & Castelfranchi 2012)?
- What are the differences, if any, between self-control and willpower, and which one can be externalized, in the vein of the extended mind hypothesis (Clark & Chalmers 1998; Paglieri 2012a, b)?
• What rationality principle is best suited to capture rational intertemporal choices, such as Ulysses' strategies with the Sirens (Parfit 1984; van Fraassen 1984, 1995; Hájek 2005)?
• Is there such a thing as too much self-control (Kivetz & Simonson 2002; Kivetz & Keinan 2006)?
• What self-influence strategies people employ and why do such tricks work – or not (Ariely & Wertenbroch 2002; Della Vigna & Malmendier 2006)?

REFERENCES