Choice blindness and the non-unitary nature of the human mind

doi:10.1017/S0140525X10002591

Petter Johansson,[a] Lars Hall,[b] and Peter Gärdenfors[b]

[a]*Division of Psychology and Language Sciences, University College London, United Kingdom;* [b]*Lund University Cognitive Science, Lund University, Sweden.*

petter.johansson@lucs.lu.se **lars.hall@lucs.lu.se**
peter.gardenfors@lucs.lu.se
http://www.lucs.lu.se/petter.johansson/
http://www.lucs.lu.se/lars.hall/
http://www.lucs.lu.se/peter.gardenfors

Abstract: Experiments on choice blindness support von Hippel & Trivers's (VH&T's) conception of the mind as fundamentally divided, but they also highlight a problem for VH&T's idea of non-conscious self-deception: If I try to trick you into believing that I have a certain preference, and the best way is to also trick myself, I might actually end up having that preference, *at all levels of processing*.

The classic paradox of self-deception is how the self can be both deceiver and deceived. Von Hippel & Trivers (VH&T) solve this

conundrum by appealing to the separation of implicit and explicit processes in the mind; I cannot knowingly deceive myself, but the non-conscious part of my mind can "deceive" me by pursuing goals that are contradictory to my consciously stated ambitions. VH&T identify and draw support from three different areas of research: explicit versus implicit memory, explicit versus implicit attitudes, and controlled versus automatic processes. None of these processes are inherently self-deceptive. Instead, as VH&T write: "These mental dualisms do not themselves involve self-deception, but each of them plays an important role in enabling self-deception" (sect. 4, para. 1).

We suggest adding a fourth set of related studies: work on choice blindness – that is, the failure to detect mismatches between a choice made and the outcome received (Johansson et al., 2005). Choice blindness is an experimental paradigm inspired by techniques from the domain of close-up card magic, which permits a surreptitious manipulation of the relationship between choice and outcome that the participants experience. The participants in Johansson et al. (2005) were asked to choose which of two pair-wise presented faces they found most attractive. Immediately after, they were also asked to describe the reasons for their choice. Unknown to the participants, on certain trials, a double-card ploy was used to covertly exchange one face for the other. Thus, on these trials, the outcome of the choice became the opposite of what they intended. Remarkably, in the great majority of trials, the participants were blind to the mismatch between choice and outcome, while nevertheless being able to offer elaborate reasons for their choices. The two classes of reports were analysed on a number of different dimensions, such as the level of effort, emotionality, specificity, and certainty expressed, but no substantial differences between manipulated and non-manipulated reports were found (Johansson et al. 2006). The lack of differentiation between reasons given for an actual and a manipulated choice shows that there is probably an element of confabulation in "truthful" reporting as well. In addition to faces and abstract patterns (Hall & Johansson 2008), choice blindness has been demonstrated for taste and smell (Hall et al. 2010 in press), as well as for moral and political opinion (Hall et al., in preparation).

Experiments on choice blindness support VH&T by providing a dramatic example of the non-unitary nature of the mind; we may have far less access to the reasons for our actions than we think we do. But experiments on choice blindness also highlight a possible problem lurking in VH&T's conception of self-deception. Is it really possible to maintain two separate sets of conscious and non-conscious goals as a technique to deceive oneself in order to better deceive someone else? For example, in one version of the experiment described earlier, the participants had to choose between the same pairs of faces a second time, as well as separately rate all the faces at the end of the experiment. This procedure revealed that the manipulation induced a pronounced, but to the participants unknown, preference change, because they came to prefer the originally non-preferred face in subsequent choices, as well as rate the face they were led to believe they liked higher than the one they thought they rejected (Hall et al., in preparation). This result is of course in line with a long tradition of studies showing the constructive nature of preferences, i.e. that we come to like what we think we like (see Ariely & Norton 2008; Bem 1967; Festinger 1957; Lichtenstein & Slovic 2006).

The crucial point is that if it is possible to get people to reverse their initial preferences by making them publicly endorse an outcome they believe they prefer, then using self-deception as a means to deceive others might result in fundamental changes to the self as well. If I try to trick you into believing that I prefer *a* over *b*, and the best way to do that is to also trick myself into believing that I prefer *a* over *b*, I might actually end up preferring *a* over *b*, *at all levels of processing*. In such a case, it would be the conscious parts of the self that makes the unconscious parts change, and in a process more akin to self-persuasion than self-deception. The apparent ease with which the participants in choice blindness experiments confabulate reasons in favor of a previously rejected alternative indicates that this form of self-persuasion is something that comes quite naturally to us.