



## The Pleasure Principle as a Tool for Scientific Forecasting of Human Self-Evolution

Victor Yu. Argonov

Pacific Oceanological Institute of the Russian Academy of Sciences

[argonov@list.ru](mailto:argonov@list.ru)

*Journal of Evolution and Technology* - Vol. 24 Issue 2 – August 2014 - pgs 63-78

### Abstract

The pleasure principle (PP) may be a verifiable fundamental law of the living matter in the universe, and this law might then be used for forecasting human self-evolution. I do not pretend to “prove” PP, but argue that it must be regarded as a scientific hypothesis. Accordingly, I formulate verifiable and falsifiable postulates of PP. Their confirmation would allow the construction of a new scientific discipline, hedodynamics, that would be able to forecast the future development of human civilization and even the probable structure and psychology of other rational beings within the universe. I suggest basic hedodynamical scenarios for human (posthuman) civilization and argue that the discovery of the neural correlate of pleasure would provide more detailed forecasts. In particular, I demonstrate how the studies of pleasure mechanisms might predict the degree of aggression in future societies. I conclude that PP may become a scientific basis for fundamental, not phenomenological (based on extrapolations), future forecasting on large timescales.

### 1. Introduction

Transhumanist philosophers argue that the most important future changes in human civilization will be related to human self-evolution (modification of human biology by technological means). Contemporary humanity modifies mostly its environment, while future humans will increasingly have opportunities to modify their own bodies: to increase lifespan, intellect, physical strength, etc. (see Bostrom 2005 and references at provided in Wikipedia’s article on Transhumanism).<sup>1</sup> This view is shared by a rapidly increasing number of people, but the most problematic issue relates to the prospect of radical changes in human psychology (in particular, our motivations). In popular culture, this has mostly been considered in the dystopian tradition of Aldous Huxley’s *Brave New World* (1932). It is often implicitly supposed that radical modification of psychology is too dangerous for human civilization. In my view, however, it is inevitable. Moreover, the process has already begun.

In modern society, education, propaganda, hypnosis, advertising, psychoactive drugs, etc., are already available to manipulate (reprogram) our motivations in a sophisticated manner. Nonetheless, today's technologies have many limitations. For example, they cannot be employed to make deliberate changes to a person's preferences in art or sex, and there are no effective methods to motivate humans to perform hard work without offering some economic advantage or threatening violence. However, we can safely suppose that in the future novel methods of artificial stimulation of pleasure (including complete elimination of human suffering without damage to health; see Pearce 2005) and reprogramming of motivations will allow any human to get pleasure from physical labor, vigorous activity, scientific research, the creative arts, and even the experience of pain (Argonov 2008; Kosarev 1997).

Motivations establish the goals of human activity, while our technologies provide only tools to help achieve them. Therefore, self-evolution of human motivations is a key to the self-evolution of humankind. And in this paper I argue that the trajectory of self-evolution is predictable: human beings will change their motivations in accordance with the *pleasure principle* (PP). Thus, I will criticize today's dominant view that PP is merely a speculative philosophical doctrine, and will show that it might provide a unique scientific basis for fundamental forecasting of the distant future of human society, and even of interactions between different civilizations in the universe. In particular, I will show that an advanced neuroscience of pleasure and motivation mechanisms might predict the degree of aggression in self-evolving civilizations.

## 2. Glossary: basic terms and values

***Subjective phenomena*** is a general term describing anything that is subjectively felt/perceived by a sentient creature: i.e., sensations, emotions, perception of time, feeling of pleasure, thoughts, and volitional acts.

***Consciousness (phenomenal consciousness, subjective reality, mind)*** is the totality of all subjective phenomena of a single creature.

Note that in this paper, I use the word "consciousness" as a synonym of subjective reality. Therefore, in principle, even primitive animals may be conscious (or sentient). In this paper, "consciousness" is distinguished from "intellect" and other complex human activity.

***Volitional behavior*** is behavior that is subjectively perceived by the creature as its "own deliberate actions," taking place by its "own choice" or upon its "own decision."

A ***neural correlate*** of a subjective phenomenon is a physical (physiological) phenomenon that is unambiguously related to the subjective phenomenon and contains complete information about it. Neural correlates of subjective phenomena are the objective phenomena, the study of which provides knowledge of the subjective reality of another being: perception, dreams, hallucinations, emotions, and the feeling of pleasure, etc.

***Materialism*** is the doctrine that each subjective phenomenon has a neural correlate and that matter contains comprehensive information about consciousness (including information that some particular organism is conscious). Note that neural correlates of some subjective phenomena are still hypothetical. The search for the neural correlates is a crucial branch of today's experimental studies of consciousness (Metzinger 2000).

***Pleasure (happiness, hedonic tone) p*** is a measure of a creature's subjective satisfaction with the current state of its consciousness regardless of the specific factors that have caused it: the measure of desirability of this state. Hedonic tone may be negative (suffering) or positive (pleasure per se).

***Comfort q*** is a hypothetical physiological parameter, unambiguously determining a creature's volitional behavior according to Postulates 1 and 4, and being a neural correlate of pleasure *p* according to Postulate 2 (see Section 3.3).

**The pleasure principle (PP)** is the hypothesis that volitional behavior is always related to the maximization of pleasure  $p$ , or, at least, the avoidance of its negative values. Rigorous reformulation of PP is given in Section 3.3.

**Planning and forecasting time  $T$**  is a period of time during which an organism is able to predict the dynamics of  $q$  and plan its volitional behavior.

**Lifespan  $T_{\max}$**  is the lifespan of a living organism, the maximal possible value of the planning and forecasting time  $T$ .

**Utility  $Q$**  is the “integral pleasure”: a product of planning and forecasting time  $T$  and average expected comfort  $\bar{q}$  during this time:<sup>2</sup>

$$Q \equiv \bar{q}T \equiv \int_0^T q dt$$

**Hedodynamics** is a quantitative theory of individual behavior and societal development based on the pleasure principle.

**Evolutionary programming of motivations (EPM)** is an evolutionary mechanism that determines which stimuli an animal feels as pleasant and which as unpleasant.

**Self-evolution** is the artificial and purposeful self-modification of rational organisms with the use of technology (affecting intellect, physical abilities of the body, lifespan, motivations, etc.).

**Artificial programming of motivations (APM)** is an artificial and purposeful regulation that determines which stimuli an organism feels as pleasant and which as unpleasant (by means of propaganda, psychological technologies, chemical and surgical invasion in brain).

### 3. The pleasure principle as a scientific hypothesis

#### 3.1 Historical overview

One of the most problematic questions in the history of human civilization is the existence of the universal motivational mechanism, which is common to all animals and underlies all particular manifestations of volitional behavior. The first *hedonistic* conception that (1) a human being has a common motivational mechanism throughout their life and (2) this mechanism is based on a striving for pleasure (happiness) dates back as far as to Aristippos of Cyrene, a student of Socrates. Later in classical antiquity, the idea was developed by Epicurus and his followers, and in modern times, interest in the problem increased after works by the major British *utilitarians* (Bentham 1789; Mill 1863). Jeremy Bentham argued that (theoretically) any good is a measurable value and any human ideas about good and evil are based on the subjective experience of pleasant and unpleasant feelings. On this approach, other rational foundations of ethics are impossible. When deprived of the feeling of pleasure and suffering, a human being cannot have motivations; he or she will be unable to distinguish between good and evil, and will not praise or value even his or her own life. Utilitarian philosophers distinguish *ethical hedonism* (utilitarianism per se, in which pleasure is the sole ethical value), *psychological hedonism* (pleasure is the sole goal of activity), and *physiological hedonism* (pleasure is the sole goal of activity, and it has purely physiological basis). In this paper, I consider only physiological hedonism (and use the term “pleasure principle,” PP, for it).

In the nineteenth century, hedonistic philosophy gave an impetus to the development of social science and economic theory (Gossen 1854; Jevons 1871; Menger 1871; Walras 1874; Edgeworth 1881). PP was also accepted and popularized by the famous psychologists Gustav Fechner (1873) and Sigmund Freud (1920). The

discovery of the limbic system's pleasure centers, and experiments on their electric stimulation (Olds and Milner 1954; Delgado 1965; Sem-Jacobsen and Styri 1972), gave further support to the idea that the distinction of pleasant and unpleasant feelings and emotions is a fundamental property of all higher animals. However, the idea that all their motivations are based on pleasure (i.e., that the level of pleasure is a universal meta-motivation) did not obtain the status of a true scientific theory. Significant numbers of researchers still reject PP as an unscientific, senseless, useless, or unverifiable idea. In this paper, however, I propose to demonstrate that they are wrong. I will reformulate the PP as a group of postulates, each of which is amenable to scientific validation. I will then show that positive verification of PP postulates would provide the basis of a new scientific domain, *hedodynamics*, with powerful predictive abilities in some fields.

### 3.2 Requirements for the scientific formulation of the pleasure principle

Let us discuss typical conceptual problems relating to PP and outline the conditions for its true scientific formulation.

**Condition 1: non-tautology.** *PP should not be justified only by definitions. It should admit of a formal logical possibility of being either true or false.*

In the relevant literature, pleasure is often defined in a purely behaviorist manner. For example, Simonov (1966, 1986) has defined a positive emotion as an emotion that an animal “wants to strengthen, elongate, and repeat,” while negative emotion is emotion that “he strives to weaken, interrupt, and prevent.” This, however, does not say anything regarding the behavior. It solely offers a terminology to describe the behavior. Bertrand Russell (1946) was one of the famous critics of the pleasure principle as a mere truism. The definition “pleasure is that which the animal strives for” automatically proves the statement “the animal always strives for pleasure.” In order to overcome the problem, we can distinguish three phenomena (see Section 2): subjective feelings of pleasure (“*pleasure*”); objective neural conditions that produce pleasure (“*comfort*”); and objective behavior maintaining these conditions. Scientific study can then detect any correlation between these phenomena. It will be empirically demonstrable, not merely an artifact of the conceptual apparatus we are using.

**Condition 2: objectivity.** *PP should be scientifically verifiable or falsifiable. Therefore, it should be related to objective phenomena.*

Behaviorist definitions of pleasure use objective terminology, but they lead to tautology. Therefore, we must regard pleasure as a subjective phenomenon, even though this leads to the problem of measurement. PP states that volitional action maximizes pleasure. Both volition and pleasure are subjective phenomena. How are we to detect them in an objective study? In order to overcome the problem of subjectivity, in Section 3.3 I set out two meta-postulates that pleasure and volition have neural correlates. Rigorous verification of these meta-postulates is problematic, but they are in a good agreement with modern trends (Metzinger 2000).

**Condition 3: practical significance.** *The theory built on the postulates of PP (*hedodynamics*) must have some advantages when compared with the existing theories.*

A new theory's predictions must be more universal, more precise, or more easily obtained in technical terms than the predictions made by the existing psychology, evolution theory or neurophysiology. In Section 4, I will show that there is at least one field where PP might be a crucial predictive tool: *self-evolution* of humans and other hypothetical rational beings in the universe.

### 3.3 Postulates of the pleasure principle and hedodynamics

**Meta-postulate 1.** *Pleasure may be measured in the objective study of the brain (pleasure has a neural*

*correlate*).

**Meta-postulate 2.** *Conscious (volitional) and unconscious behavior may be distinguished in the objective study of the brain.*

General verification of Meta-postulates 1 and 2 is a hard interdisciplinary problem. They are definitely true in *materialism*, but the verification of materialism is also a hard problem. In practice, however, these postulates are implicitly assumed in most modern studies of pleasure mechanisms. In many particular experiments, verbal reports of participants may be enough to establish the correspondence between subjective and objective phenomena. Rigorously speaking, PP is a scientific hypothesis if its meta-postulates are true. Otherwise, the condition of objectivism (from Section 3.2) is not fulfilled, so PP is scientifically unverifiable.

Now let us introduce the postulates that are specifically related to PP (they are definitely verifiable, if the meta-postulates are true).

**Postulate 1: the existence of  $q$ .** *Any volitional actions of a conscious organism are determined by a single physiological parameter  $q$  (comfort), which may take negative and positive values. In its behavior, an organism either (1) strives to maximize the current value of  $q$ , or (2) only avoids negative current values of  $q$  (the particular option is to be determined by experiments).*

The first option is Freud's conventional "*pleasure principle*" (Freud 1920). The second is his early "*unpleasure principle*" (1899) or *negative hedonism*. According to the second option, if  $q$  falls below zero, then the organism changes its behavior, but if the value of  $q$  is positive, then the organism ignores more comfortable choices (compare Voltaire's "the best is the enemy of the good"). Postulate 1 can be interpreted as the existence of fundamental meta-motivation, which cannot be reprogrammed by any modification of a creature's brain.

**Postulate 2: the correlation between  $q$  and  $p$ .** *Comfort  $q$  is a neural correlate of pleasure. Negative values of  $q$  correspond to suffering and negative emotions ( $p < 0$ ), while positive values of  $q$  correspond to pleasure per se and positive emotions ( $p > 0$ ).*

I have stated Postulate 2 separately because one can theoretically imagine the alternative situation that the behavior is regulated by a single parameter (Postulate 1 is true), but this parameter does not correspond to pleasure (Postulate 2 is false).

**Postulate 3: the universality of  $q$ .** *Comfort has a similar physiological nature in all possible conscious creatures.*

At first glance, this postulate seems to be too strong. However, it is based, at least in materialism, on very simple philosophical considerations. Let us presume that materialism is true. Hence, matter contains full information about any subjective phenomenon of any conscious creature. One might nevertheless assume that the neural correlates of pleasure are significantly different in different organisms. For example, in one animal, pleasure might be determined by the concentration of dopamine, while in another it might be determined by the concentration of opioids. These two values cannot be reduced to the common scale. However, the feeling of pleasure is similar in both animals. Therefore, at least one subjective fact (the fact of the similarity of feelings) is not manifested in objective reality. This runs counter to materialism. In other words, Postulate 3 can be proved logically in materialism. We consider it to be a postulate only because we do not know whether materialism (as defined) is true.

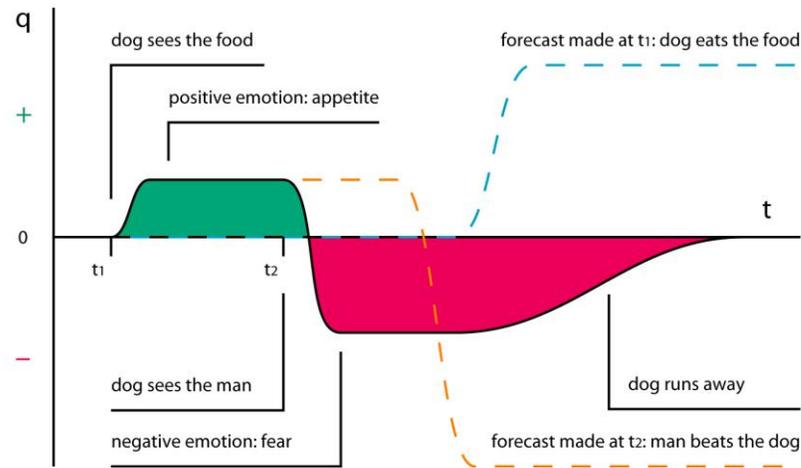


Fig. 1. Real and forecasted dynamics of dog's comfort. Dog, being in a neutral mood ( $q = 0$ ), sees the food. Predicted future  $q$  is positive, so the positive emotion develops. Then, dog sees the unpredicted factor: food is guarded by the man with a rod. Predicted future  $q$  goes negative, so dog feels fear (which is a negative emotion). Dog corrects behavior and runs away, so fear disappears

**Postulate 4: the correlation between  $q$  and  $Q$ .** *Intellect considers the maximization of utility  $Q$  (see definition and formula (1) in Section 2) as a dominating goal. There is a positive correlation between the predicted value of  $q$  and the current value of  $q$ . This stimulates an animal to maximize  $Q$ .*

In other words, anticipation of pleasure in the future causes positive emotion right now, while the negative prognosis causes negative emotion. The value of  $q$  is a cause of activity (Fechner 1873) and the value of  $Q$  is a goal. Postulate 4 is equivalent to Freud's "reality principle" (Freud 1920). Let us illustrate its action by a typical behavioral example. A dog approaches food and sees a man with a rod near it (Fig. 1). The dog thinks that the man might attack it, and the averaged expected comfort  $q$  (as well as  $Q$ ) in the near-term future is negative. To prevent this outcome, the dog's brain develops the negative emotion of fear. As a result, current  $q$  also turns negative and the dog runs away. The existence of such a stimulation mechanism is obvious. It should be noted, however, that Postulate 4 is a stronger statement. It not only declares the possibility of correlation between  $q$  and  $Q$  but also directly relates it to intellect. Postulate 4 is especially important in the study of human beings, because their planning and forecasting time  $T$  is comparable with their lifespan  $T_{\max}$ . Therefore, if Postulate 4 is true, then at least some humans strive for survival not only because of the instinctive fear of death, but also because they strive for the maximization of  $Q$  (if average expected  $q$  is positive). Suppression of instincts (by totalitarian propaganda or by technical methods) can't eliminate this factor. According to Postulate 4, only suppression of human comfort or intellect might force an individual human being to sacrifice his life for "great ideas."

### 3.4. Hypotheses relating to the neural correlate of pleasure: overview

Rigorous verification of PP postulates can become possible only after the discovery of the neural correlate of pleasure that has been intensively searched for since the mid-twentieth century. In this section, I will give a short overview of discovered mechanisms and declared hypotheses.

Most of authors relate pleasure to particular neurochemical processes. Earlier neurochemical theories, formed in the mid-twentieth century, related pleasure to the action of *catecholamines* (such as *dopamine* and *noradrenaline*) that regulate the passage of signals in the neural network. Famous experiments by James Olds, Peter Milner and others (Olds and Milner 1954; Delgado 1965) showed that electrical stimulation of some brain regions (responsible for the production of catecholamines) radically changes an animal's behavior: the animal feels an overwhelming motivation to continue stimulation infinitely. It was hypothesized that animals feel strong

pleasure. Therefore, the relevant brain regions were called “*pleasure centers*.”

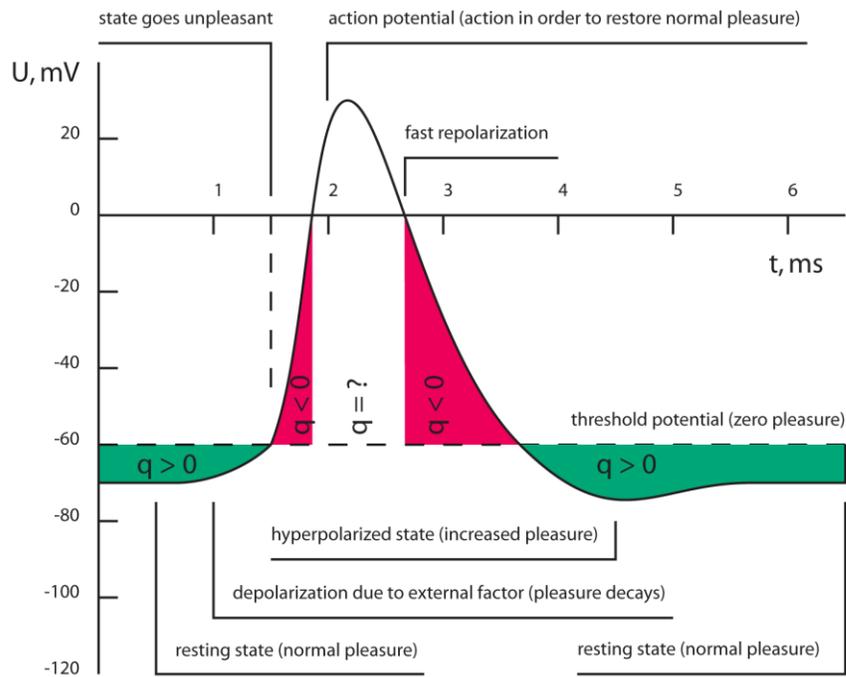
However, further experiments on humans have shown that neither their stimulation (Sem-Jacobsen and Styri 1972) nor the use of dopaminergic drugs produces rather strong pleasure. They produce obvious stimulation effects, but their hedonic effect seems to be a rather mistaken interpretation (Berridge and Robinson 1998). Catecholamines are probably responsible not for pleasure but for the estimation of its future value (Schultz et al. 1997, Borgkvist et al. 2007). Today, many authors suggest that not catecholamines but *opioids* must be regarded as a “real substance of pleasure.” Using opiate drugs, humans feel pleasure that is much stronger than was reported in old “pleasure center” stimulation experiments. New experiments (Borgkvist et al. 2007) have shown that opioids influence mice even after “switching off” major dopamine mechanisms. Kent Berridge discovered two small *hedonic “hotspots”* in a cubic millimeter of the rostradorsal region of the medial shell of the *nucleus accumbens* and in the *ventral pallidum* (Kringelbach and Berridge 2006; Smith and Berridge 2007) and showed that they produce much more clear hedonistic effect than Olds’s “pleasure centers”. He hypothesizes that opioid  $\mu$ -receptors in these “hotspots” might be real “receptors of pleasure” of the brain.

Neurochemical theories of pleasure have one serious drawback: they do not clarify the particular physiological parameter that might be unambiguously related to pleasure. There is always a chain of effects (generation of substances, their action on receptors, intracellular processes, regulation of signal propagation), and it is unknown which of them “really produces” pleasure. In response to this problem, some authors have suggested more fundamental hypotheses.

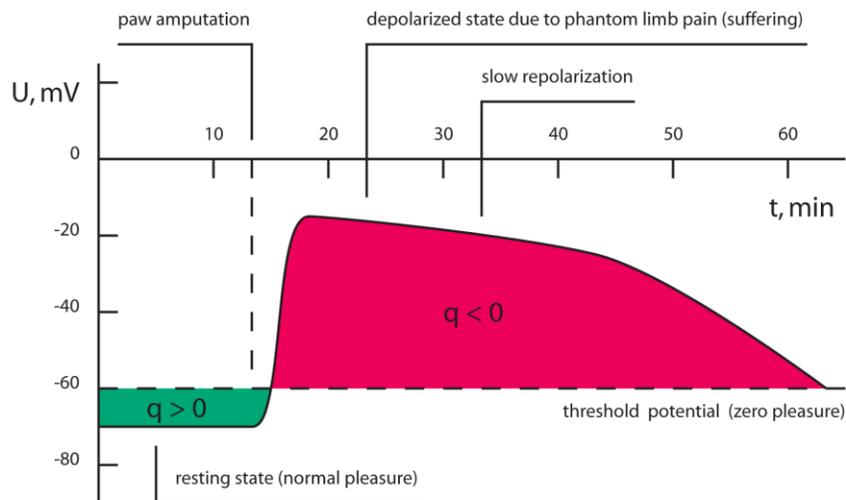
Karl Friston (2010) states that an animal always strives to maximize the predictability of its environment. The degree of predictability is estimated by computing the information analogue of *free energy*, a thermodynamic parameter introduced by Hermann von Helmholtz. According to Friston’s theory, the brain estimates the value of informational free energy (in particular, for sensory data) and tries to minimize it in order to increase predictability. Note that informational free energy is not physical free energy (which is related to metabolism and usually maximized, not minimized, by living organisms). A recent study (Friston et al. 2010) has shown that the level of dopamine might also be related to sensory prediction error and free energy.

Several authors relate pleasure to intracellular processes. This view matches well with modern tendencies to depart from old ideas of the neuron as a “mere trigger.” Many contemporary works characterize the neuron as a small but independent organism with its own memory and computing power (Koch and Segev 2000) or even individual consciousness (Edwards 2005; Sevush 2006). Soviet scientists Vladimir Shvyrkov (1995) and Yury Alexandrov (1999) developed a theory that each neuron has its own genetically preset motivations, demanding saturation with particular metabolites. Each neuron “wants” these substances and generates signals until other cells provide them. Cells that are not able to take required metabolites commit apoptosis (Alexandrov 1999), which is an analogue of suicide motivated by the absence of joy of life.

It might be hypothesized that human pleasure is a sense of “satiety” of an individual’s neurons. Sergey Murik (2006) suggests an alternative hypothesis that  $q$  is related not to cell metabolism but to cellular *membrane polarization* (its static level). Normally, the in-cell medium has a negative charge. As a result, there is a polarization between the intracellular and external medium (of the order of  $-70$  mV). The level of static polarization in animal and plant cells is related to the favorability of their metabolic condition. Decreased polarization (*depolarization*) indicates the presence of negative factors (Blake et al. 1988; Pyatygin et al. 2006). In normal conditions, in neurons, depolarization causes signal generation (Fig. 2a), and polarization restores to the normal or increased level (*hyperpolarization*). If an unpleasant factor presents too long, the cell spends all its energy, becomes exhausted and decays. Murik associates hyperpolarization directly with pleasant emotions, and depolarization with unpleasant emotions. This idea maps with the results of experiments showing a correlation between: (1) negative emotional behavior and brain polarization with electrodes (Murik 1996); (2) neural depolarization and phantom-limb pain (Wu et al. 2005) (Fig. 2b); and, (3) neural hyperpolarization and the effects of opiates (Duan et al. 1990; Chieng and Christie 1994).



## a: action potential



## b: phantom limb pain

Fig. 2. Electric processes in neurons and their hypothetical hedonistic interpretation: (a) generation of action potential as a reaction to membrane depolarization (unpleasant factor?); (b) prolonged membrane depolarization due to phantom limb pain (Wu et. al. 2005)

Today, the search for a neural correlate of pleasure continues, and there is no universally accepted theory. We have some methods of chemical control of both the feeling of comfort (with opioids and some other drugs) and the feeling of anticipated utility (with dopamine drugs) in mammals, but we have not identified the detailed underlying mechanisms of these effects. However, the most realistic assumption is that the fundamental neural correlate of pleasure is somehow related to a creature's survival at cellular or physical level. Most likely,

pleasure is a kind of indicator of the fundamental “vitality” of a neural system (expressed in terms of cellular membrane polarization, cell nutrition, or thermodynamic parameters).

#### 4. Self-evolutionary applications of hedodynamics

In this section, I will outline the basic evolutionary and self-evolutionary predictions of PP (assuming that it is a true principle). Let us presume that the PP postulates identified above (Section 3.3) are true and discuss their consequences.

##### 4.1 Hedodynamics and biological evolution

Generalizing several hypotheses from Section 3.4, we can base our analysis on the assumption that  $q$  is a microscopic indicator of a system’s vitality, and, at least for simple organisms, that it is closely related to their ability to survive in the nearest future. Most primitive organisms only maximize  $q$  or maintain it at the non-negative level ( $T=0$ ). For example, the organism might be stimulated by a deficiency of nutrients and stop acting when their normal level is restored (Fig. 3 shows typical behavior of bacteria depending on nutrient concentration). Since we do not know whether these organisms have subjective feelings, this might not be PP in a rigorous sense; it is, however, at least its basic physiological analogue (and some authors do suppose that each cell literally has subjective feelings (Edwards 2005; Sevush 2006)). Such behavior is primitive: it is based on a single motivation that is always the same (See Fig. 4a and Table 1). It is not optimal for survival of the individual or population.

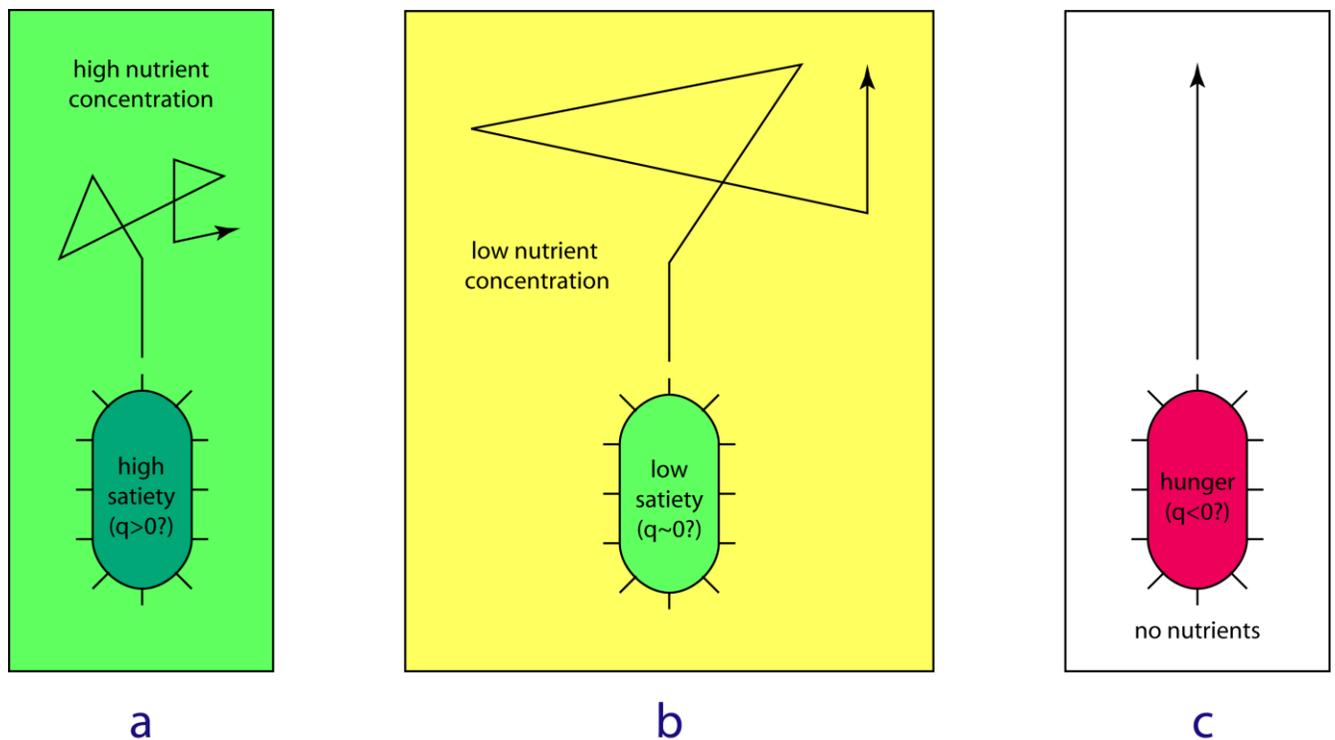


Fig. 3. Radical change in typical bacterial motion due to hunger: (a), (b) when the environment contains nutrients, bacterium moves along random trajectory and the frequency of its random turns depends on the satiety (pleasure?); (c) when no nutrients available, bacterium goes hungry and begin to move in a constant direction in order to find food (restore pleasure?) as quickly as possible

Accordingly, animals with complex neural systems display more complex behavior. They have numerous motivations other than nutrition: avoiding danger, accumulation of food supplies, building nests, sexual

behavior, etc. They are able to feel pleasure and displeasure in the absence of factors that are important for primitive organisms.

How did this become possible? According to Shvyrkov (1995) and Alexandrov (1999), it might be supposed that cells, because of certain mutations, require new sorts of metabolites that were not vital for ancient predecessors; therefore, the motivations of the whole organism change. A simpler alternative is that other processes in higher animals “deceive” consciousness, depriving some neurons of nutrients in order to achieve their activity. Today, exact microscopic motivational mechanisms are unknown, but, anyway, complex organisms are able to feel pleasure and displeasure from factors that are not directly related to their nutrition and/or survival in a short timeframe (see Fig. 4b and Table 1). Therefore, during the process of biological evolution, numerous programmable motivations replaced a single non-programmable motivation. *Evolutionary programming of motivations (EPM)* (see Section 2 for definition) became possible.

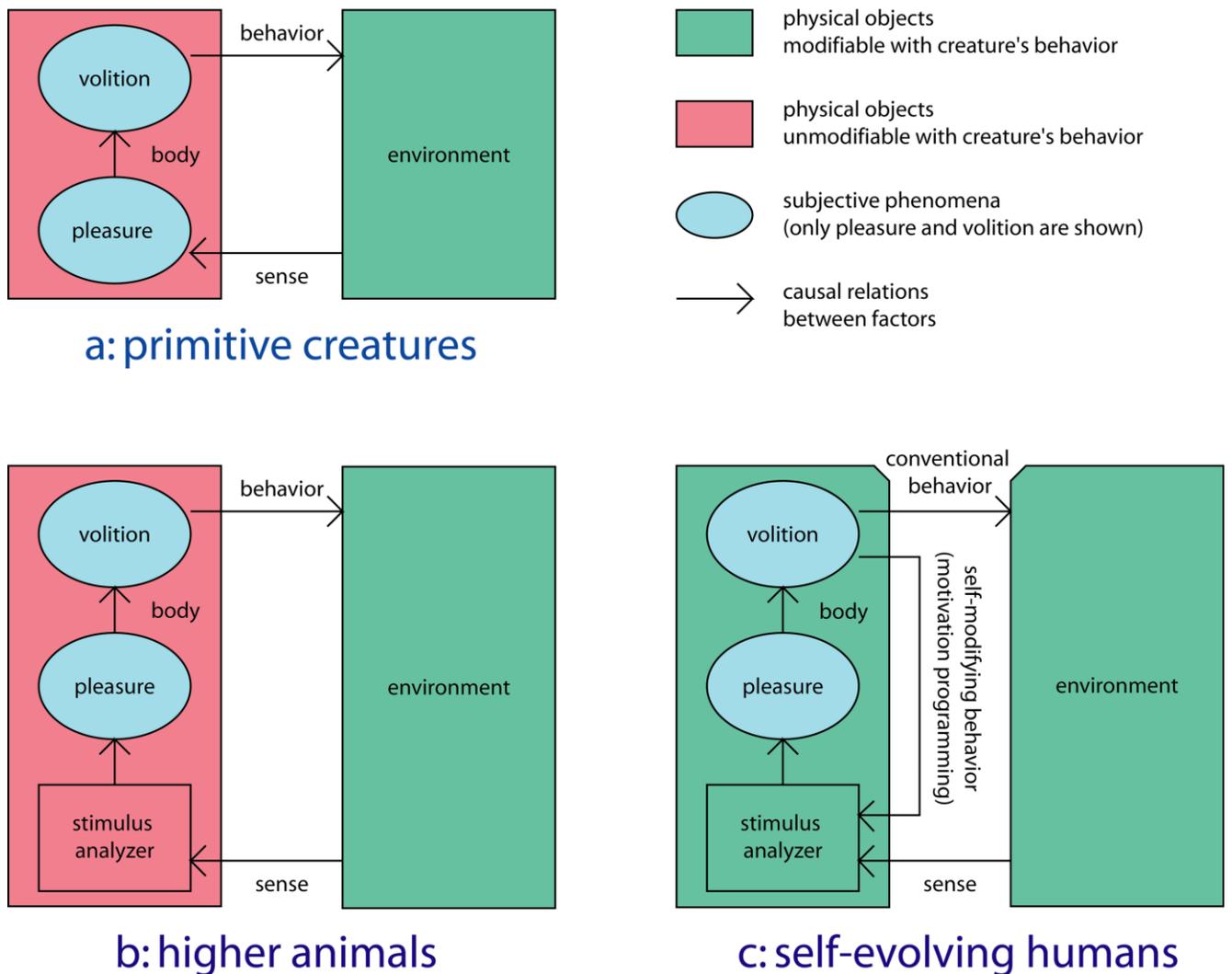


Fig. 4. Hypothetical evolution of pleasure mechanisms (assuming sentience of primitive creatures): (a) in primitive creatures, pleasure encodes some factor, important for life (for example, satiety); the goal of behavior is pleasure; (b) in higher animals, EPM-based stimulus analyzer determines which stimuli should be interpreted as negative or positive pleasure (suggesting behavior programs rewarded by pleasure); behavior is a “hunt” for particular pleasant factors suggested by analyzer; (c) in self-evolving humans, stimulus analyzer may be reprogrammed artificially, therefore, it may be considered as a part of modifiable environment, and schemes (a) and (c) are equivalent; the goal of behavior is pleasure itself, not particular pleasant factors. See also Table 1

What might hedodynamics say about the evolution of motivations and the structure of particular organisms in wild nature? It might say almost nothing. Primitive creatures can be described by their striving for maximization of some vital parameter (such as amount of nutrients), and the term “comfort” would be just another word for the same thing. In higher animals, multiple motivations emerge, but they are subordinate to natural selection. In the framework of EPM, PP is not an independent factor; it just “serves” evolutionary “needs.” It follows that hedodynamics would add nothing to existing theory of evolution. However, the situation changes radically when we consider the development of rational beings.

#### 4.2 Hedodynamics and self-evolution of rational beings

As a result of their highly developed intellects, humans have specific behavioral features that were impossible in earlier animals. Let us outline two of them. First, the human horizon of forecasting and planning  $T$  is comparable with the individual lifespan  $T_{\max}$ . As a result, humans (according to Postulate 4) try to prolong life as long as average comfort is positive (or commit suicide, if it is negative). Second, the appearance of technological progress radically changes the environment and begins to change the organism itself: i.e. *self-evolution* begins, one of the trends of which is the *artificial programming of motivations (APM)* (see Section 2 for definition).

Non-human animals act according to existing motivations, but do not try to change them. Humans act according to existing motivations, but also think about their “goodness” (one of the characteristic human motivations is to think about motivations). Human beings typically criticize some motivations as dangerous, egoistic, bad for health, etc. Modern society uses propaganda, education, drugs, and other tools to modify human motivations in a desired manner. Today’s methods of APM have serious limitations (as touched upon in the Introduction), but they are developing rapidly. In the future, the emergence of new powerful APM methods seems inevitable. While today’s APM methods are based mostly on socio-psychological manipulation, future APM may be based on direct stimulation or modification of the brain.

One simple approach to radical APM is surgical correction or destruction of certain brain centers. Such possibilities have been reported in many experiments with animals and humans. Early experiments were related to macroscopic brain invasion accompanied by multi-aspect changes in behavior. For example (in Kluever and Bucy 1938), bilateral removal of the amygdala and hippocampus in rhesus monkeys changed some of their preferences (specifically, they ceased to fear snakes), but these effects were accompanied by general brain disorders. In the second half of the twentieth century, more precise (and less destructive) methods were developed. Some of them are used today for surgical treatment of drug addiction (Kringelbach and Berridge 2006). However, these methods are also destructive. In future, a much more universal (and non-destructive) APM approach might be provided by wearable and programmable machines that observe human activity and stimulate pleasure (electrically, chemically, or by more sophisticated and less invasive methods) in desired situations (Argonov 2008; Kosarev 1997). For example, a machine might be programmed to stimulate pleasure every time an individual does physical exercises or studies for a university examination. Today, this APM approach is limited by two factors: (1) electronic devices are able to distinguish only the simplest patterns of human behavior and life situations; (2) absolutely safe pleasure stimulation is still a problem. The first limitation seems to be purely technical. The second is caused by our lack of knowledge of pleasure mechanisms. Therefore, both limitations seem to be temporary. In the future, there will be almost unlimited possibilities.

Once developed, universal APM techniques will radically change humankind. New APM methods might be the object of ethical criticism and social conflicts, but it seems almost impossible to prohibit them totally. It should be noted that some contemporary APM methods (such as surgical treatment of drug addiction) also attract ethical controversy, yet they are used in practice. After the emergence of radical APM methods, many novel social problems might appear. In particular, some people might try to use permanent pleasure stimulation (therefore, they will have no motivations for any activity). On the other hand, some authorities might try to force people to program themselves in a specific manner, thus building an “ideal” totalitarian society (these problems and their likely solutions are discussed in detail in Argonov 2008). Anyway, as a rule, according to Postulate 4, rational

people will try to modify themselves in the following two directions:

**Direction 1.** *Maximization of lifespan  $T_{\max}$*  (modification of body in order to raise its reliability, eliminate dangerous motivations, and construct useful motivations).

**Direction 2.** *Maximization of average comfort  $\bar{q}$*  (modification of brain in order to increase maximal possible level of pleasure and stimulation of pleasure in already available range).

In both liberal and totalitarian societies, there would be at least some people with enough political power to follow this optimal strategy (in agreement or in conflict with other parts of society that might be forced to choose non-optimal motivations). And this would radically change civilizational development.

With advanced forms of APM, human beings will obtain the opportunity to control  $q$  directly, much like primitive organisms (compare Figs. 4a and c). Therefore, pleasure will increasingly become the dominating factor in human development, and hedodynamics will turn into a powerful prognostic tool. More detailed scenarios of development depend on the nature of the neural correlate of pleasure (unknown today). In particular, the nature of this correlate might determine the degree of aggression in self-evolving societies. Let us examine two basic scenarios.

#### 4.3 Basic hedodynamical scenarios of civilization development

Once human civilization is completely transformed, becoming a self-evolving system (i.e. when its development depends mostly on natural resources and politics, rather than on technological limitations), the most likely scenario for further development depends strongly on the nature of the neural correlate of pleasure. In particular, the most important unknown factor is the existence, or otherwise, of fundamental limits of pleasure. There are two major alternatives:

**Alternative 1. Non-local mechanisms of pleasure.** *Comfort  $q$  is determined by the integral material values* (volume, mass, energy, electric charge, etc.). For example,  $q$  might be determined by the total amount of opioids or nutrients in brain. Therefore, comfort  $q$  can be maximized nearly infinitely via technical methods (with the consumption of resources and/or the increase of the body), and the *extensive* scenario for civilization is probable.

**Alternative 2. Local mechanisms of pleasure.** *Comfort  $q$  is determined by the average (differential) material values* (density, concentration, frequency, etc.). For example,  $q$  might be determined by the concentration of opioids or nutrients, or by the rate of computations. Therefore, comfort  $q$  might have some natural limits, and the *intensive* scenario for civilization is probable.

**Extensive scenario.** If comfort  $q$  is determined by integral, summed-up factors, the increase of the maximum level of comfort  $q$  requires accumulation of resources (in particular, an increase in the mass and volume of the organism). If the human is able to have more intense pleasure than the mouse, then the further maximization of body size might be preferable. An individual who has accumulated more resources can become happier, and this factor might induce violent conflicts (more violent than those induced by money or natural resources in the modern world). Some individuals might achieve social agreement for the sake of security; however, the global trend will be the decay of human populations, accompanied by growth in the individual power of the remaining people. If individuals are able to merge their consciousnesses without the destruction of each other, then an expansion can take place without conflicts. In the far future, super-organisms like Solaris Ocean (Lem 1961) or Utilitronium (Pearce 2014) might appear, and the total division of all available matter and/or energy in the universe among cosmic individuals is a plausible scenario. In any event, the long-term self-evolutionary dynamics will involve two parameters (because both  $T_{\max}$  and  $\bar{q}$  remain variables).

**Intensive scenario.** If comfort  $q$  is determined by average characteristics such as energy density, concentrations of substances, rate of computations per unit of volume, etc., then, at early periods of self-evolution, humans will try to increase average  $q$ . Sooner or later, however, in agreement with contemporary physics,  $q$  will reach saturation. Any further increase of  $\bar{q}$  will be impossible or unprofitable for the maximization of  $Q$ . Hence, according to Postulate 4, humans will try to maximize only a single parameter, lifespan  $T_{\max}$  (civilization returns to the one-parameter strategy, similar to that realized in the process of biological evolution). Elsewhere (Argonov 2008), I have discussed the situation in detail, arguing that the most anti-utopian political systems must be unstable. Typical individuals will subordinate all other motivations to their own survival.

In the intensive scenario, conflicts for resources would not be so prominent as in the extensive scenario, because their accumulation would not guarantee a proportional lifespan increase (while conflict itself might decrease it). Humans (feeling maximal happiness almost all the time) would need no specific entertainments. Most likely, they would program themselves for peaceful activities related to medicine, security, energy production, and so on.

Type of evolution	Groups of organisms	Motivational mechanisms	Dominating behavior strategy
Biological evolution (accidental mutability and natural selection)	Unicellular organisms and primitive metazoans	Single fixed motivation	Maximization of $q$ or maintaining non-negative $q$
	Higher metazoans	Evolutionary programming of motivations	Maximization of the evolutionary success of the species and its progeny
Self-evolution (purposeful modification)	Posthumans before the saturation of $q$	Artificial programming of motivations	Maximization of $Q = \bar{q} T_{\max}$
	Posthumans after the saturation of $q$ (intensive scenario)		Maximization of $T_{\max}$

Table 1. Hedodynamical periodization of the evolution. Here  $q$  is comfort of an organism;  $\bar{q}$  is its average value during life;  $T_{\max}$  is lifespan; and  $Q$  is utility (see Section 2 for details)

## 5. Conclusion

In this paper, I have shown that the *pleasure principle (PP)* might be not only a fundamental, verifiable law of the universe, but also a unique predictive tool for the distant future of humankind. Today's forecasts are based mostly on extrapolations of global trends ("*macroscopic,*" or "*phenomenological,*" methods). Such methods demonstrate good results in many fields from demography to computer science, but all of them are applicable to the near future only. Forecasting of civilizational development over centuries and millennia is impossible without alternative methods that might be called "*microscopic,*" or "*fundamental.*" Such methods must be based on fundamental laws of civilizational development. In this paper, I have argued that PP might be such a law.

I have proposed a group of PP postulates, each of which requires scientific validation (and is formulated to acknowledge that need in principle). The cornerstone of PP is pleasure's *neural correlate*, a physiological parameter having a one-to-one correspondence with the level of pleasure subjectively experienced by the creature concerned. The discovery of this parameter, and with it confirmation of PP postulates, would lead to the establishment of a new quantitative theory, called "*hedodynamics.*" PP might be especially useful for describing the behavior of rational beings, whereas in wildlife it is subordinate to natural selection and, thus, does not add anything new to our knowledge. When a civilization of rational beings develops technologies of purposeful

modification of their biology (enters the stage of *self-evolution*), hedodynamics becomes a useful forecasting tool, able to predict not only the psychology of new creatures, but also their structure and the global tendencies of their civilization's development.

In the future, detailed scientific knowledge of fundamental pleasure mechanisms (common for humans and animals) may be used for the construction of detailed forecasts. In particular, it may become possible to determine, whether the future self-evolutionary scenario is extensive (unlimited aggressive increase of bodies of individuals) or intensive (peaceful activity aimed at lifespan extension). These surprising applications of neuroscientific knowledge illustrate the extreme importance of interdisciplinary studies in modern science.

## Notes

1. See <http://en.wikipedia.org/wiki/Transhumanism> (last accessed June 14, 2014). Like all Wikipedia articles, this undergoes a process of continual editing.

2. In this paper, I neglect the difference between objective and subjective time (considering  $t$  as objective time). However, in practice, the *subjective time rate*  $\gamma = d\tau / dt$  (a ratio between the “speeds” of subjective and objective times) is not a constant. The value of  $\gamma$  may be determined by some periodic neural processes, for example, by *gamma synchrony* (Crick 1994). During sleep,  $\gamma$  decreases, while during intense action it increases. Drugs may also change the value of  $\gamma$ . In order to generalize our approach to the case of variable subjective time rates, Postulate 2 must be reformulated as follows: “*Physiological comfort  $q$  is a neural correlate of the product of the pleasure  $p$  and the subjective time rate  $\gamma$ . That is:  $q = \gamma p$ .*” Other postulates may remain unchanged, because  $\gamma$  has a positive value.

## References

Alexandrov, Yu. I. 1999. Psychophysiological regularities of the dynamics of individual experience and the “stream of consciousness.” In *Neuronal bases and psychological aspects of consciousness*, ed. C. Teddei-Ferretti and C. Musio, 201–235. Singapore, N.Y., London, and Hong-Kong: World Scientific.

Argonov, V. 2008. Artificial programming of human motivations: A way to degradation or rapid development? *Questions of Philosophy* (in Russian) 12: 22–37.

Available online <http://transhumanism-russia.ru/content/view/392/144> (accessed June 6, 2014).

Bentham, J. 1789. *An introduction to the principles of morals and legislation*. London: T. Payne.

Berridge, K. C. and T. E. Robinson. 1998. What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews* 28: 309–369.

Blake, K., W. T. Clusin, M. R. Franz, and N. A. Smith. 1988. Mechanism of depolarization in the ischaemic dog heart: Discrepancy between T-Q potentials and potassium accumulation. *Journal of Physiology-London* 397: 307–330.

Borgkvist, A., A. Usiello, P. Greengard, and G. Fisone. 2007. Activation of the cAMP/PKA/DARPP-32 signaling pathway is required for morphine psychomotor stimulation but not for morphine reward. *Neuropsychopharmacology* 32: 1995–2003.

Bostrom, N. 2005. A history of transhumanist thought. *Journal of Evolution and Technology* 14(3): 1–25. Available online <http://jetpress.org/volume14/bostrom.html> (accessed June 9, 2014).

Chiang, B. and M. J. Christie. 1994. Hyperpolarization by opioids acting on mu-receptors of a sub-population of rat periaqueductal gray neurones in vitro. *British Journal of Pharmacology* 113: 121–128.

- Crick, F. 1994. *The astonishing hypothesis: The scientific search for the soul*. New York: Charles Scribner's Sons.
- Delgado, J. M. R. 1965. *Evolution of physical control of the brain*. New York: American Museum of Natural History.
- Duan, Sh. Zh., N. Shimizu, A. Fukuda, T. Hori, and Y. Oomura. 1990. Hyperpolarizing action of enkephalin on neurons in the dorsal motor nucleus of the vagus, in vitro. *Brain Research Bulletin* 25: 551–559.
- Edgeworth, F. Y. 1881. *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. London: Kegan Paul.
- Edwards, J. 2005. Is consciousness only a property of individual cells? *Journal of Consciousness Studies* 12: 429–457. Available online <http://www.ucl.ac.uk/jonathan-edwards/publications/conscpropcells> (accessed June 9, 2014).
- Fechner, G. 1873. *Einige ideen zur schöpfungs- und entwicklungsgeschichte der organismen*. Leipzig: Breitkopf & Härtel.
- Freud, S. 1899. *Die traumdeutung*. Leipzig: Franz Deuticke.
- Freud, S. 1920. *Beyond the pleasure principle*. London: Hogarth.
- Friston, K. 2010. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience* 11: 127–138.
- Friston, K., J. Daunizeau, J. Kilner, and S. J. Kiebel. 2010. Action and behavior: A free-energy formulation. *Biological Cybernetics* 102: 227–260.
- Gossen, H. H. 1854. *Entwicklung der gesetze des menschlichen verkehrs, und der daraus fließenden regeln für menschliche*. Braunschweig: Vieweg.
- Huxley, A. 1932. *Brave new world*. London: Chatto & Windus.
- Jevons, W. S. 1871. *Theory of political economy*. London: Macmillan.
- Kluever, H. and P. C. Bucy. 1938. An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to “psychic blindness.” *Journal of Psychology* 5: 33–54.
- Koch, C. and I. Segev. 2000. The role of single neurons in information processing. *Nature Neuroscience* 3: 1171–1177. Available online [http://www.anc.upmc.fr/teaching/UE\\_NB045/docs/KochSegev2000.pdf](http://www.anc.upmc.fr/teaching/UE_NB045/docs/KochSegev2000.pdf) (accessed June 9, 2014).
- Kosarev, V. 1997. Who will live on the Earth in XXI century? *Neva* (in Russian) 1997(10): 135–149. Available online <http://vlad-kos.narod.ru/NEVA1.html> (accessed June 9, 2014).
- Kringelbach, M. L. and K. C. Berridge. 2006. Towards a functional anatomy of pleasure and happiness. *Trends in Cognitive Sciences* 13: 479–487.
- Lem, S. 1961. *Solaris*. Warsaw: Wydawnictwo MON.

- Menger, C. 1871. *Grundsätze der volkswirtschaftslehre*. Wien: W. Braumüller.
- Metzinger, T. (ed.). 2000. *The neural correlates of consciousness*. Cambridge, MA: MIT Press.
- Mill, J. S. 1863. *Utilitarianism*. London: Parker, Son, and Bourn.
- Murik, S. E. 1996. The relation of emotions to polarization processes in sensory systems. *International Journal of Neuroscience* 88: 185–197.
- Murik, S. E. 2006. *General neural mechanisms of motivations and emotions* (in Russian). Irkutsk: Irkutsk State University Press.
- Olds, J. and P. Milner. 1954. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology* 47: 419-427.
- Pearce, D. 2005. *The hedonistic imperative*. Available online <http://www.hedweb.com/hedethic/hedonist.htm> (accessed June 9, 2014).
- Pearce, D. 2014. *Utilitronium shockwaves? Unsorted Facebook postings*. Available online <http://www.hedweb.com/social-media/pre2014.html> (accessed July 19, 2014).
- Pyatygin, S. S., V. A. Vodeneev, and V. A. Opritov. 2006. Depolarization of plasma membrane as the universal primary bioelectric response of plant cells to action of different factors. *Advances in Current Biology* (in Russian) 126: 493–502.
- Russell, B. 1946. *History of western philosophy*. London: Allen and Unwin.
- Schultz, W., P. Dayan, and P. R. Montague. 1997. A neural substrate of prediction and reward. *Science* 275(5306): 1593–1599.
- Sem-Jacobsen, C. W. and O. B. Styri. 1972. Depth-electrographic stereotaxic psychosurgery. In *Psychosurgery*, ed. E. Hitchcock, L. Laitinen, and K. Vaernet, 76–82. Springfield: C.C. Thomas.
- Simonov, P. V. 1966. *What is emotion* (in Russian). Moscow: Science (in Russian: Nauka).
- Simonov, P. V. 1986. *The emotional brain: Physiology, neuroanatomy, psychology and emotion*. New York: Springer.
- Sevush, S. 2006. Single-neuron theory of consciousness. *Journal of Theoretical Biology* 238: 704–725. Available online <http://stevensevush.com> (accessed June 9, 2014).
- Shvyrkov, V. B. 1995. *Introduction to objective psychology: Neuronal basis of mind* (in Russian). Moscow: Institute of Psychology of the Russian Academy of Sciences Press.
- Smith, K. S. and K. C. Berridge. 2007. Opioid limbic circuit for reward: interaction between hedonic hotspots of nucleus accumbens and ventral pallidum. *Journal of Neuroscience* 27: 1594–1605.
- Walras, M.-É.-L. 1874. *Éléments d'économie politique pure, ou théorie de la richesse sociale*. Lausanne: Corbaz.
- Wu, M. F., Zh. P. Pang, M. Zhuo, and Z. C. Xu, 2005. Prolonged membrane potential depolarization in cingulate pyramidal cell after digit amputation in adult rats. *Molecular Pain* 1: 23-27.