# Evolving Persons and Free Will

Rüdiger Vaas, Stuttgart

## 1. Self-models and persons

„Man kann den eigenen Sinnen mißtrauen, aber nicht dem eigenen Glauben", Ludwig Wittgenstein (1953, p. 514) once remarked. „Gäbe es ein Verbum mit der Bedeutung ‚fälschlich glauben', so hätte das keine sinnvolle erste Person im Indikativ des Präsens." But what if such a first-person entity is indeed a fiction in certain respects – or if it has at least false beliefs without being able to distrust them?

Human beings are masters of deception if they want to appear superior to others and to suggest that they have everything under control (see, e.g., Fingarette 2000, Mele 2000). Such self-delusions might be advantageous, because those are the most successful liars who believe their own lies. Although it seems paradoxical at first (for he who does not tell the untruth intentionally is, strictly speaking, not a liar at all), it rests upon a much more radical self-deception which is quite useful – a systematic and continuous illusion regarding ourselves. Higher-order forms of self-consciousness, namely *I-consciousness*, are based on a feature which is called a *self-model*. This is an episodically active representational entity (e.g. a complex activation pattern in a human brain), the contents of which are properties of the system itself. It is embedded and constantly updated in a global model of the world, based on perceptions, memories, innate information etc. (Metzinger 1993). But because self-models *cannot* represent their own representations as their own representations as their own representations and so on *ad infinitum*, they are semantically transparent, i.e. on the level of their content they do not contain the information *that* they are models. Thus, such systems are not able to recognize their self-model as a self-model (Van Gulick 1988). The result is an *ego-illusion*, which is stable, coherent, and cannot be transcended on the level of conscious experience itself.

Such a self-model is not an epistemic luxury. It plays a role for the system, it is a weapon developed in the course of biological evolution. As Marcel Kinsbourne (1988, p. 249) said: „If the concept of self evolved, it did so on account of adaptive advantage, not because it reflects some objective truth. The concept of self reifies the organizing activity of a cybernetic device that incorporates its history (‚experience') into the basis for its actions. It is the construct around which are organized impressions and intentions that reach awareness."

Neuroscience suggests that mind coincides with certain brain processes (Vaas 1999) and has evolved as a goal-oriented device that implements predictive interactions between the organism and its environment. The capacity to predict the outcome of future events – necessary to orchestrate and express their movements successful active movements – seems to be the ultimate and most common of all global brain functions; consciousness and thinking can be viewed as an evolutionary internalization of movement, and a self-model is the centralization of prediction (Llinás 2001, Vaas 2001a).

Such an approach for self-models – if it is basically correct, and it cannot be argued here that it actually is – has essential implications for our understanding of what is it to be a person. Rationality, higher-order intentionality, I-consciousness, intentional stance, autonomous agency, transcendence of the presence (i.e. a concept of time), language, altruism and morality are the main criteria for characterizing persons, but they are not totally independent, and it is controversial whether they are all necessary or sufficient. Central at least is I-consciousness and some sort of autonomy – terms which are also ambiguous. However, they do not necessarily subscribe to *ego theories* of persons which hold that there are irreducible ontological, immaterial substances endowed with certain properties like free will; but they are compatible with *bundle theories* assuming that persons are based on simpler, e.g. psychological, computational or physiological processes (cf. Vaas 1996, 2001b).

## 2. Free will as a useful illusion

The *mind body problem* can be taken as a trilemma where any of the following three premises is excluded by the two others: (1) *dualism* – the mental is not the physical, (2) *mental causation* – the mental does causally influence the physical (and is affected by it), (3) *physical closure* – there is no nonphysical influence on the physical. The so-called *mystery of consciousness* consists in the explanatory gap between (1) and (2), i.e. how could matter (if at all) create mind? The *problem of free will* is the incompatibility of (2) and (3) if (2) requires (1). Here, the main opponents are: first, *libertarianism* – (3) is wrong, there are contracausal or nonphysical originations; second, *incompatibilism (determinism)* – (1) is wrong; and, third, *compatibilism* – (1) is wrong, but contradictions between (2) and (3) can be explained away.

Strong arguments (which cannot be defended here) show that the philosophical problem of free will cannot be *solved*, for this would require the reconciliation of apparently inconsistent premises; but it might be *dissolved* by eliminating one of the premises, namely the claim that there are irreducible entities like free-floating selves or Cartesian egos with the ability to act due to their own non-physical power, for this cannot avoid the dilemma of either plunging into an *infinite regress* or leading to a *mysterious causa sui*. Ultimately our reasons, beliefs and volitions are non-consciously determined – by earlier experiences, heredity, physiology or external influences – and therefore not *ultimately* up to us (Honderich 1988, Kane 2002, O'Connor 1995, Vaas 2001c, Strawson 1986, Walter 1998, Wegner 2002).

Nevertheless our misleading *conception* of being such selves with impressions of having free will has to be explained! We do conceive ourselves, at least sometimes, as being free, i.e. that we can decide between alternatives. This feeling depends on complex abilities of voluntary movement (Vaas 2001d), second-order emotions (without which we cannot act and choose in complex situations despite of rationality), a non-deprived development, non-predictability or epistemic indeterminism (we cannot know the future for certain, especially not our own future; Wittgenstein 1921, 5.1362), rationality (the ability to reflect and reason), planning (and hence higher-order thoughts, a concept of the future), higher-order volitions, and sanity.

These features are compatible with naturalism and determinism. Therefore it is not to deny a weaker form of free will. But this does not imply the existence of the kind of freedom for which Libertarianism is arguing.

Libertarians still insist that our subjective impression of freedom is a powerful argument for free will. Thus, a sceptic should be able to explain such an impression within the framework of naturalism. And this is what an evolutionary perspective might achieve.

Many zoo and field experiments as well as behavioral studies in the wild have shown that apes can respond differently according to the desires and beliefs of other individuals – rather than according only to the other's overt behavior. Hence, they probably have what Daniel Dennett (1973) called the *intentional stance*: They ascribe intentions to others and take them into consideration for their own actions (Taylor Parker et al. 1994, Whiten and Byrne 1997) – at least we do this. Evolution shaped our minds respectively our brains to cope with our complex social lives. The social environment might have been a significant selective pressure for primate intelligence (Humphrey 1976) and the rapid expansion of our ancestors' neocortex. This cortical enlargement – about a factor of four during the last five million years – is otherwise hard to explain; and there is evidence for a correlation between neocortical size and group-size or social complexity (Barton and Dunbar 1997). Since better access to food, a safer place to sleep or a higher rank in complex hierarchies normally increase the probability of producing more offspring than other group members, social intelligence pays off pretty well. The elaborated mental abilities of higher primates are conceived as the product of an evolutionary cognitive arms race leading to more and more sophisticated representational capabilities (representation of complex social relationships, higher-order intentional stance, mind-reading, primitive theory of mind).

*We are forced by our very nature to interact with other people in a fundamentally different way than to interact with, say, stones and sticks* (Strawson 1962). In cognitive psychology, there is plenty of evidence now that in attempting to make sense of other people, perceivers regularly construct and use categorical representations to simplify and streamline the person perception process (Macrae and Bodenhausen 2000). This is advantageous only in so far as it influences one's own actions.

Thus, the evolution of an intentional stance is at the center of our impression of having free will: *Ascribing intentional states to others necessarily includes ascribing volitions to them and assuming that they have the power to transfer their volitions into actions somehow*, because this is the only way to get advantages from the intentional stance at all. For, if other beings are thought to have intentions but they would be causally inert, i.e. their behavior has nothing to do with their volitions, this ascription of intentions and hence volitions simply wouldn't matter. However the intentional stance is not an irrelevant luxury. It is a powerful tool to get along with the complexity of the social world and even an anthropomorphically-conceived nonsocial world (up to highly restricted activities – e.g. in playing computer chess nowadays it is common and useful to think and act as if the computer „wants" and „plans" something). Individuals endowed with this tool are better prepared for the struggle of social life. And it is advantageous to assume the volitions of others as somehow being independent of the environment or the past – not absolutely independent of course, but in an approximate sense, because this makes it a lot easier to deal with them due to the fact that complex organisms can act (or react) quite differently in similar circumstances and quite similar in very different circumstances.

There is another reason to take a concept of volition as evolutionarily advantageous, and this is just the other side of the coin: To deal with other individuals in a complex way means also *to plan one's own actions carefully in an explicit way and evaluate their effects*. This presupposes some kind of awareness of one's own volition, hence a concept of will and self. Higher-order representations also take one's own mental states into account – not only for decisions and follow-up analyses but also as a parameter in the plans of others regarding oneself. Thus, it is reasonable or even necessary to ascribe volitions to oneself, too – because otherwise one cannot reason about the mental states of others who are presumably dealing with oneself. This makes one's own volitions explicit – and much more flexible. At least since the point from which there has been language with an inbuilt grammatical structure, distinguishing between subjects and objects, active and passive, present and future, such concepts of volition, actions and self-notions have been flourishing (Vaas 2000).

This was not only the case in contexts of cheating, however! In the course of time *co-operation* became more and more important among our early ancestors. And the existence of some form of language already implies a high degree of co-operation – spoken language would never have emerged unless most people, most of the time, followed conventional usage. But co-operation in complex, not inherited forms also presupposes an intentional stance and the capacity to ascribe volitions to others.

From this it is no big step to a notion of free will which is a powerful tool to act in consonance with or opposition to others and to establish some kind of moral responsibility – a very effective way to influence the behavior of others and justify punishments. Thus, free will even succeeded to become an entity of religious, philosophical or political theories and a postulate for jurisdiction. Of course we need not dismiss an intentional and personal stance. It is, obviously, crucial for our survival. We cannot leave our subjective standpoints, turning exclusively to an objective, perspectiveless view (cf. Nagel, 1986). We may accept that we have, ultimately, no free choice. Nevertheless, in our everyday life we think and act as if we did. Even sceptical philosophers do – or they might find themselves out of the race quickly.

## References


Barton, R. A. and Dunbar, R. I. M. 1997 „Evolution and the social brain“, in Whiten and Byrne (eds.) 1997, 240-263.

Dennett, D. 1973 „Mechanism and Responsibility“, in T. Honderich (ed.) *Essays on Freedom of Action,* London: Routledge and Kegan Paul, 157-184.

Fingarette, H. 2000 *Self-deception,* Berkeley: University of California Press.

Honderich, T. 1988 *A Theory of Determinism,* Oxford: Clarendon Press.

Humphrey, N. 1976 „The Social Function of Intellect“, in N. Humphrey 1998 *Consciousness Regained,* Oxford: Oxford University Press, 303-317.

Kane, R. (ed.) 2002 *The Oxford Handbook of Free Will,* Oxford: Oxford University Press.

Kinsbourne, M. 1988 „Integrated Field Theory of Consciousness“, in A. Marcel and E. Bisiach (eds.) *Consciousness in contemporary Science,* Oxford: Clarendon Press, 239-256.

Llinás, R. 2001 *I of the Vortex,* Cambridge: MIT Press.

Macrae, C. N. and Bodenhausen, G. V. 2000 „Social Cognition“, *Annual Reviews of Psychology,* 51, 93-120.

Mele, A. R. 2000 *Self-deception unmasked,* Princeton: Princeton University Press.

Metzinger, T. 1993 *Subjekt und Selbstmodell,* Paderborn: Schöningh.

Nagel, T. 1986 *The View from Nowhere,* New York: Oxford University Press.

O'Connor, T. (ed.) 1995 *Agents, Causes, Events,* Oxford: Oxford University Press.

Strawson, G. 1986 *Freedom and Belief,* Oxford: Clarendon Press.

Strawson, P. F. 1962 „Freedom and Resentment“, *Proceedings of the British Academy,* 48, 187-211.

Taylor Parker, S. Mitchell, R. W. and Boccia, M. L. (eds.) 1994 *Self-Awareness in Animals and Humans,* Cambridge, Cambridge University Press.

Vaas, R. 1996 „Mein Gehirn ist, also denke ich“, in C. Hubig and H. Poser (eds.), *Cognitio humana, Vol. 2,* Leipzig, 1507-1514.

Vaas, R. 1999 „Why Neural Correlates Of Consciousness Are Fine, But Not Enough“,

*Anthropology & Philosophy,* 3, 121-141.

Vaas, R. 2000 „Evolving language, I-consciousness and free will“, in J.-L. Dessalles and L. Ghadakpour (eds.) *Evolution of Language,* Paris: Ecole Nationale Supérieure des Télécommunications, 230-235.

Vaas, R. 2001a „It Binds,Therefore I Am!“, *Journal of Consciousness Studies,* 8, 4, 85-88.

Vaas, R. 2001b „Persönlichkeit und Personalität“, in *Lexikon der Neurowissenschaft, Vol. 3,* Heidelberg and Berlin: Spektrum Akademischer Verlag, 53-63.

Vaas, R. 2001c „Willensfreiheit“, in *Lexikon der Neurowissenschaft, Vol. 3,* Heidelberg and Berlin: Spektrum Akademischer Verlag, 450-459.

Vaas, R. 2001d „Das Gehirn in Aktion“, *Universitas,* 56, 924-940.

Van Gulick, R. 1988 „A functionalist Plea for Self-consciousness“, *The Philosophical Review,* XCVII, 149-181.

Walter, H. 1998 *Neurophilosophie der Willensfreiheit,* Paderborn: Schöningh.

Wegner, D. M. 2002 *The Illusion of Conscious Will,* Cambridge: MIT Press.

Wittgenstein. L. 1921 *Tractatus logico-philosophicus,* Frankfurt am Main 1984: Suhrkamp.

Wittgenstein. L. 1953 *Philosophische Untersuchungen,* Frankfurt am Main 1984: Suhrkamp.

Whiten, A. and Byrne, R. (eds.) 1997 *Machiavellian Intelligence II,* Cambridge: Cambridge University Press.