

On the Alleged Illusion of Conscious Will

Marc van Duijn and Sacha Bem

The belief that conscious will is merely “an illusion created by the brain” appears to be gaining in popularity among cognitive neuroscientists. Its main adherents usually refer to the classic, but controversial ‘Libet-experiments’, as the empirical evidence that vindicates this illusion-claim. However, based on recent work that provides other interpretations of the Libet-experiments, we argue that the illusion-claim is not only empirically invalid, but also theoretically incoherent, as it is rooted in a category mistake; namely, the presupposition that neuronal activity causes conscious will. We show that the illusion-claim is based on the behaviorist ‘input-output’ paradigm, and discuss the notions of ‘self-organization’ and ‘self-steering’ to provide an alternative perspective on the causal efficacy of conscious will. In the final sections, a tentative theoretical picture is sketched of conscious will as an instance of self-steered self-organization. We conclude that the subjective experience of conscious will is not a misguided one, but rather that the mechanisms supporting conscious will are considerably more complex than mainstream cognitive neuroscience currently acknowledges.

Keywords: Conscious Will; Benjamin Libet; Self-Organization

1. Introduction

Science and common sense have a rather long history of conflict. A modern version of this classical dispute can be found in current debates on the function of conscious will in behavior. Recently, some renowned cognitive neuroscientists have argued that the subjective experience of our conscious will is fundamentally misguided, and that there is empirical evidence to support the claim that conscious

Correspondence to: Marc van Duijn, Filosofisch Instituut, Oude Boteringstraat 52, 9712 GL, Groningen, NLD.
Email: M.van.Duijn@rug.nl

will is in fact an illusion created by the brain (e.g., Gazzaniga, 1998; Nørretranders, 1998; Roth, 2003; Wegner, 2002, 2003). This “illusion-claim” is based on the finding that neuronal activity that sets-up a consciously willed action, occurs prior to the conscious decision to engage in that action (e.g., Libet, 1985). We are normally not aware of this discrepancy, so the claim goes, because the brain retrospectively fabricates a coherent story that is accessible to subjective experience. Introspection is thus asserted to provide us only with access to an illusory causal path from thought to action (Wegner, 2003); the causal efficacy of conscious will is therefore only apparent (Nørretranders, 1998; Roth, 2003; Wegner & Wheatley, 1999).

The empirical evidence supporting the illusion-claim is partly based on the classical, pioneering experimental work by physiologist Benjamin Libet. His research suggests that a conscious decision to engage in an action is preceded by neuronal activity that sets up this action. However, Libet’s work has been controversial from the start; his experiments have also been claimed as evidence supporting pre-cognition, retro-causality, “quantum-consciousness,” and dualism (Bierman, 2003; Eccles & Popper, 1977; King, 1997; Penrose, 1989; Radin, 2003; Wolf, 1989). Over the last 20 years or so, criticisms related to Libet’s own interpretations of his data have accumulated, and other, less radical interpretations have been suggested that appear equally compatible with Libet’s data (e.g., Churchland, 1981a; Glynn, 1990; Gomes, 1998; Pockett, 2002). Given these criticisms and alternative interpretations, can Libet’s findings still be seen as supporting the illusion-claim?

In this article, we contest the empirical and theoretical validity of this illusion-claim, and discuss the notions of ‘self-organization’ and ‘self-steering’ to provide an alternative perspective on the causal mechanisms that support conscious will, using the language of dynamic systems theory. In the final sections, we provide a tentative theoretical picture of conscious will as an instance of self-steered self-organization, based on recent empirical work that targets self-organizing processes in the brain. We conclude that the subjective experience of conscious will is not a misguided one, but rather that the neuronal mechanisms that govern conscious will are considerably more complex than mainstream cognitive neuroscience currently acknowledges.

2. The Libet-Experiments

Benjamin Libet is regarded as a true pioneer in the empirical study of human consciousness. Two of his experimental approaches have been much discussed over the years. The first approach, which we refer to as the ‘back-referral experiments’ (Libet et al., 1964; Libet et al., 1967; Libet et al., 1979), is based on the assumption that there exists a substantial delay before cerebral processes achieve what Libet (1985) calls “neuronal adequacy”—i.e., according to Libet (1991), the brain requires a considerable neuronal build-up period, up to 500 ms, in order for us to become conscious of peripheral stimuli. Still, the subjective timing of the sensory experience

appears to be without a significant delay, which is due to an alleged back-referral mechanism (e.g., Libet et al., 1979). Based upon the findings of these back-referral experiments, Libet and his collaborators came up with their much discussed second experimental approach, the ‘readiness potential’ or ‘RP-experiments’ (Libet et al., 1982; Libet et al., 1983; Libet, 1985).

The readiness potential (RP) is an electrophysiological indicator for cortical movement preparation. It normally precedes a voluntary motor act by approximately 500–1500 ms (Libet et al., 1982; Haggard & Eimer, 1999). The appearance of these preparatory cerebral processes at such surprisingly long times before spontaneous acts led Libet to question whether the conscious intention to act occurs equally far in advance of the actual movement. Libet argued that if the RP-onset occurs before the time of a conscious decision to perform a certain action, consciously willed decisions therefore also require a substantial build-up period to achieve “neuronal adequacy.”

In order to test this hypothesis, Libet et al. (1983) devised an experiment to time objectively subjects’ awareness of their “urge to move” when performing spontaneous movements, and to relate this subjective timing to the onset of the RP. This was achieved by instructing his subjects to fix their gaze on the circular screen of a cathode ray oscilloscope (CRO). The CRO had a spot of light that revolved clock-wise near the circumference of the screen with numbers at each five-second position, similar to an ordinary clock. While the RPs were recorded on the subjects’ scalps, the subjects were directed to abruptly and spontaneously flex the wrist or fingers of the right arm, at any time after the spot on the CRO completed one full revolution. Before the start of the experiment, the subjects were asked to remember the clock position at the moment they were aware of an “urge” or a “feeling to make movement.”

The results of the RP-experiments showed that preparatory neuronal processes, measured in the form of RPs, preceded the awareness of the conscious intent by about one third of a second. The onset of the averaged RP was found to be about 550 ms before the arm-movement, whereas the subjects reported their “urge to move” only 200 ms before the action, 350 ms *after* the RP-onset.

Libet (1985) concluded from these RP-experiments that although conscious will does not appear to control the initiation of actions, it still acts as a veto-mechanism: as his subjects were aware of an “urge to move” 200 ms before the act itself, the veto to move or to refrain from this act was under conscious control during a brief 100–200 ms time-window just before the act was executed. To Libet then, conscious will does not initiate behavior but is a binary control device that censors unconsciously generated actions.

As said before, although Libet’s views have been criticized heavily over time (e.g., Churchland, 1981a,b; Flanagan, 1997; Glynn, 1990; Gomes, 1998, 2002; Keller & Heckhausen, 1990; Pockett, 2002), some have adopted his experiments as evidence in favor of the idea that conscious will is merely “an illusion created by the brain” (e.g., Gazzaniga, 1998; Nørretranders, 1998; Roth, 2003; Wegner, 2002, 2003). For instance, neuroscientist Gerhard Roth (2003) recently concluded the following from Libet’s

RP-experiments:

These studies demonstrate that the ‘act of will’ occurs several hundred milliseconds after the onset of the . . . readiness potential. This means that the ‘act of will’ cannot cause voluntary actions and that the freedom of the ‘act of will’ is an illusion. (pp. 129–130)

On this view, even Libet’s (1985) conscious veto would only be an epiphenomenon of unconscious neuronal activity (Dennett & Kinsbourne, 1992; Gomes, 1998). Recently, however, other interpretations of Libet’s work have been offered, which also provide an alternative perspective on the validity of the illusion-claim (e.g., Glynn, 1990; Gomes, 1998; Pockett, 2002).

3. Other Interpretations of the Libet-Experiments

Many have questioned the ecological validity of the Libet-experiments: Can Libet’s finding that neuronal activity precedes a conscious willed decision be generalized to normal, every day conscious decisions? Keller and Heckhausen (1990) and Gomes (1998, 2002) argue that conscious intent was “artificially imposed” by Libet’s experimental set-up: Libet et al. (1983) instructed their subjects to pay attention to an “urge” or a “feeling to move,” and to initiate the willed act when this feeling occurred. However, Keller and Heckhausen (1990) claim that unconscious motor responses to a particular sensory stimulus can be detected through a mechanism of selective attention. The “urge to move” in the RP-experiments, then, could have functioned as an internal stimulus to which the subjects responded by executing the predefined act. While subjects were prepared to make a movement, introspective attention thus “picked up” neuronal activity that is normally unconscious. The ecological validity of the Libet-experiments is therefore rather questionable (Gomes, 1998).

Another much discussed problem related to the Libet-experiments deals with whether the conscious decisions were made by the subjects on every single trial of the RP-experiments, or only once, at the beginning of the experiment. For example, van de Grind and Lokhorst (2001) argue that when the subjects were given instructions on how to act in the experiment, they also consciously prepared themselves to participate in the experiment, thereby activating or “pre-programming” the appropriate response (see also Flanagan, 1997; Gomes, 1998; Joordens, van Duijn, & Spalek, 2002; Keller & Heckhausen, 1990; Zhu, 2003). After that, automatic unconscious processes took over, initiating pre-set, automatic movements on each trial.

On this view, conscious will only guides behaviors in a global manner and monitors when it is time to stop or initiate unconscious “motor programs” (van de Grind & Lokhorst, 2001). This view is further corroborated by recent research that also suggests that the role of conscious will in behavior is indeed an active one (e.g., Baars, 1997; Elsner & Hommel, 2001; Jahanshahi & Frith, 1998; Jeannerod, 2000). Given the limited capacity of the conscious system, the bulk of processes need to be automatic and unconscious. By far the majority of our actions are thus prepared and executed automatically, which guarantees a fast and accurate execution

of the act. Conscious control on the other hand operates at a much slower pace, but prevents us from getting “stuck in set” by solving failures that occur at the automatic levels, by creating the context for understanding an action, and by rearranging the means to achieve a certain goal if necessary (Flanagan, 1997; Jahanshahi & Frith, 1998; Jeannerod, 1997). Conscious will therefore appears to provide a meta-level of behavioral control that creates, edits, and selects global action plans, which in turn enables us to engage in more strategically complex behaviors (e.g., Baars, 1997; Jeannerod, 2000). On this view, it is not necessary to assume a direct conscious cause for every single voluntary movement. By far most of our actions are automatically controlled and executed but have to be understood in a broader context in which conscious will influences behavior by setting up the automatic action-routines in advance.

With this alternative interpretation in mind, Libet’s findings of the appearance of RPs before the conscious initiation of an act are hardly surprising. However, although these alternative interpretations of Libet’s findings and of conscious will appear to resolve most of the counterintuitive consequences of Libet’s (1985) conclusions, they still do not exclude the illusion-claim. After all, it could still be argued that even though conscious will appears to globally set-up and activate “motor programs” (van de Grind & Lokhorst, 2001), even such a global conscious decision would be preceded by preparatory neuronal activity (Gomes, 1998). On this view, the genuine causal processes that support conscious will would thus only reside on the neuronal level, and not on the conscious one. The preparatory neuronal activity that sets up a willed action would also give rise to an after-the-fact illusory subjective experience of a causally effective will (e.g., Nørretranders, 1998; Roth, 2003; Wegner, 2002).

In sum, although critics to Libet’s work have provided a convincing case against the validity of his main claims, we need to dig deeper into the background assumptions of the illusion-claim in order to reveal its theoretical incoherencies.

4. The Input-Output Paradigm

Although behaviorism is no longer fashionable in the modern cognitive neurosciences, some of its theoretical assumptions still appear to be of influence today. In particular, the stimulus-response paradigm, or ‘input-output model’, has remained quite tenacious (e.g., Anderson, 1995; Freeman, 1999b). The behaviorists looked upon perception and action as two separate mechanisms, and acquired associations between stimulus and response were seen as deterministic, one-way causal constructs. Action was thus regarded as a linear and reflexive effect of perception, while feedback loops between action and perception were disregarded (Hurley, 2001). The cognitivists elaborated on this behaviorist model by wedging “cognition” in between the “perception” and “action” boxes (Hurley, 1998). This way, the behaviorist reflex paradigm became incorporated into cognitivist psychology.

Because of the influence of the input-output model, psychological phenomena are still sometimes explained in a strict, linear causal manner (Anderson, 1995).

Behavior is then captured in terms of the classical Newtonian picture of causality: events are placed on an imaginary time-line, succeeding each other in a discrete, and conveniently arranged linear sequence (Freeman, 1999a). These linear causal models abstract away from their context, and characteristically require a “first cause”: one discrete occurrence at the origin of a discrete causal cascade of events.

The influence of the input-output paradigm and linear causal thinking can also be recognized in Libet’s work. In order to account for his results, Libet (1985) placed the RP-onset at the beginning of the linear causal chain, as to constitute the first cause. The RP-onset is followed by the conscious veto, which in turn leads to the execution of the act itself. From a strict linear causal perspective, it seems that the RP-onset actually causes the consciously willed decision, and that therefore the causal work is performed by the first event in the sequence, the RP-onset, effectively robbing conscious will of its causal powers.

Libet’s notion of the “conscious veto” attempts to escape from this epiphenomenalistic threat by restoring the idea that conscious will still matters to the behavioral outcome just prior to the execution of the act. However, this apparent “veto solution” results in a rather artificial conception of conscious will as a reactive veto-mechanism that only has a say in whether or not unconsciously generated actions can proceed—a solution that does not convincingly restore the idea of a causally potent conscious will (Dennett & Kinsbourne, 1992; Gomes, 1998).

Unfortunately, the view that conscious will only guides our global behavior and not every single action we take (Flanagan, 1997; Keller & Heckhausen, 1990; van de Grind & Lokhorst, 2001) also succumbs to this strict, linear causal way of thinking, as it simply places the conscious decision at the beginning of the linear time-line, before the RP-onset. This theoretical move is incomplete because it results in a fruitless, never-ending discussion: What comes first, the conscious decision to initiate a certain action, or the preparatory neuronal activity that sets-up and causes the conscious decision?

It has become apparent that linear causal models cannot provide a satisfying answer to this discussion, and that a fundamentally different approach is required. In the remainder of this paper, we develop an alternative perspective on the causal efficacy of conscious will.

5. Self-Organization

During the last few decades, self-organization has been increasingly recognized as an important subject of scientific inquiry (e.g., Haken, 1987), and researchers from various scientific disciplines now study it. ‘Self-organization’ is generally defined as a process that, given certain boundary conditions, gives rise to increasing order in a particular system by spontaneous synchronization of system parts, without a central executive that helps to set-up this self-organization (Jantsch, 1980). Heylighen (2001) discusses (ferro)magnetization as a conservative example of self-organization¹: potentially magnetic materials can be seen as to consist of a collection of tiny, individual magnets called “spins,” which represent local magnetic domains. At very

high temperatures these spins show no coherence, i.e., their individual orientation is random so that they cancel each other out. However, when the material cools down and reaches a material specific transition temperature, the so-called “Curie point,” the individual spins spontaneously begin to organize and to align themselves. When the material solidifies, the given system reaches a so-called “attractor state,” locking the collective spin-dynamics in-place, and thereby providing that material with “macroscopic” magnetic properties (Heylighen, 2001).²

Self-organization is more complex in so-called “dissipative structures” or “far-from-thermodynamical-equilibrium-systems.” These dynamical systems can be described in terms of changes in the values of a set of interdependent variables (e.g., Port & van Gelder, 1995), which together constitute a high-dimensional state-space or phase-space, with each variable counting as a distinct dimension. The function of the dynamics of the system are depicted as a trajectory through an n -dimensional state-space containing attractors and repellents. The attractor minima correspond to states of minimum energy and maximum stability, while the repellent maxima represent states of maximum energy and minimal stability in a given system. The trajectory of the system can then be drawn and settle into multiple kinds of attractors that may vary in complexity.

Self-organization can be seen to operate as an attractor in a dynamical system. Under the proper boundary conditions (i.e., a particular temperature or energy level), systems can then be drawn into self-organization (Heylighen, 1989). In complex, dynamic, open systems pumped with energy, like our atmosphere, self-organization can technically be seen as an entropy/extropy cycle that supports a net flow of order into the system—enabled, on the one hand, by maintaining extropy or order import and production; and, on the other hand, by dissipating or exporting the acquired thermodynamical entropy by, e.g., giving of heat. The amount of entropy that a system is able to disperse is then a boundary condition for its organizational growth (Jantsch, 1980).

Besides maintaining the energy flow to sustain self-organizing processes, harnessing and utilizing this energy is what life forms do best. This makes self-organization not only a particularly relevant phenomenon for biology (e.g., Ben-Jacob, 1997; Ho, 1996), but also for the brain and behavioral sciences (Bressler, 2003; Haken, 1995; Jordan, 2003; Kelso, 1995; Krippner & Combs, 2000).

6. Self-Steered Self-Organization

A particularly complex form of self-organizing, namely, “self-steered self-organization” (Keijzer, 2003), has strong connotations with intentionality and conscious will (see also Jordan, 2003; Kelso, 1995). Self-steering processes involve internal control parameters that modulate the dynamics of self-organizing patterns into a particular adaptive direction, characterizing dynamical biological processes such as metabolism, morphogenesis, and ontogenesis (Ben-Jacob, 1997). DNA, for example, can be seen as an internal control parameter that perturbs the self-steered

dynamics of the morphogenesis of organisms, without offering a strict prescription of the overall outcome (Keijzer, 2001).

The presence of self-steering systems allows the hierarchical organization of the body to remain relatively flat, making its organization with all its different, relatively autonomous subsystems more efficient (Aulin, 1979; Heylighen, 1989). The more complex organisms become, the more they rely upon internal, self-steering factors in order to maintain internal stability (Beer, 1974). Self-steering processes can therefore be said to constitute an evolutionary compromise in complex biological systems between the importance of internal stability on the one hand, and effective organizational design on the other.

The link between self-steered self-organization and conscious will becomes more salient in the so-called “bimanual phase transition paradigm” (e.g., Lee, 2004). For example, Kelso (1995) demonstrated that when subjects performed bimanual finger movements to a metronome, both the in-phase and the out-of-phase pattern of finger twiddling can be equally stably maintained at low frequencies. At higher frequencies the out-of-phase pattern soon becomes unstable, as subjects have a tendency to synchronize to the metronome. However, the conscious wish to maintain an anti-phase pattern results in a more prolonged stability of the out-of-phase pattern at higher frequencies, and thus prevents the phase change to the in-phase pattern. This implies that the conscious wish to maintain a certain behavioral pattern is accompanied by the stabilization and destabilization of the already inherent coordination dynamics. Kelso (1995) introduced “specific parametric influences” or “coordination variables” to refer to these intentional forces: intrinsic information that alters the dynamics of a system, steering it towards the intended pattern.

Although self-steering phenomena occur in many biological and psychological contexts (Keijzer, 2003), the notion of ‘self-steering’ may help provide us with an alternative perspective on conscious will. On this tentative view, conscious will can be seen as perturbing the intrinsic behavioral coordination dynamics of brain, although it cannot directly override its inherent constraints. That is, intentional switching between behavior patterns is dependent on and constrained by the stability and strength of existing patterns. Conscious will could then be envisaged as a self-steering neural mechanism generated by the collective, self-organized activity of neurons, providing the brain with a tool to rapidly switch between existing coordination dynamics, and to stabilize them when necessary, in order to achieve a particular behavioral goal. As many such self-steering self-organizing processes take place within our body and brain, the self-steering influences that govern large-scale, self-organizing neural phenomena should not be regarded as particularly special or mysterious. Nevertheless, it is clear that the neuronal mechanisms that support conscious will are still largely beyond the grasp of contemporary cognitive neuroscience.

The idea that neurons cooperate on a macroscopic level to sustain psychological functions is not a new one, however, and can be traced back to the works of Pavlov, Luria, and Wernicke (Bressler, 2003). Donald Hebb (1949) already spoke of

“neuronal cell-assemblies” that act as a “diffuse structure comprising cells in the cortex and diencephalon... capable of acting briefly as a closed system” (p. 73). Whereas the proper tools to understand these integration mechanisms were not present in those times, we believe that relatively novel concepts such as self-organization and self-steering may help pave the way for a better understanding of conscious will, including its causal efficacy. At least it allows for an alternative perspective on the illusion-claim, on which we elaborate in the next section.

7. Levels and Causality

As discussed earlier, the illusion-claim is rooted in the philosophical position that the genuine causal processes that underlie conscious will only take place at the neuronal level. On this view, the dynamics of self-organized, macroscopic neuronal patterns can, at least in principle, be understood and explained by a complete account of the causal processes at the level of neurons and synapses. The modern cognitive neurosciences, however, study a wide range of phenomena, from the workings of NMDA-receptors and ion channel dynamics, up to neuronal oscillation rhythms, and, of course, behavioral and psychological events.³ In other words, it is an enterprise that spans an entire hierarchy of mechanisms across multiple levels of description (Craver, 2002). That is, levels are domains at a particular spatiotemporal resolution scale, so that events at one level obey particular characteristic laws or regularities that cannot be found at other levels, which in turn allows a clear demarcation of such levels (Juarrero, 1999). Additionally, not only do levels form a particular kind of a part-whole relation, there is also the restriction that the components on the microscopic level “are organized tighter to produce the behavior of the mechanism as a whole” (Craver, 2002, pp. 7–8). In other terms, the microscopic parts of a level are constrained to behave in a certain way; order or organization is therefore another key component in distinguishing levels (Bechtel & Craver, 2003).

A part-whole or stratified picture of ‘levels’ is common in most scientific disciplines (Bem & Looren de Jong, 1997). For example, it is generally acknowledged that the laws of thermodynamics capture regularities that do not occur at the atomic level (Freeman, 1990). However, it appears that in the cognitive neurosciences the influence of the input-output paradigm and linear causal thinking has hampered the assimilation of this mereological perspective. Linear causal frameworks in particular have problems dealing adequately with hierarchies of mechanisms, as they can only mold hierarchical relations into strict, linear cause and effect relations. According to Juarrero (1999), the category mistake to ascribe linear causal powers to each individual explanatory level is then easily made. Such a category mistake lies also at the root of the illusion-claim: neuronal activity is seen as causing conscious will, whereas they only form a constituency relation. Saying that neuronal activity causes conscious will is therefore very much like saying that H₂O molecules cause water.

Instead, the notion of self-organization comes with a distinction between a macroscopy and a microscopy as two different levels of aggregation (Keijzer, 2003),⁴

and it is here that the circular nature of self-organizing phenomena becomes more apparent, as we will show in the next section. The notion of self-organization then provides a handle on better describing interlevel relations, and, as we will argue later on, a more natural understanding of the causal efficacy of conscious will.

8. Interlevel Constraints

Interlevel causation is a rather controversial topic in modern philosophy. One reason for this is that interlevel causation is often erroneously captured in terms of strict, linear cause and effect relations between levels (e.g. Bechtel & Craver, 2003; Juarerro, 1999). How then do higher- and lower-level phenomena relate if not in a strict linear, causal, bottom-up and top-down manner? Clark (1997) argues that on the one hand, the elementary units of a given system may constitute macroscopic patterns, while at the same time these patterns can be said to constrain their microscopic elements to behave in ordered patterns. This two-way or reciprocal relation between a micro- and a macro-level is then often captured in terms such as “circular causation” (e.g. Freeman, 1999a; Kelso, 1995), “continuous reciprocal causation” (Clark, 1997, p. 163), and “inter-level constraints” (Juarrerro, 1999). In this context, the so-called cooperative phenomena in physics are highly relevant. For example, both water and ice are made of the same micro-level components, which lack properties such as liquidity or solidity. Water and ice as macroscopic phenomena, in turn, each comes with different constraints on the organization of H₂O molecules. These constraints then sustain system properties that exhibit causal powers that cannot be found at the level of its individual parts (Bechtel & Craver, 2003). It is thus the organized interaction between a multitude of units that yields novel system properties.

Self-organization is also a process that imposes order or constraints on system parts. For example, Haken’s (1987) laser gives rise to coherency of emitted wavelengths that compels all elements to behave in a coherent manner. The self-organized coherent bundle of photons comes with novel system properties, such as the ability to cut through certain materials, whereas a bundle containing collectively out-of-phase photons lacks these properties. The photons are thus said to be “enslaved” into order by self-organization, so that all parts no longer behave independently but as a collective (Kelso, 1995). This self-organized coherency expands the state-space of the system, bringing along more degrees of freedom and additional causal powers for the system as a whole, while simultaneously constraining the number of possible microscopic states.

We believe that Haken’s (1987) laser, and self-organizing phenomena in general, provide a good analogy for thinking about the causal efficacy of conscious will, as there is a similar reciprocal relation between macroscopic neuronal states and their microscopic neuronal constituents: Global macroscopic mental states are constituted by neurons on the microscopic level, while simultaneously these mental states organize the activity of the individual neurons (Keijzer, 2001). On this view, the relatively slow and coordinated mass-activity of large populations of neurons can act as a constraint device that imposes order onto the firing patterns of the individual

neurons (see also Freeman, 1999a). The self-organizing neuronal states that underlie wake and sleep rhythms and synchrony patterns in the gamma band, for example, can thus be seen as harnessing the individual behavior of neurons into ordered patterns. These patterns in turn may exhibit properties that cannot be found at the micro-level, and new dynamic principles are thus required to understand these global patterns (Kelso, 1995).

Conscious will, as a self-steering component of self-organizing neuronal processes, may similarly impose restrictions on existing coordination patterns, thereby acting as a “dynamic constraint mechanism” that steers self-organized patterns of neuronal activity. On this view, it no longer makes any sense to pinpoint a “genuine” causal origin of a chain of events that leads to a willed action on one particular level of neural organization. That is, neither macroscopic mental states nor microscopic neural states should be seen as harboring the “ultimate causes” for all of our actions. In the penultimate section we shall relate our tentative interpretation to modern and classical views on conscious will.

9. The Efficacy of Conscious Will

William James (1890) maintained that primary reflexes and involuntary movements evoke complex impressions of looks, feelings, sounds and kinesthetics, which leave a multimodal image in our memory. The mere thought of a movement’s sensible effects then becomes sufficient to result in the instant automatic triggering of the necessary actions to achieve those effects. James (1890) argued that consciously willed acts differ from voluntary ones, as the former require an additional conscious element that has to intervene in order to overcome competing inhibitory forces. This element, which James termed the “volitional mandate” or “mental fiat,” has the power to overrule conflicting goals that keep some of our desired acts from execution.

In James’s (1890) classical ideo-motor theory, our “mental images” are directly linked to movement codes, which suggests that conscious thoughts and actions are fundamentally intertwined. Recent research subscribes to and elaborates on James’s introspective account, and shows that there is indeed evidence that cognitive codes of movement patterns, which make up a “motor program” and anticipated action effects, or goals, automatically form a bilateral association. When performing voluntary actions, we actively use this automatically formed relationship for goal formulation and selection (Elsner & Hommel, 2001). Research in developmental psychology shows that successful inhibition of competing actions is key in the development of consciously willed behavior (e.g., Diamond, 2002). However, there might be more to conscious will than mere inhibition and excitation of existing coordination dynamics. It is likely that it also operates as a particular kind of matching mechanism that compares goal-states with visual or kinesthetic feedback. The self-steering aspect of conscious will then come into play when there is a mismatch between external feedback and the internal goal state. Conscious will

can be seen to organize or self-steer behavioral patterns towards the goal-state by matching it with perception-action feedback (e.g. Lee, 2004).

However, consciousness only has access to global information in the brain, so it only has a global influence on our actions; specialized parallel systems deal with far more detailed and complex information (van de Grind & Lokhorst, 2001). Automatic control processes correct actions by means of simple feedforward processes when there is a mismatch between a present state and desired goal-state. When for some reason the goal-state is not reached, the representation of the desired state becomes available to higher-level control processes, which enables conscious access to the goal and awareness of the previous failure to reach that goal (Jeannerod, 2000). Conscious will can then override the unconscious systems and minimize the error between the goal-state and the sensory feedback, by reorganizing the behavioral patterns to reach the goal-state, thereby preventing us from getting “stuck in set” (Jeannerod, 1997).

In this article, we argued that ‘self-organization’ and ‘self-steering’ are useful notions to come to a better understanding of the cooperative neuronal mechanisms that underlie conscious will, although of course more research is needed to corroborate such a view. At least it provides us with a more natural alternative to the illusion-claim that allows modern scientific views to converge with good old-fashioned common sense.

10. Conclusion

In this paper, we discussed the empirical and theoretical validity concerning the illusion-claim, and made explicit some of its main philosophical presuppositions. By analyzing each of these three corner stones of the illusion-claim, it has become apparent that:

1. The empirical evidence that is seen as vindicating the illusion-claim is not persuasive. Rather, the evidence points in favor of the common sense idea that the conscious will plays an active role in behavior, although consciousness does not provide us with an all-access view to all the specific details that are involved in the underlying processes of these willed actions.
2. The theoretical incoherencies of the illusion-claim are rooted in the behaviorist input-output paradigm, which, by the use of strict, linear causal models, induces the category mistake of looking at neuronal activity as the cause for conscious will.
3. The philosophical position that the genuine causal processes that underlie conscious will only take place at the neuronal level lies at the roots of the illusion-claim. However, as we have argued, this theoretical position is based on faulty premises. Instead, we showed that with the help of notions such as ‘self-organization’, a better understanding of the integration mechanisms in the brain can be obtained, and with that a more natural picture of the causal efficacy of conscious will. This serves not only as to satisfy philosophical curiosity, but also for its pragmatic relevance: a better empirical and theoretical framework on the global integration mechanisms in brain functioning might also help advance the understanding and treatment of clinical disorders such as schizophrenia, and

Parkinson's disease (e.g., Jahanshahi & Frith, 1998). This marks the importance of conscious will as a serious object of scientific investigation, rather than a blind spot in orthodox neuroscience.

Acknowledgments

We thank William Bechtel, two anonymous reviewers, Fred Keijzer, Daan Franken, Jaap Bax, and all PCCP-members for their helpful comments and suggestions. This work is supported by NWO-VIDI research-grant no. 016.038.301.

Notes

- [1] Other relatively well-known examples of self-organization include (1) the hexagonal Rayleigh-Bénard cells, characteristic convection rolls that form in uniformly heated liquids (Jantsch, 1980), and (2) the Belousov-Zhabotinski reaction, a complex, prolonged chemical reaction.
- [2] Not all forms of self-organization lead to macroscopic order that displays novel so-called "emergent" system properties.
- [3] In modern philosophy of mind, there is still much discussion over the question of whether or not neuronal phenomena and psychological ones are really part of different levels (e.g., Kim, 1998). A view along these lines can be found in Keijzer and Schouten (2005).
- [4] The relation between self-organization and 'levels' is not a straightforward one. Not all macroscopic properties of self-organizing systems are generally regarded as being part of a new level of description.

References

- Anderson, J. R. (1995). *Cognitive psychology and its implications* (4th ed.). New York: W. H. Freeman & Co.
- Aulin, A. (1979). Law of requisite hierarchy. *Kybernetes*, 8, 259–266.
- Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. New York: Oxford University Press.
- Bechtel, W., & Craver, C. F. (2003). *Explaining top-down causation (away)*. Manuscript submitted for publication.
- Beer, S. (1974). *Designing freedom*. Toronto, Canada: Canadian Broadcasting Corporation.
- Bem, S., & Looren de Jong, H. (1997). *Theoretical issues in psychology*. London: Sage Publications Ltd.
- Ben-Jacob, E. (1997). From snowflake formation to growth of bacterial colonies, Part II: Cooperative formation of complex colonial patterns. *Contemporary Physics*, 38, 205–241.
- Bierman, D. J. (2003). Does consciousness collapse the wave function? *Mind and Matter*, 1, 45–57.
- Bressler, S. L. (2003). Cortical coordination dynamics and the disorganization syndrome in schizophrenia. *Neuropsychopharmacology*, 28, 25–39.
- Churchland, P. S. (1981a). On the alleged backwards referral of experiences and its relevance to the mind-body problem. *Philosophy of Science*, 48, 165–181.
- Churchland, P. S. (1981b). The timing of sensations: Reply to Libet. *Philosophy of Science*, 48, 492–497.
- Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
- Craver, C. F. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science Supplemental*, 69, 83–97.

- Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences*, *15*, 183–200.
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of Frontal Lobe Function* (pp. 466–503). London: Oxford University Press.
- Eccles, J. C., & Popper, K. R. (1977). *The self and its brain*. Berlin: Springer-Verlag.
- Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 229–240.
- Flanagan, O. (1997). Conscious inessentialism and the epiphenomenalist suspicion. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 357–373). London: MIT Press.
- Freeman, W. J. (1990). On the fallacy of assigning an origin to consciousness. In E. R. John (Ed.), *Machinery of the mind: Data, theory, and speculations about higher brain function* (pp. 14–26). Boston: Birkhauser.
- Freeman, W. J. (1999a). Consciousness, intentionality, and causality. *Journal of Consciousness Studies*, *6*, 143–172.
- Freeman, W. J. (1999b). Comparison of brain models for active vs. passive perception. *Information Sciences*, *116*, 97–107.
- Gazzaniga, M. S. (1998). *The mind's past*. Berkeley, CA: University of California Press.
- Glynn, I. M. (1990). Consciousness and time. *Nature*, *348*, 477–479.
- Gomes, G. (1998). The timing of conscious experience: A critical review and reinterpretation of Libet's research. *Consciousness and Cognition*, *7*, 559–595.
- Gomes, G. (2002). Problems in the timing of conscious experience. *Consciousness and Cognition*, *11*, 191–197.
- Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, *126*, 128–133.
- Haken, H. (1987). Synergetics: An approach to self-organization. In F. Yates (Ed.), *Self-organizing systems: The emergence of order* (pp. 417–437). New York: Plenum Press.
- Haken, H. (1995). Some basic concepts of synergetics with respect to multistability in perception, phase transitions and formation of meaning. In M. Stadler & P. Kruse (Eds.), *Ambiguity in mind and nature* (pp. 23–44). Berlin: Springer.
- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- Heylighen, F. (1989). Self-organization, emergence and the architecture of complexity. *Proceedings of the 1st European Conference on System Science* (pp. 23–32). Paris: AFCET.
- Heylighen, F. (2001). The science of self-organization and adaptivity. In L. D. Kiel (Ed.), *Knowledge Management, Organizational Intelligence and Learning, and Complexity*. In *The Encyclopedia of Life Support Systems* (pp. 253–280). Oxford, England: Eolss Publishers.
- Ho, M. W. (1996). The biology of free will. *Journal of Consciousness Studies*, *3*, 231–244.
- Hurley, S. (1998). *Consciousness in action*. Cambridge, MA: Harvard University Press.
- Hurley, S. (2001). Perception and action: Alternative views. *Synthese*, *129*, 3–40.
- Jahanshahi, M., & Frith, C. D. (1998). Willed action and its impairments. *Cognitive Neuropsychology*, *15*, 483–534.
- James, W. (1890). *Principles of psychology*. New York: Henry Holt.
- Jantsch, E. (1980). *The self-organizing universe*. Oxford, England: Pergamon Press.
- Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford, England: Blackwell.
- Jeannerod, M. (2000). Consciousness of action and self-consciousness. A cognitive neuroscience approach. In J. Roessler & N. Eilan (Eds.), *Agency and self awareness: Issues in philosophy and psychology* (pp. 71–86). Oxford, England: Oxford University Press.
- Joordens, S., van Duijn, M., & Spalek, T. M. (2002). When timing the mind one should also mind the timing: Biases in the measurement of voluntary actions. *Consciousness and Cognition*, *11*, 231–240.

- Jordan, J. S. (2003). The embodiment of intentionality. In W. Tsachacher & J. Dauwalder (Eds.), *The dynamical systems approach to cognition: Concepts and empirical paradigms based on self-organization, embodiment, and coordination dynamics* (pp. 201–228). London: World Scientific.
- Juarrero, A. (1999). *Dynamics in action: Intentional behavior as a complex system*. Cambridge, MA: MIT Press.
- Keijzer, F. A. (2001). *Representation and behavior*. Cambridge, MA: MIT Press.
- Keijzer, F. A. (2003). Self-steered self-organization. In W. Tsachacher & J. Dauwalder (Eds.), *The dynamical systems approach to cognition: Concepts and empirical paradigms based on self-organization, embodiment, and coordination dynamics* (pp. 243–260). London: World Scientific.
- Keijzer, F. A., & Schouten, M. (2005). *Embodied cognition and wide mental causation*. Manuscript submitted for publication.
- Keller, I., & Heckhausen, H. (1990). Readiness potentials preceding spontaneous motor acts: Voluntary versus involuntary control. *Electroencephalography and Clinical Neurophysiology*, 76, 351–361.
- Kelso, J. A. S. (1995). *Dynamic patterns*. Cambridge, MA: MIT Press.
- Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. Cambridge, MA: MIT Press.
- King, C. (1997). Quantum mechanics, chaos and the conscious brain. *Journal of Mind and Behavior*, 18, 155–170.
- Krippner, S., & Combs, A. (2000). Self-organization in the dreaming brain. *Journal of Mind and Behavior*, 21, 399–412.
- Lee, T. D. (2004). Intention in bimanual coordination performance and learning. In V. K. Jirsa & J. A. S. Kelso (Eds.), *Coordination dynamics: Issues and trends* (pp. 41–56). Berlin: Springer.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences*, 8, 529–566.
- Libet, B. (1991). Conscious vs. neuronal time. *Nature*, 352, 27–28.
- Libet, B., Alberts, W. W., Wright, E. W., Jr., Delattre, L. D., Levin, G., & Feinstein, B. (1964). Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex. *Journal of Neurophysiology*, 27, 546–578.
- Libet, B., Alberts, W. W., Wright, E. W., Jr., & Feinstein, B. (1967). Responses of human somatosensory cortex to stimuli below threshold for conscious sensation. *Science*, 158, 1597–1600.
- Libet, B., Gleason, C. A., Wright, E. W., Jr., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (Readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*, 106, 623–642.
- Libet, B., Wright, E. W., Jr., Feinstein, B., & Pearl, D. K. (1979). Subjective referral of the timing for a conscious sensory experience: A functional role for the somatosensory specific projection system in man. *Brain*, 102, 193–224.
- Libet, B., Wright, E. W., Jr., & Gleason, C. A. (1982). Readiness-potentials preceding unrestricted “spontaneous” vs. pre-planned voluntary acts. *Electroencephalography and Clinical Neurophysiology*, 54, 322–335.
- Nørretranders, T. (1998). *The user illusion: Cutting consciousness down to size*. New York: Penguin Books.
- Penrose, R. (1989). *The emperor’s new mind*. New York: Oxford University Press.
- Pockett, S. (2002). On subjective back-referral and how long it takes to become conscious of a stimulus: A reinterpretation of Libet’s data. *Consciousness and Cognition*, 11, 144–161.
- Port, R. F., & van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.
- Radin, D. (2003). Time-reversed human experience: Experimental evidence and implications. *Journal of Nonlocality and Remote Mental Interactions*, 2, 256–279.

- Roth, G. (2003). The interaction of cortex and basal ganglia in the control of voluntary actions. In S. Maasen, W. Prinz, & G. Roth (Eds.), *Voluntary action: Brains, minds, and sociality* (pp. 115–132). New York: Oxford University Press.
- van de Grind, W. A., & Lokhorst, G. J. C. (2001). Hersenen en bewustzijn: van pneuma tot grijze massa. In F. Wijnen & F. Verstraten (Eds.), *Het brein te kijk: verkenning van de cognitieve neurowetenschappen* (pp. 217–246). Lisse, The Netherlands: Swets en Zeitlinger.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Science*, 7, 65–69.
- Wegner, D. M., & Wheatley, T. P. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54, 480–492.
- Wolf, F. A. (1989). On the quantum physical theory of subjective antedating. *Journal of Theoretical Biology*, 136, 13–19.
- Zhu, J. (2003). Reclaiming volition: An alternative interpretation of Libet's experiment. *Journal of Consciousness Studies*, 10, 61–77.