Perspective Article

Which is Necessary for Cognition, “Free Will” or “Free Will Illusion”? 

Enrico Bignetti, MD*

Department of Veterinary Sciences, University of Parma, Via del Taglio 10, Parma 43126, Italy

SOFT SCIENCES

Over a long period of time, the issue concerning free will (FW) has been a much debated topic. In general, being fully aware of ourselves, we are convinced that our conscious will can freely control a “voluntary” action. However, natural events that cannot be predicted, such as earthquakes, cyclones, etc., frustrate our self-esteem and affects the process of thinking in men. To address this concern, by comparing the action decision-making of voluntary actions owing to our inability to escape such “supernatural” events, Searle1 was of the opinion that: “The persistence of the problem of traditional free will in philosophy seems to me something of a scandal”; nevertheless, many have studied this issue and many papers have been written focusing on this area; however, it appears that limited progress in related research has been made.

Among the inexplicable events occurring all the time in our daily lives, there are particular events that occur at random, known as the aleatory events. Throwing a dice is identified as an aleatory event in which a number is selected at random; the same occurs when determining the gender of an unborn child due to the random exchange of genetic material. Both the events discussed cannot be predicted in advance; these events are also known as “Markovian” since their occurrence is independent of the preceding results. Talking about these processes can be upsetting having realized that we do not have control over these events. In case of a dice with 6 faces, the degree of probability by which we can predict a number is very low (about 16%). A deterministic prediction is associated with an approximation close to 1, but, to do so, we should be able to widen our perspective and observe the innumerable causes that converge to determine a Markovian event; so that we may consider aleatory events as truly free, unconditioned events. The perception of FW in the minds of people is far from being related to this concept.

In the wide landscape of different religions and cultures moving from the West to the far East, one can find many modalities by which FW is defined; at least, two of these categories that stand at the mental antipodes have been defined: 1) the first one typically belongs to the monotheistic religions — Christianity, Judaism and Islam; 2) the second one corresponds to Taoism.

1) The social hierarchy decided by God which positions mankind as a subordinate to Him but dominant with respect to the universe, is a view common to the first category religions. This form of privilege offered to the mankind over the physical world is justified by the presumption of possessing a soul-inhabited self. Since the religious dialectic can grasp the theories of metaphysics by managing philosophical and psychological argumentations better than empirical science, belief in the soul and spiritual dimension of humans are naturally sustained and inflated. In principle, the idea of FW that pervades this context is a true FW, i.e., a FW that allows mankind to claim the sense of agency and to take the responsibility for any action-decision-making, but this decision can exert only a conditional freedom since, ultimately, the action outcomes are judged ex post facto by the laws imposed by a superior entity of infinite wisdom. Thus, having the perception of a soul-inhabited self is a prerequisite for a sense of guilt when the action decision-making surpasses the ethical constraints posed by the transcen-
dental rules. According to Freud8 people’s belief in God is like a form of illusion; in reality resting on the basis that the idea of God acts as a shortcut to the fulfillment of human wishes. In this regard is his well-known affirmation: “...we call a belief an illusion when a wish-fulfillment is a prominent factor in its motivation, and in doing we disregard its relations to reality, just as the illusion itself sets no store by verification...”. The more he was critical about religion, the less clearly he encountered the question of FW. The goal of Freud’s psychoanalysis was to bring repressed thoughts and feelings to consciousness; this is where FW seemed to be a necessary tool to place an individual at the centre of his own life, a perspective in disagreement with the deterministic view against religions. Freud’s hypothesis on FW remained ambiguous; after his death, two completely opposite interpretations for the goals of therapy were proposed: a) it allowed for the patient to develop a stronger ego; b) it led the subject to acknowledge his or her inability to satisfy the most basic desires. In the first case, FW functions as a real instrument in the subject’s hands, while, in the second case, FW is a mind confusing form of illusion.

2) Just as the orthodox Hindu Darśana (Advaita Vedanta, Shamkya, Yoga, etc.) and heterodox philosophies (Buddhism and Japanese Zen, etc.), Taoism primarily explores the human mind and its attributes.8,9 Though the final message conveyed is more radical: the mind is somehow an imperfect tool lying about or hiding the true nature of reality that is non-dual in nature, thus the definition of the self, negatively implies a dual mind-soul perspective. The realization of the non-dual self must be postponed following the enlightenment phase. Noticeably, a detailed study on epistemology presents the sages with the right to unveil the limits of the mind. The self-awareness of the individual's inconsistency and futility, on one side, and the impossibility to experience the divine (or vital principle), if any, through the chattering of the mind, on the other, urge the followers to put aside a sense of ego and deny seeking the truth using their minds. In contrast to many of the Darśanas cited above, Taoism is an atheistic philosophy which interprets our daily life experiences as a global net of synchronous series of events. The global interconnection between the events occurring simultaneously gives a sense of “timelessness” both to bad and good events. The synchronicity denies the classic deterministic cause-effect dependence between the actions of the past, present and the future; therefore, leading to the belief that life is a stream of tri-dimensional frames against which we cannot do anything, except becoming aware about it and, in principle, learning it upon deeper introspection. Tao is a philosophy much closely related to the concepts of modern physics than any other religion or philosophy.8,8 As a matter of fact, Taoism considers FW as a tool of no purpose and to this aim, one of the most important precepts of Taoism is to meditate in genuine non-action (Wu-Wei) in order to attain the right attitude to be able to avoid desires, i.e., to reach a state of happiness in clear contrast to the philosophy of Buddhism. In the Taoist concept of life, FW is a synonym for desires, falling into the trap of becoming a prey of time. A constant battle between time and our mind implies that some form of duality continues to control our life. A clear synthesis of the statements enunciated above can also be found in the book of the ancient Taoist Shu Ching10:

«The human mind is dangerous; its selfish tendencies lead to error and crime and its affinity with Tao is small»

At the highest spiritual level, a disciple refusing an attempt to showcase the true nature of divinity, may paradoxically assess a relationship with him (whatever he is), based on the knowledge of wisdom or gnosia. In contrary to the belief of Freud, Jung was positively attracted by the psychological role that religions (in particular of Far East) may play in human life.11 His psychological, philosophical and sociological interests towards the philosophies of the Far East led him to establish a deep connection with Chinese Taoism. His central idea of the unconscious was fundamentally influenced by Taoism, especially on the individual self that emerges from the a causal coincidence of events (later renamed with the notion of “synchronicity”). In this regard, he collaborated with Richard Wilhelm on the translation and the preface of the divinatory book “I King: or book of changes”.12 Through a new investigation of the unconscious, he formulated a complex concept of the self that stands on an existing parallelism between the inner and the outer realms of experience. The temporal coincidence between the theories of modern physics and the discovery of Eastern metaphysics played a significant role in the evolution of Jung’s thinking process from psychoanalysis to analytical psychology.

In conclusion, we may infer that the ego-sense and FW are psychological by-products of the mind that claims a territory of intervention. In the early lessons of Yoga, in one of the Darśanas cited above, the beginners are posed with the question “who are we? To find an answer, they try to become aware of nature and the extension of their bodies providing an opportunity to explore their existence both from inside and the outside. By means of this practice, the judgement is momentarily suspended; in particular they cannot vividly relate to the world inside or with the outside environment. The beginners’ perception of the individual self thus, tends to vanish.

In our opinion theistic religions facilitate the opposite, i.e., they reinforce personal identity and by-products of the mind such as a sense of ego and FW. On the basis of the historical, political, cultural and social environment, all religions tend to strongly support the dichotomy between the brain and mind riding the wave of dissent within the scientific community, so that we fear that the cultural prejudices in this context will bias any positivistic theory of mind even though supported with strong experimental validations. In contrast to this pessimistic view, although illuminating, controversial books by Robert Wright explain this perspective in much detail.13 Wright shows that, though starting from an erroneous primordial view on God, the evolution of our ideas leaves room for a transcendental perspective of divinity.

Cognitive sciences have raised the question of Self, FW and the sense of agency, in different contexts. Dennett’s
cognitive theory presents the analogy of self with “a centre of narrative gravity.” Self is not a physical entity, but a purely abstract object, a sort of folk-physics that is soberly known as phenomenology. Dennett observed that the insistence that consciousness must turn out to be something inexplicable, irreduncible, transcendental sometimes rises to a fever pitch.” For many, Dennett’s representation of mind has no self, no central witness and only an abstract for “centre of narrative gravity,” which appears to be a convenient fictional perception to avoid metaphysical or religious issues concerning decision-making. For Dennett, it is not a case of the emperor having no clothes, but rather that the clothes having no emperor.

Taken together, in Dennett’s view, the soul has been eliminated, which is a view that is in odds with the general way of thinking about the relationship between science and religion. A clear conclusion to be drawn here is that the soul is not visible through the brain’s eyes. If Dennett denies the self, then two important questions need to be asked: first, do we need a central agent in charge with the direct responsibility of decision-making, and second, is the agent really or only apparently free when exhibiting a purposeful action? Imagining and evaluating are considered as pre-mental states that correspond to the states of brain responsiveness that is already present when affected by events that trigger actions or intentions. This indicates a goal-directed process, which stands on antecedent determinants. This consideration can be elucidated with the example of the car and its engine. The engine will always be prompted to burn fuel and transform the fuel into motion. However, it is unclear as to who steers the car. Thus, the question of whether ego is a virtual driver of our mind becomes an issue that is far more interesting than that relative to the existence of FW. First, FW is a useless tool when deciding an intentional action that has already been predetermined and conditioned by antecedent determinants; and second, the existence of FW is fictitious if we consider that FW is a product of ego, which is, in itself a virtual driver. By chance, are we entering the duality of the new electronic era?

So what are the thoughts of people on FW? Are they in agreement with Dennett? According to a series of psychophysical tests reviewed by Nichols, several causes contribute to our senses of agency and responsibility in decision-making as well as the prediction of an action. As an example, if one perceives through feedback sensory signals that any body movement is carried out as predicted by one’s decision, then any form of related movement is considered as a voluntary action. According to another example, the same feeling of agency can be perceived if there is a time interval between an external cue and the action, as if an individual requires a proposed period of time necessary for thinking of a voluntary response to the cue. Interestingly, one does feel that one’s own actions are free, not necessarily those carried out by others. Moreover, if we analyse how the existence of FW is perceived by the population of different ages, the question becomes even more complex. To this regard, Nichols’ review reports that people might statistically shift from “determinism” to “indeterminism” and from “compatibilism” to “incompatibilism”, with different nuances.

However, on the basis of the experiments on the way people think about “I”, cognitive scientists as Dennett, infer that the idea of possessing FW is incompatible with “determinism”. People widely accept the fact that internal and external cues contribute to the sense of agency and strongly support the opinion of the self-being a free agent. The idea of being a deterministic machine does not lie in the belief of the people of being free agents; so that, inputs or programs of a computer or even a psychological mechanism cannot substitute for the mind in drawing conjectures, elaborating on thoughts and taking relevant decisions. This people’s way of thinking is referred to as the subjective “1st-person or perspective”(1st-PP), a well-rooted belief that cannot be undermined by any scientific or objective “3rd-person perspective” (3rd-PP) demonstration that the “sense of agency” is just like any other psychological mechanism that can be impaired by various neurological and psychiatric disorders. To this regard, Nichols’ review reports an interesting experiment that demonstrates that if a concrete case of a man killing his family, is presented to the subjects’ test, they tend to attribute full agency and responsibility to that man, even if affected by neurological disorders.

Incidentally, the possibility that people might consciously embrace a combination of the views of determinism and incompatibilism, would entail considerable risks of social and ethical nature, since no one could be considered morally responsible for his actions.

HARD SCIENCES

In biophysics, “integrate-and-fire” is the most widely accepted mechanism of neuronal computation of information processing. According to Rolls and Deco, this mechanism can explain how action decision-making may assume alternative directions in a probabilistic way. On considering that a random spiking assembly of neurons might resonate among a Poisson distribution of firing states; each state might be attractive for only one decision, i.e., the only one that fits at best in the attractor basin at that moment. In summary, if one knows which attractor basin is functioning during the stimulus arrival, one can predict in advance the kind of decision the mind will make. To this, the take-home-message of Rolls and Deco is that “the system has so many degrees of freedom that it operates effectively as a non-deterministic system (philosophers may wish to argue about different senses of the term deterministic, but it is being used here in a precise and quantitative way, which has been defined within the framework of stochastic neuro-dynamics).”

According to our knowledge, the way Rolls and Deco interpret the dynamics of their simulated systems, enter in conflict with thermodynamics and, in particular, with entropy implications. One of the most common way of measuring the energy involved in entropy loss is to calculate the number of states of equivalent energy through which a system can resonate, i.e., the interchangeable states that exhibit the same probability (see as an example the resonating formula of benzene that has the same energetic content); therefore, when different attractors induce
the same behavior they must exhibit the same energy content; conversely, when different attractors exhibit different strengths, the system behaviors will also be different. Moreover, the attractor with the longer life span will be also the one with the higher probability to intercept the arriving stimulus. Hence, the idea that system behaviours may vary from trial to trial, based on statistical fluctuations of the attractor within the states of the same energetic probability, is thermodynamically incorrect.

A similar argument was faced in a preceding paper, while discussing the paradox of the so-called “Buridan’s ass”. This paradox was used to support the thesis that, if the mind of a hungry ass is strictly “deterministic” or “mechanistic”, the ass will be unable to decide between the two perfectly identical sacks of hay, thus starving to death. This question could be answered by saying that the hard-deterministic brain does not exist; actually, the ass would not starve since a voluntary action is the outcome of the cooperation between probabilistic and deterministic activities that would help the ass overcome any obstacle in the best way possible. For the sake of simplicity, we may envisage the timeline of the ass’s behavior as separated in two sequential steps:

1) by the means of a probabilistic trial-and-error behaviour, the ass will find the first sack, recognize it and learn how to eat it;

2) since the efforts spent by the ass in the first step are finally rewarded, it will deterministically engrave in short-term memory the experienced paradigm, a paradigm that will be very easily replicated in front of the second sack.

The two-step sequence describes a typical learning-through-experience process of a probabilistic-deterministic brain; it is not by chance that the sequence reminds of the two sequentially described events predicted by Bayes’ theory of information processing. Rolls and Deco, tried to describe about half of the first sequence, i.e., action decision-making, on the basis of the probabilistic mechanism but entirely ignored the deterministic events underlying the cognitive processes. Moreover, Rolls and Deco do not consider at all why everybody inwardly manifests the idea of possessing FW: according to these authors, the so-called “freedom of choosing” mechanism is an impersonal mechanism attributed to a probabilistic oscillation between the states of an unconscious mind. If we ask people (including me) their opinion about what Rolls and Deco would mean by the above inference (i.e., the subjective 1st-person perspective in comparison with the objective 3rd-person perspective) probably they will reply saying: “nonsense! My FW means another thing! In fact, when my mind consciously decides what to do, it activates an inner speech that silently forewarns the actions that I will make, in a compulsory sequence” (see the role of inner speech in cognition elsewhere).

As far as it concerns the fundamental roles of reward and FW in voluntary actions, Schultz stresses on the theory that human intelligence depends on rewards while, facing the FW question, he claims that there exists no knowledge of its origin. If we assume as Kant said, that God is the only depository of true FW, the question for us is unsolvable. However, he tries to approach the question by observing that the unconscious mind presents alternative intentions in the conscious field, so that conscious FW can freely choose or veto one of these intentions. Some authors have proposed FW as the mechanism of being able to make choices; Schultz accepts the mechanistic definition of FW but he restricts the area of choices only to imaginable or available options. Among a long list of constraints posed by Schultz, rewards have the highest priority since they are a survival signal; different constraints may span from a deterministic world that nullify any form of FW (i.e., the actions are pre-determined in advance) to an unrestrained aleatory composition of stimuli that release any capricious choice of conscious FW. According to him, the truth lies in between, but this idea is absolutely risky for world survival. Hence, according to him, we can only choose between anything imaginable that can be enumerated in a list of restrictions e.g., education, personal experience, social pressure, emotions, etc. However, we firmly disagree with him in this regard; can we imagine the risk that democracy might encounter if someone decides on a list of options which everyone is expected to conform to?

Apart from this personal comment, Schultz’s FW is a conditional FW and people personally know what this means. However, the main difference is that people initially build up their own list of wishes (affordable or not) and then, by means of personal experience, they skip over those that are unattainable; paradoxically, this might be another incentive for reinforcing the idea of possessing FW.

It is quite obvious even to a non-professional that a decision is thought elaborated in response to the outer or inner stimuli to satisfy desires and ensure general well-being; in other words, it is a reaction to put ourselves again in equilibrium with the environment (“homeostasis” is the technical term widely used in chemistry, biology and psychology, etc.). Most probably, the inference that the idea of possessing FW is ontologically linked with the growth of our psyche might be generally accepted; but the existence of FW is still a much debated topic. Therefore, the question arises as to whether we can assess a benchmark on the issue that hard-sciences has not yet dealt with?

Since the pioneering work of Katz on quantal release of the neurotransmitter vesicles in the neuromuscular junction, we have learnt that each molecular and cellular component of the nervous system, generates aleatory responses to a stimulus. On the basis of this evidence, how can we rely on the information processing of our mind? In his famous book, Katz proposed that by integrating in space and time a sufficiently large number of active vesicles, the end plate potential may reach the threshold for a statistically reliable response. Generally speaking, the relationship between a stimulus and the appropriate response exhibits a causal and non-casual dependence, with a probabilistic-deterministic mechanism; this modality is functional in all highly-organized, biological and physical-chemical
systems. This evidence leads to the important inference that a neuron or a neuronal assembly exhibits a rational/deterministic modality only if it is appropriately stimulated from the outside. A first logical consequence is that the mind when totally isolated from its environment cannot work (even dreams are evoked by some sensations that are picked up from the paradoxical correlations of the memory archives). However, an even more relevant consequence is that the conscious mind cannot take any decision nor execute an action “on its own”, i.e., in the absence of any input. In other words, a logic and finalistic action is causally dependent on the nature of the outer stimuli. This inference is in contrast with the idea that the mind can elaborate on its own decisions autonomously and, ultimately against FW (see the discussion of Bignetti on the “free won’t” of Libet).

Is it possible that TBM might conciliate all the scientific positions?

How can we then acknowledge people’s belief in FW even though we know it is clearly an illusion? The apparently “nonsensical approach” might be resolved if we assume that, due to a psychological evolution of the abilities of the mind, the illusion of possessing FW plays a fundamental role in fostering cognitive processes. This hypothesis which is a compromise between the need of believing in FW of the 1st-person perspective and the evidence of FW nothingness of the 3rd-person scientific perspective, is enunciated in “The Bignetti Model” (TBM). Elaborated many years ago but formalized point by point only from 2014 onwards, TBM describes the sequence of events underlying the so-called “voluntary action” and the associated cognitive processes, in 5 compulsory steps:

1. The so-called “voluntary” action is decided and performed by the agent’s unconscious mind (UM) by the means of probabilistic responses to the inner and outer stimuli.

2. After a slight delay, the agent becomes aware of the ongoing actions through feedback signals (somatosensory, etc.) that are conveyed to the brain as a consequence of its performance. Thus, the agent’s conscious mind (CM) always lags behind unconscious activity.

3. Owing to this delay, the CM cannot understand the unconscious work that precedes awareness; thus, the CM erroneously believes it has freely decided the action. Though objectively false, this belief is subjectively perceived as true (FW illusion). It is so persistent and deep-rooted in the mind that the CM is unwilling to abandon it.

4. The FW illusion satisfies a psychological need to secure the arousal of the sense of agency (SoA) and of responsibility (SoR) of the action. Both SoA and SoR inevitably lead the CM to self-attributing reward or blame depending on the performance of actions and its outcome.

5. Both reward and blame are motivational incentives that foster learning and memory in the CM; the updating of knowledge will provide new information and the skill required for further action (restart from point 1).

To conclude, one might fear that the agent conscious but without FW, would not have the sense of morality; so important implications of TBM in ethics might be raised. Obviously an individual who is dangerous for himself and for the community must be accordingly dealt with. However, FW illusion is the basis for cognition, so the solution to the moral question stands on how moral values can be imprinted by the familial and the social environment. Therefore, the acceptance of TBM in principle would result in a much heavier implication on the rehabilitating methods in jails.

In TBM, every system, be it a neuron or a network of neurons or even the brain in-toto, can be considered as a “sparingly opened system” from a thermodynamic point of view since it can communicate with its environment. According to TBM, both sensory inputs from outside and “pieces of thought” coming from a nearby area, trigger an adequate response according to a paradigm that is picked up from memory archives, among those utilized in similar or identical experiences. To this regard, if our memory archives are like a “tabula rasa”, a trial-and-error mechanism will be activated to restore the homeostasis between the system considered and its environment (purely probabilistic mechanism); conversely, if we have repeatedly experienced the same stimuli many times in our life, the correct paradigm for an adequate response is already available in the memory, so that an automatic, instinctive response will be put in action (purely deterministic mechanism). When this model was tested by means of a classic press/no-press psychophysical task, a learning curve could be observed in response to trials with the same stimuli rehearsal, in which the probability of success hyperbolically rose from 0% up to 100%. Moreover, the learning process could be impaired by introducing distractors along the task which changed the nature of the paradigm. In accordance with TBM, these results show a probabilistic-deterministic cognitive process. In this process, the behavioural paradigm is progressively updated and contextually uploaded to memory archives for future actions, thus explaining the observed inter-trial priming effect. This ex-post updating was envisaged as the “updating factor” in Bayes’ equation applied to cognition. Then, according to TBM, the role of memory archives in information retrieval and upload carried out with UM and CM, respectively, seems to be fundamental in cognition. Obviously, in agreement with Schultz, blame or reward remain as the determinants in fostering a learning process; in fact, paradigms that are satisfied best by expectations are rated by a reward and thus, memorized.

The main question that should be addressed now is: “Who is conferring blame or a reward and who is to be blamed or rewarded?” According to TBM, every time an individual is hit by an outer stimulation, it is submersed by a psychological truth, perceived as the presence and absence of desires, of imagination, of an identity engendered by the inescapable will of illusion which is the root of existence. This illusion is fundamental since it projects on the explicit frame of the brain the idea
of being a true individual “I” with the responsibility of acting “in my name”. However, the Gestaltic arousal of possessing a Self is not sufficient to self-attribute the senses of agency and responsibility; in other words, the illusion of having a Self not necessarily implies the possession of FW; rather, it works the other way around, i.e., the illusion of FW (e.g., the illusion of controlling our daily lives) implies the false idea of an individual self independent of outer conditioning (i.e., a thermodynamically closed system capable of autonomous voluntary actions). To this regard, TBM proposes that this illusion is a pillar of human cognition. In fact, by means of this illusion, CM self-attributes the senses of agency and responsibility by which means one’s own actions are legitimized and might be ranked as a blame or a reward. This step is necessary to foster a learning-through-experience process. Actually, the “driver of the car” is non-existent; however with this trick, the mind deludes itself as if it were.

CONCLUSION

On the basis of soft-sciences we may infer that this illusion is psychologically connatural with the human mind (1st-person perspective). All of us can recognize ourselves in a popular definition of FW that recites “FW is an art for a particular sort of capacity for the rational agent to choose a course of action from among various alternatives”. Actually, the term “...rational...” might not have a clear meaning for all, but the verb “...choose...” is surely comprehensible and extremely attractive to our Psyche. According to soft-sciences, the idea of possessing FW is connatural with a subjective experience of conscious will [18,19,32,43-46]; then, the possibility that our Self might choose a voluntary action sounds appealing to TBM’s CM, the “explicit” part of the mind. This would imply that we may differentiate ourselves from a robot or a mechanical device.

Whereas, from hard-sciences, we may infer that FW must be an illusion (3rd-person perspective). On the other hand, we have evidences from hard-sciences that Free-will is not compatible with the intrinsic activity of neurons and neuronal networks as well, since these are not self-sufficient thinking systems in the absence of an environmental input that might cause an adequate and efficient response. As we inferred from hard-sciences, the theory about the continuity of mind claims that brain in the absence of an environment is nothing but a noisy electronic circuit. To this regard, Spivey [48] has published a systematic overview of how perception, cognition, and action are partially overlapping segments of one continuous mental flow.

TBM proposal seems to conciliate the two positions above: on the one hand, the mind adopts a psychological trick based on FW illusion, in order to activate a learning-through-experience circuit; on the other hand, it manages this circuit by means of a well-known biophysical computational mechanism, e.g., integrate-and-fire.

REFERENCES


