12

What's the Use of Consciousness?

How the Stab of Conscience Made Us Really Conscious

Chris D. Frith and Thomas Metzinger

Before the birth of consciousness,
When all went well,
None suffered sickness, love, or loss
None knew regret, starved hope, or heart burnings.
— Thomas Hardy, *Before Life and After* (1909)

Regret is the most bitter pain, because it is characterized
by the complete transparency of all one's guilt
— Søren Kierkegaard, *Either/Or* (1843/1992)

Abstract

The starting assumption is that consciousness (subjective experience), rather than being an epiphenomenon, has a causal role in the optimization of certain human behaviors. After briefly outlining some of the critical properties of consciousness, this chapter reviews empirical studies that demonstrate how much can be achieved in the way of action and decision making in the absence of relevant conscious experience. Thereafter, it considers, in detail, the experience of action and suggests that this has two key components: the experience of being an agent, which causes events in the world, and the belief that we could have done otherwise. Such experiences enable us to justify our behavior to ourselves and to others and, in the longer term, to create a cultural narrative about responsibility. Finally, the experience of regret is explored (i.e., the recognition that one could and should have acted otherwise). Regret is a powerful, negative emotion that is suggested to integrate group norms and preferences with those of the individual. The transparent and embodied nature of the experience of regret ensures that cultural norms become an inescapable part of the self-narrative. The conclusion is that conscious experience is necessary for optimizing flexible intrapersonal interactions and for the emergence of cumulative culture.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

Introduction

What's the use of consciousness? By asking this question, we indicate that we approach the problem of consciousness mainly via biology and evolutionary theory. Our first assumption is that the appearance and maintenance of the phenomenon of consciousness in humans and other animals implies that there is some continuing evolutionary advantage to consciousness. If we assume that consciousness evolved, then it is also reasonable to assume that some creatures, such as humans, are more conscious than others. It also follows that, at least in our world and under the laws of nature holding in it, we do not believe in the possibility of zombies, those philosophical constructs, functional isomorphs that behave exactly like humans, but in the absence of consciousness. There are some things that such zombies would not be able to do. Our task is to identify these things.

Much previous work on consciousness has concentrated on perception, especially vision (Crick and Koch 1995). The natural antidote to this biased perspective requires that we cease to focus on perception, or action, or cognition in isolation. Thus, to answer the question "What's the use of consciousness?" we need to relate consciousness research to the underlying principle that connects all three elements: the action-perception loop. If there was a formal framework capable of unifying all three aspects under a common principle, and if that framework turned out to be empirically plausible, then it would be natural to describe conscious experience by using the conceptual tools offered by it. For example, conscious experience could then be a single, generative model of reality including a model of the self as currently acting, perceiving, and thinking (Friston 2010).

If consciousness gives an advantage to humans, then it must causally enable humans to achieve more optimal behavior. What class of optimization problems does consciousness enable us to solve? What new types of action does it enable? These potential uses of consciousness are particularly relevant to the current pragmatic turn in cognitive science.

Consciousness and Its Properties

Levels and Contents of Consciousness

At this point we need to make a gesture in the direction of defining consciousness. One distinction is between *levels* of consciousness and *contents* of consciousness. Levels of consciousness relate to the distinction between being awake or asleep as well as between being a man or a mouse. Consciousness comes in degrees (e.g., of wakefulness and alertness). We can ascribe the property of "consciousness" to whole persons or biological systems, and we might distinguish such systems according to their overall level of wakefulness, the

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

presence of an orientation reaction, etc. This is sometimes called *creature consciousness* (Metzinger 1995a). Consciousness, however, can also be viewed as a property of individual states (e.g., representational states in the central nervous system of an organism). This is sometimes called *state consciousness* (Rosenthal 1986). The content of consciousness refers to what our conscious experience is *about*, what we are currently conscious *of*. This relates, for example, to the distinction between being conscious (aware) of the face in our field of view and not being conscious of that face (important in the search for the neural correlates of consciousness; see, e.g., Beck et al. 2001), while the overall level of consciousness does not differ between these two states of the subject in the experiment. Accordingly, we can look for the global neural correlate of consciousness (i.e., the set of physical or functional properties corresponding to the totality of an individual's experience) or the correlate of specific kinds of content; for useful conceptual distinctions, see (Chalmers 2000).

Properties of Conscious Experience

In terms of the contents of conscious experience, three properties are of particular relevance (for further details, see Metzinger 1995b):

1. There is a pure subjective experience, the *phenomenal content* of our mental states. These subjective states have a certain *feel*: there is something *it is like* to be in these states.
2. These phenomenal states are frequently *transparent*. We do not experience these states as representations of reality; we experience them directly *as* reality.
3. Conscious experience is always perceived as part of the current moment, whereas phenomenal experience is characterized by the subjective character of *presence*. What is present is always a whole situation or single, unified world model.

In addition, under standard conditions these states and their contents are experienced from a *first-person perspective*: they are the inner experiences of an individual and seem to be private. This last property raises well-known epistemological problems for the scientific study of consciousness (Jackson 1982; Levine 1983): How can scientific objectivity be applied to something that is subjective and only available as an individual first-person perspective (cf. Nagel 1974, 1986)? This problem reveals an interesting paradox: my conscious experience appears to be private and inaccessible to anyone else, yet it is the only aspect of my mental life to which I have seemingly direct access, about which I possess maximal certainty, and which I can potentially report to others. Perhaps the core problem in consciousness research consists in finding out what exactly a "first-person perspective" is and if it can, at least in principle, be naturalized.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

The Problem of Report

In practice the study of subjective experiences depends on the report of the person having the experience. Strictly speaking, there is no such thing as "first-person data." Scientifically we can only access first-person reports, but never the experience itself, in its subjectivity (for further discussion, see Metzinger 2010). These reports need not be verbal; some experiences may be too novel or complex for suitable words to be available. Further, *if* verbal reports are available, they are in constant danger of being "theory-contaminated"—a process by which subjective reports are influenced by the scientific or philosophical theories the subject believes to be true, by specific psychological needs, by social context, or by cultural background assumptions. For the subjective experience of agency, which is the primary focus here, we believe this point to be of particular relevance.

A Novel Proposal

We believe that conscious content may have played a decisive role in the emergence and stabilization of complex societies. This is one prime example for a function of consciousness. To ground our proposal below, we will look at a range of biological and cognitive functions for which conscious processing is not a necessary prerequisite. Then we will consider the experience of action and introduce the notion of "regret," first describing its phenomenological profile, then offering a brief representationalist analysis, and proceeding to isolate its hypothetical function. Finally, we will present a brief sketch of an argument as to why this specific kind of phenomenal experience would have been advantageous under an evolutionary perspective.

Functions for Which Consciousness May *Not* Be Necessary

There is so much empirical evidence in favor of consciousness, viewed by itself, as having little role in our behavior, that we might conclude that it is no more than an epiphenomenon. Huxley (1874) proposed that consciousness, although real and created by the brain, was an epiphenomenon with no influence on behavior:

> Consciousness…would appear to be related to the mechanism of the body simply as a collateral product of its working, and to be as completely without any power of modifying that working as the steam-whistle which accompanies the work of a locomotive engine is without influence upon its machinery.

Humans, he suggested, are "conscious automata." Even if we accept that consciousness does have a role in decision making, there are many cases where better decisions are made when people forgo conscious control (for a review, see Engel and Singer 2008). What, then, is the use of consciousness? Before

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

presenting our own speculations, we will consider and dismiss several candidate processes for which it has been proposed that consciousness is necessary.

Sensory Integration and Global Informational Access

Here is a good candidate for a function of consciousness: Our conscious experience of the world typically involves objects and actions, rather than isolated sensations and movements, and it is plausible that consciousness is necessary for integrating information (Tononi 2008) and for broadcasting information between different processing modules (Baars 1988; Dehaene and Naccache 2001). However, we are doubtful. There is increasing evidence that sensory integration happens even at the earliest stages of sensory processing (e.g. Watkins et al. 2007; Lemus et al. 2010); even high-level, cross-modal integration of symbols can occur without awareness (Faivre et al. 2014). For many activities there is clear need for "global availability" of information. But why should this global access be associated with subjective experience? Furthermore, decisions which require the integration of many sources of information seem to be better made in the absence of conscious reflection (Dijksterhuis and Nordgren 2006), perhaps because appropriate weighting of multiple sources of information is disrupted by conscious deliberation (Levine et al. 1996; Engel and Singer 2008). A similar phenomenon can be observed in the performance of highly skilled acts (Beilock et al. 2002).

Sophisticated and Flexible Top-Down Control

Many accounts suggest that consciousness is necessary for a high-level supervisory system that modulates lower-level automatic processes, especially when unexpected problems arise and when novel skills must be developed (e.g., Norman and Shallice 1986). Again, we find this to be a very plausible suggestion. However, given that there is a hierarchy of sensorimotor control (e.g., Friston 2005), this formulation requires that we specify at which level in this hierarchy consciousness emerges. While there is, as will be described below, good evidence for a role for consciousness in top-down control, we suggest that the level in the hierarchy, at which this operates, is higher than previously supposed. Control of considerable sophistication and flexibility can occur at lower, "automatic" levels. For example, it is well established that much low-level control of action can occur without awareness. This is true for hand movements (Fourneret and Jeannerod 1998) as well as for locomotion involving the whole body (Kannape et al. 2010).

Consider two examples in which monitoring and control occur at an even higher level in the absence of awareness. In a study of walking (Varraine et al. 2002), people were given arbitrary and unpracticed instructions as to how they should change their walking pace when they detected a change in the responsiveness of the treadmill upon which they were walking. Remarkably, they

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science," Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18, series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

changed their pace correctly for about six seconds before they reported detecting the change. In another study, skilled typists slowed down after they had made errors, but not after they experienced errors inserted by the experimenters, even though their verbal report showed that they were not conscious of the distinction between these types of errors (Logan and Crump 2010). Here, monitoring and control (metacognition) of the low-level process of typing occurred outside consciousness.

Emotion and Motivation

Does the motivation created by affective states, such as pleasure and pain, depend on awareness of these states? Perhaps emotions do not have to be conscious to make people act in particular ways. The following study gives an example of an unconscious emotion: Smiling faces, presented subliminally, caused thirsty people to pour and consume more drink, even though they were unaware of any change in their emotional state (Winkielman et al. 2005). It is obvious that consciousness is required for us to *talk about* an emotion, and further research is needed to identify those aspects of emotion that enable functional availability for verbal report (Metzinger 2003). However, certain kinds of conscious emotion, such as regret, do have effects on behavior (Filiz-Ozbay and Ozbay 2007). Unlike more basic emotions, such as happiness and anger, regret involves counterfactual thinking: "Things would be different, if only I had behaved differently."

Representing the Mental States of Self and Others

Do we need consciousness to account for the mental states (e.g., beliefs, perceptions and intentions) of others and of ourselves (Humphrey 1999; Graziano and Kastner 2011)? We certainly need consciousness to *talk* about our mental states, but can we take account of the mental states of others without awareness? For example, our behavior is affected, automatically, by the action goals of others (Sebanz et al. 2003) as well as by the perceptual knowledge of others (Samson et al. 2010). In Samson et al.'s study, they showed that it took people longer to report the number of targets when another person with a different viewpoint saw a different number of targets, even when this was entirely irrelevant to the task being performed. This effect was not altered when a cognitive load was applied (Qureshi et al. 2010), suggesting that the process of taking another person's knowledge into account was automatic and unconscious.

Do We Need Consciousness to Make Free and Flexible Decisions?

The ability to make free and flexible decisions is a role for consciousness that is most relevant to action. Yet ever since Benjamin Libet's classic experiment, doubt has been cast on this role (Libet et al. 1983). Results from his studies,

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

replicated more recently using fMRI (e.g., Soon et al. 2008), suggest that the awareness of initiating an action comes too late to have any causal role in the decision. From fMRI data, patterns of brain activity, occurring well before a participant reports making the decision, can be used to determine, somewhat better than chance, which action will be chosen.

One problem with these studies is that the decision (e.g., whether to move the left or the right finger) is neither taxing nor of much relevance to real life. However, the discovery of choice blindness, by Johansson et al. (2005), confirms the fragile relationship between decisions and awareness in situations of much greater ecological validity (e.g., Hall et al. 2012). In these studies, participants could be persuaded to justify a decision that they had not actually made. They seemed unaware of the decision that they had actually made.

Being in Control: The Experience of Agency

A striking paradox is revealed by the above-mentioned studies: awareness of decision making seems to have little or no role in causing decisions, yet the vivid feeling of being the author of one's own actions—the sense of *agency*—is a large component of conscious experience. Indeed, it is only because we have this clear experience of being in control that experiments like those of Libet are possible. People have no problem when asked to report the precise moment at which they made a decision. There is strong awareness of mental agency, yet, at the same time, very little awareness of bodily agency. Why should awareness of mental agency be given such salience, unless it has some function?

Since at least the time of Epicurus (Bobzien 2006), the experience of being in control of our actions, our sense of agency, is considered to have two key components: (a) the sense that it is I that am doing a particular action (i.e., I am in control) and (b) the sense that I *could* have done something else (i.e., the counterfactual element). The latter component is critical for our experience of regret: I would have done better, if I had chosen the other option.

Below we outline several aspects of the experience of agency before addressing the question regarding the salience of mental agency awareness.

Intentional Binding

Research on the sense of agency received a dramatic boost from the discovery of intentional binding by Haggard et al. (2002). Libet's technique was used to indicate the subjective timing of an action: the initiation of the action (a button press) and the outcome (a sound) of the action. For an action in which the person was the author (i.e., a deliberate button press), the subjective time between these two events was shorter than the physical time. For an action in which the person was not the author (e.g., finger movement is caused by transcranial

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science," Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18, series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

magnetic stimulation) the subjective time was longer than the physical time. This suggests that (a) our experience/perception of actions is composed of two components (the initiation and the outcome) and (b) the subjective time between these two components is a marker of intentionality.

The Experience of Being in Control

As is generally the case with perception (Kersten et al. 2004), our experience of action depends on our expectations as well as on evidence from our senses. Thus, in the case of action, there are prospective and retrospective influences on the degree of intentional binding (Moore and Haggard 2008). Prospective influences (i.e., expectations) can arise from learning about the probability that an outcome will follow the movement. Retrospective influences arise from the nature of the outcome. As a result, the time at which a person experiences initiating an action can be influenced, retrospectively, by whether or not the outcome occurs (see also Lau et al. 2007).

These results indicate a considerable malleability for our experience of being in control. As with other kinds of perception, illusions of control can arise, typically through manipulation of expectations. Such illusions have been documented in detail by Wegner (2003) and include people believing that they were controlling an action when they were not, and vice versa. Beliefs about control, caused by instruction, can also alter intentional binding (Dogge et al. 2012).

Responsibility and the Sense of Agency

Whether or not the conscious experience of agency (being in control) has any well-circumscribed causal role in the action currently being performed, the experience has an important role in culture. For example, verbal reports about this specific type of phenomenal experience can now become "theory-contaminated" and begin to drive cultural evolution. As we pointed out above, "theory-contamination" is a process by which subjective reports are influenced by the scientific or philosophical theories the subject believes to be true, by social context, or by cultural background assumptions. Here, our point is that this obvious fact is not only a deep epistemological problem for the philosophy of mind, but that it can also figure in the scientific explanations of the formation of "sociocultural priors" (i.e., the emergence of *new* cultural background assumptions). Cultural beliefs about the nature of agency, such as "free will is an illusion" or "self-control is like a muscle," affect not only our experience of agency, they also impact our behavior (Job et al. 2010; Rigoni et al. 2013).

In addition to specifying the key components of agency, Epicurus believed that agency was the basis for moral responsibility (Bobzien 2006). Critical to this aspect of agency is the extension of the self across time: responsibility cannot be denied simply because an action was carried out in the past. Today, our

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

beliefs about free will are intimately connected with the idea of responsibility (Nahmias et al. 2005). When behavior is caused by conscious states, people tend to judge that the agent acted freely. In contrast, when behavior is caused by unconscious states, people judge that the agent did not act freely (Shepherd 2012). We can only be held responsible for our actions if these have been chosen freely.

The concept of responsibility has a major role in Western legal systems. If we are capable of controlling our actions, then we are responsible for these actions. If, by reason of mental illness, for example, we are not capable of controlling our actions, then our responsibility is diminished. Young children and animals are also generally considered unable to exert control and are therefore not considered responsible for their actions. However, it is very difficult to judge when and to which extent they can control their actions and must take responsibility. Public views vary and have changed over time. On occasions, animals have been tried in court (Humphrey 2002), and the age at which children become legally responsible for their actions varies widely, even within present-day Europe (Hazel 2008).

What Use Is the Ability to Detect Agency? How Does It Influence Our Social Lives?

The importance of beliefs about agency for social cohesion has been explored in the laboratory. Experimental studies of economic exchanges show how easily cooperation within groups can be subverted by the appearance of free riders, people who benefit from the willingness of others to share resources, while not sharing themselves. Cooperation can be maintained by the introduction of sanctions through which free riding is punished (Fehr and Gächter 2002). Furthermore, people prefer to join institutions in which such sanctions are applied (Gürerk et al. 2006). Importantly, however, punishment is only applied when it is believed that free riders are acting deliberately of their own free will. Punishment (or reward for good behavior) was not applied to people believed to be behaving in accord with instructions given by the experimenter, even though the consequences of their behavior was no different (Singer et al. 2006). Here then is an experimental demonstration of a link between perceived responsibility, derived from the perception or belief of deliberate agency, and contingent social regulation. Furthermore, this responsibility is associated with identifiable individuals rather than acts. The experience of agency and responsibility can optimize personal-level interactions between individuals within groups.

Regret

Individual perception is critical for the human experience of regret: I would have done better, if I had chosen the other option. The experience of regret has

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science," Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18, series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

several important implications for our understanding of consciousness, especially self-consciousness. First, the experience of regret implies an extension of the self across time: backward in time, because the action I am regretting happened in the past, as well as forward in time, because my anticipation of regret will affect my actions in the present (Filiz-Ozbay and Ozbay 2007). Second, the experience of regret emphasizes the importance of cultural factors for consciousness. It may have exactly been the emergence of the specific form of self-conscious suffering, which today we call "regret," that opened the door from biological to cultural evolution. Interestingly, feelings of regret are especially intense when the chosen action has flouted some cultural norm. This normative aspect of regret reminds us that, pre-Descartes, the concept of consciousness was to a large degree synonymous with the concept of conscience.

The Phenomenology of Regret

Regret is a form of suffering (Metzinger 2016). The first defining feature is that regretting something is a distinctly negative form of phenomenal experience, one that we will try to avoid and which we will try not to repeat or intensify (Reb and Connolly 2009). Second, regret is an *embodied* form of conscious experience: phenomenologically, it is predominantly an emotional experience (Gilovich and Medvec 1995) possessing aspects like despair (what has been done can never be changed), shame (one would like to conceal what one has done from the public or one's conspecifics), and guilt (one is acutely aware that one's past actions are immoral in the sense of having caused concrete suffering in others or having violated group interests). Very little is known about the physiological correlates, but the phenomenology of regret itself is frequently described as having interoceptive components (e.g., it can be *heart wrenching*). Third, it typically involves a cognitive aspect as well: a consciously experienced element of understanding or a sudden insight into the inadequacy of one's own past behavior. Fourth, the phenomenology of regret is always one of acutely enhanced self-awareness. In regret we experience ourselves as attaining a form of self-knowledge, which we previously did not have: we have done something morally wrong (or stupid) in the past, and we had the choice of doing otherwise. Interestingly, while the sense of agency is represented as something we possessed in the past, the state of regret itself does not itself involve a sense of agency. While the phenomenology of ownership is crisp and distinct (I *identify* with my regret, it is an integral part of *myself*), regret itself is not an action. It is a kind of inner pain that simply appears in us. This, therefore, is the fifth defining characteristic.

The phenomenology of regret can be described as a *loss of control* over our personal narrative, and in this sense it is also a threat to our integrity. It

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

is a threat to the integrity of our autobiographical self-model, because, on the personal level of description, we become aware of an irrevocable damage to our life narrative. Because it is an emotional and frequently also an embodied experience, perhaps with heart-wrenching qualities, we cannot *distance* ourselves from it—another important way in which regret involves a loss of control. This is not only about our autobiography, but also about our current and future inner life. In this sense, the cognitive aspect mentioned above is "counterfactually rich": if I necessarily will regret what I have done for the rest of my life and if, therefore, I will try very hard never to have this experience again, then a very large number of possible and future states of myself are automatically affected.[1] Regret is something that can overshadow or "color" all other phenomenal experiences that a human being can have.

The Representational Content of Regret

In regret, we have a transparent self-model, whereby the system necessarily identifies with its content. First, this self-model portrays the organism as an *agent* in a strong libertarian sense. It can initiate and control actions, and it can deliberately choose an action on the basis of its desires and values. Second, if such desires and values are represented in the transparent part of the conscious self-model, then the organism necessarily *identifies* with these values and desires, leading to the distinct phenomenology of ownership sketched above. Third, many conscious "acts of deliberation" just appear in the conscious self-model, without any introspectively available precursors. That is, there are specific action-related representational states (e.g., dynamic, conscious goal-representations) which are portrayed as spontaneously occurring and subjectively experienced as *uncaused mental events*. Fourth, there is, therefore, a phenomenology of ultimate origin (free will) grounded in the self-model, depicting the organism as having a certain, crucial *ability*: the ability to initiate spontaneously new causal chains and thus to do otherwise. The individual self is represented in the brain as possessing a plurality of futures open to it, which are fully consistent with the past being just as it was. Fifth, there is a strong representational fiction of *sameness across time*. The agent, as consciously portrayed, possesses a sharp transtemporal identity, it is always the *same* entity that acted in the past, which acts now, and which will act in the future.

[1] It is interesting to note how, phenomenologically, feelings of regret are highly "present": they are hard to suppress and continuously *re*-present themselves to the subject of experience. For the case of conscious perception, Seth (2014) proposed that "counterfactually rich" generative models encode sensorimotor contingencies related to repertoires of sensorimotor dependencies, with counterfactual richness determining the degree of perceptual presence associated with a stimulus. It is intriguing to extend his idea to the emotional layer of the self-model: the greater the counterfactual richness of an emotion, the greater its experiential degree of "presence."

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

Therefore, action consequences will always be attributed to one and the same entity: the fictional self is *responsible* for its actions.[2]

In addition, there is a novel, and much stronger representation of the social dimension: Other agents exist who have preferences too, which can be frustrated, for example, by actions for which one is responsible. These agents are also sentient, and they have the ability to suffer in many ways. In particular, a *group* exists, one's own group, and this group possesses interests and preferences as well. The group is not a sentient being, but it is a superordinate entity to which *preferences* can be attributed. There is a representation of group interests, which can be violated by individual agents, and of group preferences, which may stand in conflict with individual preferences and can accordingly be frustrated by individual actions.

In departing from theological and ancient philosophical models of regret, we propose that the representational content of regret may result from a failed integration of group preferences and individual preferences. Obviously, we also have the capacity to regret having been the cause of individual suffering, and it is often the case that the individual in question is identical to ourselves. Nevertheless, regret always has to do with conflicting sets of preferences and its representational content is inherently social. In essence, regret results from applying mechanisms of social control to oneself, namely, retribution (self-punishment) and reputation loss (self-blame). Societies are complex, self-modeling systems too, which self-regulate their activity via distributed control mechanisms that include many individual agents. Every good regulator of a social system must be a model of that system (Conant and Ashby 1970; Friston 2010; Seth 2015).

Importantly, for any organism that has acquired the capacity to feel regret and whose behavior is determined by this very special form of conscious content, the self-model and the group-model have become functionally integrated in a much stronger way. As soon as desires and values of the group are represented in the transparent part of the conscious self-model, the organism necessarily *identifies* with these values and desires (Metzinger 2003). This enables an organism to suffer *emotionally* from a self-caused frustration of group preferences. This further creates a permanent and never-ending source of conflict in its inner life. However, this source of conflict simultaneously acts as a strong source of motivation to strive continuously for social cohesion in one's own group. We believe that the conscious experience of regret marks out a critical transition in the internal dynamics of our model of reality: A functional platform for automatic self-punishment has been created. The

[2] The term "virtual identity formation" was introduced to refer to this process (Metzinger 2013:5) and it is speculated that one function of mind wandering is the constant creation and functional maintenance of the representation of a transtemporal, fictional "self." Only if an organism simulates itself as being *one and the same* across time will it be able to represent reward events or the achievement of goals as happening to the same entity, as a fulfillment of its *own* goals.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

group-model has invaded the organism's self-model to such a degree that the conflict between group and individual interests is now *internally* modeled in a way that includes (a) *sanctions* by the group (regret is internal self-sanctioning) and (b) dynamic *competition* between group and individual interests, which takes place not only on the level of overt, bodily actions but also on the self-model of the individual. In this way, social interactions and group decisions are optimized.

The Causal Impact of Conscious Processing

Viewed in isolation, the conscious experience of agency seems to occur too late to have any causal role in the action with which it is associated. Nevertheless, experience in relation to action can now affect future choices of action, as with anticipated regret. Personal-level experience, therefore, does appear to have a role beyond an individual action. It affects cultural practices, such as moral codes and laws, and shapes the sense of self, by generating beliefs about self-control, thus giving rise to concepts such as responsibility, intentionality, accountability, culpability, and mitigating circumstances. These cultural beliefs are fed back to influence the behavior of the person. This suggestion raises the interesting possibility that the sense of agency and the idea of voluntary action are acquired through cultural learning. The causal link between the group level and the individual level is constituted by the conscious self-model, in which group preferences are increasingly reflected as social complexity increases.

Wolf Singer (pers. comm.) has made the interesting observation that were this cultural learning process to take place before the formation of autobiographical memory, it would appear as "*a priori*": agency and responsibility would appear as a simple, given property in the child's autobiographical self-model as it matures. Taking this point further, we could describe the experience of agency and responsibility as an "abstract prior," a stable hyperprior guiding the process of conscious self-modeling. We treat children as responsible by rewarding and by punishing them. They grow up embedded into a cultural practice of being *held* responsible. Accordingly, their self-model always predicts that they, themselves, will be held responsible, because their autobiographical narrative unfolds in a cognitive niche which assumes that they are in control of their actions and have the ability to do otherwise.

We have already mentioned the wide cultural variations in beliefs about the age at which responsibility should be assigned. There is also some evidence for variation in beliefs about the relevance of self-control and their effect in cultural practice. Among the Mopan Mayas of Central America, perpetrators of crimes are punished according to the degree of damage that they inflicted rather than the degree to which the act was committed intentionally (Danziger 2006). As a result, the defense "I didn't mean it!" is considered irrelevant, and therefore seldom attempted. In terms of legal preconditions of criminal guilt

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

and liability to punishment, this culture has adopted a "consequentialist" (as opposed to a "deontological") approach to justice, in contrast to the test for mental competence and a "guilty mind" (*mens rea*) that is typically applied in "developed" societies. Most types of deontology hold that choices cannot be justified by their effects at all. No matter how morally good their consequences, some choices are morally forbidden, and what makes a choice the right choice is its conformity with a moral norm. Moral norms, very simply, are to be obeyed. This example again illustrates one of our main points: the phenomenal experience of agency becomes theory-contaminated by the way it is verbally described; different meta-ethical theories lead to different "socio-cultural priors" that determine which action counts as a *good* action and which agent counts as a *moral* agent. Suprapersonal models of moral agency (those shared by a society) then exert a top-down, causal influence on personal-level behavior by shaping the self-model of individual group members.

Effects of Cultural Beliefs on the Experience and Control of Action

It is unknown whether the unusual beliefs about responsibility of the Mopan Mayas have had an impact on their personal experience of action or on empirical measures such as intentional binding. However, many experiments show how manipulation of beliefs about agency can alter behavior in the laboratory.

In these studies, some participants are presented with statements such as "most rational people now recognize that free will is an illusion" (Crick 1994), while others see statements that do not involve free will. Participants who are led to doubt the existence of free will show increased aggression and reduced helping behavior (Baumeister et al. 2009). They are also more likely to cheat in exams (Vohs and Schooler 2008). Effects can be observed even on more basic aspects of action. It is well established in reaction time tasks, where participants have to be as accurate and as fast as possible, that response times increase immediately after an error (for a review, see Dutilh et al. 2012). This post-error slowing is reduced in participants who have been led to doubt the existence of free will (Rigoni et al. 2013). Furthermore, the amplitude of the brain's readiness potential, measured with EEG, which precedes voluntary responses, is reduced (Rigoni et al. 2011).

Empirical studies of the effects of regret are still in their infancy. Regret can lead to ruminative thoughts and is associated with anxiety and depression (Roese et al. 2009). Furthermore, the experimental activation of regret can lead to delayed sleep onset and insomnia (Schmidt and Van der Linden 2013). It is not surprising, therefore, that we will take action to avoid regret (Reb and Connolly 2009). When we consider the options before us, we will factor in how much regret we anticipate feeling if any particular choice turns out to be suboptimal. This anticipated regret affects our choices (Filiz-Ozbay and Ozbay 2007).

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

The Connection between Consciousness and the Evolution of Regret

We have discussed how various types of report about subjective experience can serve as data for developing an empirically constrained theory of consciousness and how such reports can be strongly "theory-contaminated." For many centuries, Western theories about regret had to do with purifying the inner life of the soul, with philosophical self-knowledge, and with man's relation to God. In the Greek philosophical and biblical tradition, important technical concepts were "compunction, "contrition," and "repentance." For example, the experience of regret could be something that leads a human being to a specific type of social action, called "confessing her sins." Here, by considering regret, we want to show how a fresh perspective of these concepts can be gained by connecting a data-driven (socio)biological approach with the more general question of what the central evolutionary functions of conscious experience might have been.

In the history of ideas, we find two main themes dominating theories of consciousness: *integration* (e.g., consciousness as a mental function that creates a union of the senses) and *higher-order moral knowledge* (inner knowledge about one's own bad actions and desires). Interestingly, the first semantic element has been strongly preserved in current research on consciousness (Metzinger 1995b; Tononi and Edelman 1998) whereas the second meaning of "conscious awareness" is almost completely absent.

In more than twenty centuries of Western theorizing on consciousness, an extremely interesting connection is found between phenomenal experience and moral cognition. The English word "conscience" is derived from the Latin *conscientia*, originally defined as jointly knowing, knowing together with or co-awareness, as well as consciousness and conscience. Here, the first point of interest is that throughout most of the history of philosophy, consciousness had a lot to do with conscience. Descartes was the first to separate conscience and consciousness and to constitute the modern concept of consciousness in the seventeenth century. Before modern times, being unconscious meant lacking a conscience. Even today, most people believe that moral considerations should only be applied to acts that are consciously intended (Shepherd 2012). The Latin term *conscientia,* in turn, stems from the Greek term *syneidesis*, which refers to moral conscience, co-awareness of one's own bad actions, inner consciousness, accompanying consciousness, joint knowledge, or disconcerting inner consciousness. Early thinkers were always also concerned with the *purity* of consciousness, with taking a normative stance, and especially with the existence of an inner witness. Democritus and Epicurus philosophized about inner torture associated with the bad conscience (Bobzien 2006) and Cicero formed the matchless term, *morderi conscientiae* (Hödl 1992): in English, the pangs of conscience (agenbite of inwit; Joyce 1922) or, in German, *Gewissensbisse*.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science," Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18, series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

Even before Christian philosophy, the idea existed that conscience is a form of inner violence, a way to persistently hurt oneself.

In many early writings, consciousness as *conscientia* is part of the conscious person as an inner space, into which sensory perception cannot penetrate. It is an inner sanctum which contains hidden knowledge about one's own actions and private knowledge about the contents of one's own mind. Importantly, it is also a point of contact between the ideal and the actual person. In Christian philosophy, this contact is established by testifying or *bearing witness* to one's own sins. All of these concepts from early philosophy suddenly sound completely different when they are not read from the perspective of the later addition of the Christian metaphysics of guilt, but rather when they are read in a fresh and unbiased manner from the perspective of an evolutionary approach to consciousness.

A second interesting idea, found in many early philosophers, is that agents share their knowledge with an ideal observer, typically God. Never, however, was there a convincing argument for saying that this ideal observation is necessarily conducted by a person or one kind or another of individual self. Here, we propose that the "ideal observer"—which lies at the origin of moral cognition and moral behavior—is a *mental representation of group interests*. This is the emergence of a "first-person plural perspective" (Gallotti and Frith 2013). Self-consciousness served as a functional platform for the representation of group preferences in the brain of individual organisms. Upon this platform, individual and group interests could compete. The mechanisms which constitute self-consciousness are often subpersonal; the representational content is suprapersonal.

Consciousness as the Interface between the Person and Culture

Our actions and the brain systems through which they are implemented depend on a hierarchy of top-down control (Felleman and Van Essen 1991; Friston 2005; Koechlin and Summerfield 2007). This hierarchy of control, however, does not stop inside the person. In the examples given above, and, indeed, in most experiments, the highest level of top-down control comes from the instructions given to the participant by the experimenter (Roepstorff and Frith 2004) and, ultimately, from culture.

In many experiments, including those discussed above, instructions are designed to manipulate the beliefs of participants. For example, in economic games participants learn, by trial-and-error, that some of their partners can be trusted to make fair returns of the money invested in them, whereas other partners cannot be trusted. Participants also learn about the trustworthiness of information given by the experimenter (Delgado et al. 2005) or through gossip from other participants (Sommerfeld et al. 2007). Such information changes

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

the participants' behavior, even though there is no actual difference in the behavior of their partners.

In these examples, acquiring a new model of reality (or in traditional Bayesian terms, a *belief* about the world)[3] causes changes in behavior, even when it is false. Interestingly, this can still count as an example of mental causation, because the representational content of the self-model accounts for the shift in behavioral profile, and also because conscious experience itself has the critical role of causally enabling the transfer of a model from one mind (the experimenter's) to another (the participant's). In other words, change in behavior is a causal consequence of shifts in the functional profile of the participant's phenomenal self-model brought about, in this case, by the instructions of the experimenter.

Mechanisms of Suprapersonal Top-Down Control

The learning process that occurs in trust games is nicely captured through a Bayesian mechanism. When we invest money in a partner, we can predict how much of our money will be returned on the basis of our degree of trust (a prior belief). If we get more than expected (positive feedback), our degree of trust increases. If it is less (negative feedback), our degree of trust decreases. However, if we are given prior information about trustworthiness, much greater weight is given to the prior information than to direct experience. This effect has been observed in terms of brain activity (Fouragnan et al. 2013) as well as behavior (Sommerfeld et al. 2007).

We suggest that beliefs arising from instructions, or from culture more generally, exert their effects by modifying prior expectations at the highest level of the personal hierarchy of control. Effects of these modifications demonstrably penetrate deeply into the hierarchy of control, affecting the monitoring of low-level cognitive processes (Rigoni et al. 2013) and associated brain activity (Rigoni et al. 2011).

A similar process might explain the effects of manipulating (or first installing) beliefs about free will. Our basic urge, we believe, is to be selfish, to gain advantages at the expense of others. This is one of those "abstract priors" that emerges through very early cultural learning. To overcome this urge we have to exert self-control (Metzinger 2015). Free will is necessary to exert such control (Nahmias et al. 2005). It is this intentional, top-down control that enables us to behave in a moral fashion. Without such top-down control, we might as well give in to our basic urges and gain all the (short-term) advantages that this

[3] It is important to note how the largest part of our model of reality cannot be adequately reconstructed as a set of beliefs (where, according to the standard definition, a belief would be the relation between a person and a proposition). Neural representations in human brains do not come in a propositional format, as they do not have the necessary properties of systematicity and productivity—the information expressed by a Bayesian model in the biological brain is a *subsymbolic* representation of probability distributions.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

might bring. This leads to an increase in cheating and general antisocial behavior. Ironically, telling people that there is no free will alters their very behavior, thus providing another example of mental causation and the effective role of conscious self-representation.

Sharing Experiences

We have discussed how instructions and culture influence the person, but there is, of course, traffic in both directions (Sperber 1996). The explicit metacognitive mechanisms that enable us to be influenced by the ideas of others also allow us to influence them. This permits control, not just at the personal level, but also at the suprapersonal level (Shea et al. 2014).

In the choice blindness paradigm discussed above (Johansson et al. 2005), participants are easily persuaded to accept that they have made a different decision from the one they actually made. This phenomenon is part of a larger set of examples showing that we have remarkably poor access to the mental processes underlying our behavior (Nisbett and Wilson 1977). In spite of such meager knowledge, people are more than happy to talk about and justify the decisions they have just made.

Although the conscious experience of agency may have little causal role in the action with which it is associated, the experience will be very relevant to any attempt to justify the action after it has been made. We would be able to claim, for example, that our action was accidental rather than deliberate. By justifying our actions and discussing with others why we do things, a consensus is built about the mental basis of action. Whether or not this is a *true* account of the mental processes, such consensus is likely to be an important basis for cultural norms about responsibility. Thus, consciousness of action enables us to develop a folk psychology critical for the regulation of social behavior (McGeer 2007).

We not only tell each other about our experiences of action, we also share our perceptual experiences. In a series of experiments, Bahrami et al. (2010) have shown how such discussion can create group advantages. In these studies, two people jointly perform psychophysical signal detection tasks. After giving individual reports about the presence of a signal, disagreements are resolved by discussion, leading to a joint decision. If the abilities of the partners are roughly equal, then the joint decision is consistently better than that of the more skillful person working alone. Discussion is crucial for optimizing this group advantage and requires that the partners talk about their confidence in their experience of the signal (Bang et al. 2014). Through such discussion they develop a verbal scale for rating their levels of confidence. Group advantage depends on the development of such a scale (Fusaroli et al. 2012).

We suggest that these group advantages, which depend on the experience of and ability to report confidence in a perception, constitutes another case where consciousness has an important and possibly necessary function. Transparency

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

is the phenomenological equivalent of maximal confidence. In this case, the explicit report of confidence enables optimization of joint decisions. It remains to be seen if elegant new paradigms can be created to show that even these aspects of our mental lives may occur without conscious awareness. For example, do some phenomena associated with hypnosis (e.g. Smith et al. 2013) indicate that instructions can have their effects without awareness?

We conclude that, in relation to both action and perception, a particular kind of self-consciousness arises at the point in the hierarchy of control where the person interacts with other minds. This is the level at which instructions work. At this interface, between the person and culture, there is two-bidirectional traffic (Sperber 1996), such that the person can be influenced by other minds and the person can in turn influence others. So, what use does consciousness fulfill? We propose that at least one kind of consciousness functions to enable explicit communication about subjective experience. This, in turn, causally influences behavior and enables the growth of cumulative culture. This growth is dependent on the development of norms about appropriate behavior. This kind of consciousness creates the social cohesion and cumulative culture that has proved such an immense advantage to humans.

Regret and Regret Prediction: The Argument from Transparency and Modal Competence

Regret is a very specific kind of representational content: it carries information that a biological organism can utilize to optimize future decisions and enables group preferences and norms to have a direct influence on the behavior of individuals. Returning to the question posed in the introduction, what is it that, in our world, a zombie could never do? In our world, a maximally similar but unconscious creature could never be a true functional isomorph. Why? Because it would lack the representation of "realness," and thus it could not compare real and counterfactual states of self and world, and because it would not possess the enormous motivational force that comes from identifying with the contents of one's self-model.

Regret carries self-related information, which often refers to specific *social facts*. The evolutionary advantage of representing this information under the very specific, neurally realized data format of a transparent, egocentric model of reality, as described above and elsewhere (e.g., Metzinger 2003, 2009) is that it forces a biological organism to:

1. Experience the relevant kind of fact about the world as irrevocably *real* (e.g., damage to the interests of its own group, or itself, has been done, the organism itself was the cause of this damage, and it could have done otherwise). Let us call this the "principle of phenomenally transparent representation."

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science," Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18, series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

2. Identify with this damage by integrating it with its internal self-representation. We could call this the "principle of transparent self-modeling."

Regret is a particularly powerful form of conscious experience, because it represents the group's interests *in* the individual's transparent self-model, thereby creating a new form of suffering from which the organism cannot distance itself—for the simple reason that the relevant form of representational content has now been functionally integrated with an internal representation of *itself*. The sense of agency is the decisive causal prerequisite, because it introduces the phenomenal experience of "I *could* have done otherwise!" (whether true or not) into the self-model.

Let us define "modal competence" as the ability to represent mentally the operators of modal logic and their function: □ (It is *necessary* that…) and ◊ (It is *possible* that…), but also *F* (in deontic logic, It is *forbidden* that…) or *P* (in temporal logic, It *was the case* that…). Modal competence is a naturally evolved form of intelligence, which comes in many degrees. In our context, the mental ability to represent successfully some things as *possible* and other things as *real* (i.e., as actual facts) is of highest importance. If a biological organism is to develop higher forms of intelligence like episodic memory, future planning, or counterfactual reasoning, it needs a simple form of modal competence. To develop these forms of intelligence, it needs a functional mechanism that reliably distinguishes between what is real and what is only possible or what happened in the past; for example, the animal must avoid episodic memories from turning into hallucinations and manifest daydreams, or, as in future planning, it must find an optimal trajectory from a model of the world reliably marked out as "given" into a second model of the world portrayed as "possible and desirable." Only conscious representation has this remarkable functional property and, on the level of self-representation, it is exactly this property that causally enables the phenomenal experience of "I *could* have done otherwise!"

Under the Bayesian predictive coding framework, we assume unconscious inferential processes which lead to a continuous, dynamic representation of probability distributions (Friston 2010). Only conscious experience, however, can represent something as *real* and as taking place *now* (Metzinger 2003; Lamme 2015a, b; Melloni 2015), and only self-consciousness provides a singular unit of identification. There could be unconscious models of the organism as a whole, of individual and group preferences, and so on, and they could certainly be characterized by a high degree of Bayes optimality. But only misrepresenting the probability of a hypothesis as 1.0 and simultaneously flagging it as a fact holding *now* via a window of presence turns a possibility (or a likelihood) into a reality. This is what makes the zombie conscious. The argument from transparency is that conscious experience must be exactly the functional mechanism that "glosses over" subpersonal Bayesian processes by assigning "realness" to them—that is, by misrepresenting them as exemplifying an *absolutely* maximal likelihood or maximum posteriori probability. It is this step

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

that turns a process into a thing, a dynamical model into an internal reality, and a self-model into a self.

Therefore, it is only conscious experience that enables suffering and the enormous motivational force that comes with representing something as an irrevocable and untranscendable fact and at the same time as a threat to one's *own* integrity. We believe that it is the conscious self-model that causally integrates the continuous, low-level biological process of sustaining the organism's existence with a specific dynamic representation of the system, namely, a generative self-model that continuously strives to find evidence for the system's very existence (Hohwy 2014; Friston 2013). If this internal self-model has the capacity to integrate social facts (e.g., the frustration of group preferences) then it creates a new biological phenomenon: the causal integration of the individual's striving for self-sustainment with the group's need for cohesion and stability. This is a culturally shaped form of self-consciousness, linked with the idea of identity (see Kyselo 2014). It enables new types of actions aimed at the satisfaction of group preferences, because it makes a new set of facts globally available for introspective attention, verbal communication, and behavioral self-control.

At the outset, we also asked: Which class of optimization problems does consciousness enable us to solve? A well-known neuroscientific concept is "reward prediction" (Hollerman and Schultz 1998; Schultz and Dickinson 2000; Tobler et al. 2006). We want to point out that in complex biological nervous systems the opposite capacity might also exist, and we dub it "regret prediction." If a system has the capacity to distinguish between its own actual and possible future states, then it could also begin predicting future regret (Filiz-Ozbay and Ozbay 2007; Coricelli et al. 2005). It could simulate future states of the self-model that resemble the current one. If it has a self-model that misrepresents it as possessing a precise transtemporal identity, then it will also represent such future regret events as potentially happening to the *same* biological system, to itself. The prediction of future suffering of the kind we have sketched in this chapter allows for the comparison of future states with present states, and opens the possibility of seeking trajectories into more desirable situations. We believe that this new biological capacity—regret minimization—will dramatically have increased the motivational force behind prosocial behavior. The search for one's own coherence turns into the search for group coherence.

The experience of regret is intimately associated with the experience of agency: the experience that I did it and that I could have done otherwise. In closing, we wish to draw the reader's attention to a specific logical possibility. Ultimately, regret, like the experience of being an agent, may be a form of self-deception, a naturally evolved, but functionally adequate form of misrepresenting reality. Exactly this form of "theory-contaminated self-deception" may have provided a mechanism for cultural evolution and the sustaining of social cohesion, therefore providing advantages for the group as a whole (von

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

Hippel and Trivers 2011; Trivers 2011). Kierkegaard (1843/1992) made a similar point in *Either/Or*: "The deceived is wiser than one not deceived."

Ackowledgments

CDF acknowledges support from the Wellcome Trust and Aarhus University. We are grateful for comments from Cecilia Heyes, Holk Cruse, and Marek McGann.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.