An Interdisciplinary Analysis of Decision Variability and the Illusion of Free Will

Christian Ekstrand*

May 2, 2024

Abstract

We demonstrate how Darwinian evolution enhances decision-making via experiential learning and game-theoretical strategies. To an external observer, the resultant moderate intra-individual variability is indistinguishable from free will. We conclude that this evolutionary outcome is the simplest explanation for decision-making, thus being preferable according to Occam's razor, and implying that free will is an illusion. Furthermore, we argue that the perception of free will exists due to evolutionary benefits.

1 Free will and variability

In this paper, we define an *agent* as an entity—be it a biological organism or a machine—that can generate an output in response to an input, a process we term *decision-making*. The notion of *free will* in an agent implies an autonomy or control over its decisions, distinct from simple reflexive responses or deterministic outcomes. Free will is a multifaceted concept often characterized in philosophical discourse as the capacity to make choices unconstrained by external determinants. Philosophically, it hinges on the ability to act differently under the same circumstances (condition of alternative possibilities).

Each decision can be represented by a map $x \mapsto y$ from the space X of inputs to the space Y of outputs. For an agent endowed with free will, repeated decisions with identical inputs could yield different outputs y, implying a probability distribution in Y for each x. Determining if an agent has free will involves examining whether this probability distribution deviates from a singular value. This dispersion is called intra-individual variability, or more briefly, variability, and can be quantifiable in terms of a variety of statistical measures such as the Shannon entropy if Y is a finite space or the variance if it is a Euclidean space.

Variability is a necessary but not sufficient condition for free will: If an agent has free will, there exist at least one x for which there is variability.

*christian.ekstrand.phd@gmail.com

However, quantum mechanical experiments, such as the double-slit or Stern-Gerlach experiments, also display variability even though we do not associate them with free will. Due to its vague definition, the concept of free will poses challenges for direct scientific investigation. Variability, on the other hand, can be strictly defined and is therefore amenable to scientific analysis. In this paper, we explore variability as a quantifiable property and demonstrate how insights derived from such an investigation allow us to draw conclusions regarding the existence of free will.

2 Empirical variability

As humans, it feels obvious that we have free will and can display variability. For instance, a person might choose different routes to work on different days. However, attributing this variability to free will is problematic. Circumstantial changes—whether external, like varying weather conditions, or internal, such as a poor night's sleep—mean that the inputs each day are not identical, and it remains uncertain whether we can truly consider the agent (in this case, the person) as consistent across different instances.

To focus more directly on decision-making rather than circumstantial influences, consider the example of two chess players engaged in multiple games in an isolated room. We assume they play blitz games, which typically means less than ten minutes per game. This is long enough time for the player to give some thoughts to their moves rather than relying on instinct, yet brief enough that factors like the players' blood sugar levels remain relatively stable between games.

In such controlled conditions, each chess game will unfold differently, potentially diverging already at the first move. While it is possible to claim that these variations result from minor environmental changes, as someone who has both played and observed numerous chess matches, I, the author, along with my chess-playing acquaintances, claim that the variability in our games would persist even under completely identical conditions. We identify three reasons for the variability: learning, game theory, and enjoyment.

Firstly, moderate to advanced players know that variation leads to an improvement of their skills. Secondly, introducing unpredictability into gameplay can confound an opponent, or at least avoiding prediction and preparation against our moves, increasing the likelihood of winning. Lastly, variation in play is inherently enjoyable; the monotony of repeating identical games holds little appeal. Another observation from chess is that games between two players will display moderate variability. This makes sense as too low variability will lead to reduced learning, more predictability and less enjoyment. Too high variability, on the other hand, will include poorer moves and more losses, diminishing enjoyment.

Chess serves as an illustrative example, but these conclusions extend to general human behavior: we do display variability and the general reasons are learning, game theoretical and enjoyment. It explains why we engage with different friends, watch various genres of films, travel to new places, and choose diverse outfits. For completeness, we would like to point out that there are special situations where the variability does not originate in these three reasons. For instance, dietary choices often vary due to nutritional needs, and shifting sleep positions prevents bedsores.

While we have identified learning and game theory as contributors to moderate variability, mathematical arguments for this assertion are necessary and will next be addressed. Enjoyment of variations has a different relation to variability and will be discussed separately.

3 Experiential learning

When life forms make decisions, the outputs are often associated with a benefit, such as gaining access to food. The solid graph illustrates a simplified scenario where both the benefit and the output are one-dimensional, and the input remains constant. The primary objective for the life form is to identify the action (output) that maximizes the benefit.

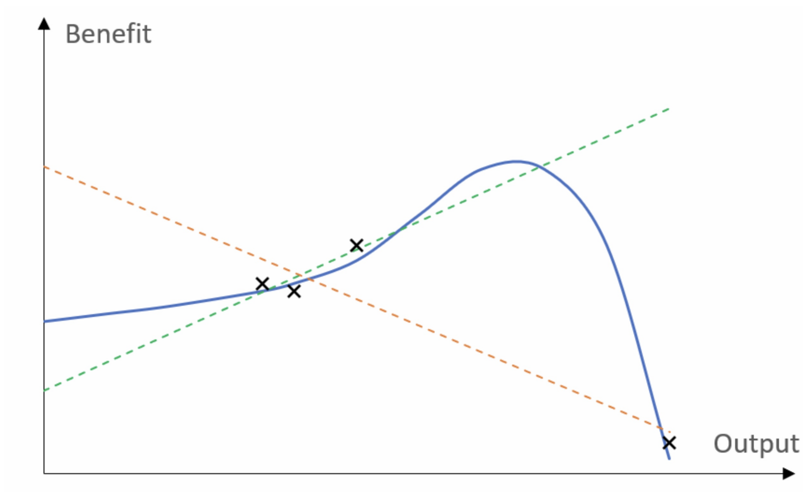

Figure 1: Illustration how experiential learning can find the action that results in optimal benefit.

In this context, the life form can employ a method of trial and error, known as experiential learning, to solve the problem. Imagine that the life form conducts three trials, represented by the crosses towards the center of the figure. The reason for the deviation from the solid line is that there can be additional factors, apart from the action, impacting the benefit. The solid line represents the benefit after averaging over such noise. Based on these trials, the life form could infer that an increase in action will enhance the benefit, as suggested by the dotted line with a positive slope.

The cross to the far right illustrates the situation when the life form applies to much variability in the learning process. We see from the dotted line with negative slope that it could lead to the incorrect prediction that the optimal region is to the left of the initial three data points. Furthermore, the additional data point led to a particularly bad benefit, which potentially could be harmful. Too low variability, on the other hand, can be illustrated by only using the two furthermost left data points, which again incorrectly would indicate a maximum towards the left.

This analysis shows that the amount of variability in experiential learning must be carefully balanced. It should be enough to overcome the noise without being so extensive that it leads to harmful outcomes or incorrect conclusions. We have thereby proven that experiential learning not only induces variability but also necessitates that it is of moderate magnitude to ensure effective decision-making.

4 Game theory

Game theory provides a structured framework for analyzing strategic interactions among decision-makers. Let us explore this with a straightforward example: a game where two players simultaneously display either one or two fingers. The payoff for each player depends on whether the combined total of fingers is odd or even. If the total is odd, player 1 wins an amount equal to the total number of fingers in dollars; if even, player 2 wins that amount. The payoff matrix for player 1 is shown in the table.

Player 1 \ Player 2	1	2
1	-\$2	\$3
2	\$3	-\$4

Table 1: The pay-off matrix for a simple game of holding up one or two fingers.

The matrix suggests that player 1 could reason that holding up 1 finger might be preferable, as it could result in either a win of \$3 or a loss of \$2—arguably better than a win of \$3 or a loss of \$4 when holding up 2 fingers. However, this approach is simplistic as player 2, recognizing the pattern, might adjust the strategy to counteract, often choosing 1 finger to minimize losses or maximize gains. A more nuanced strategy involves mixing the choices. If player 1 uses a mixed strategy, holding up 1 finger with a probability of 7/12 and 2 fingers with a probability of 5/12, calculations show that on average, player 1 gains \$1/12 per round, independently of player 2's choices.

To verify the optimality of this strategy, consider if player 2 adopts the same approach. In this scenario, player 2 would average a loss of \$1/12 per game, regardless of player 1's decisions. As player 2 can limit the loss to \$1/12, player 1's strategy must be optimal. Therefore, this mixed strategy proves more advantageous than adhering to a single, predictable approach of always choosing one or two fingers. This simple game exemplifies how game theory can predict and rationalize moderate-sized variability in strategic decision-making.

5 Darwinian evolution of decision-making

Variability, like any mental or physical trait observed in life forms, has to be a product of Darwinian evolution. It means that there must be an evolutionary advantage of incorporating a seemingly random component when taking decisions. We will explore this phenomenon within the broader context of the evolution of decision-making—a crucial trait vital for survival, reproduction, and gene propagation. Thus, significant evolutionary resources must have been invested into its enhancements. To elucidate the Darwinian role of variability, we will outline the key stages through which decision-making could have evolved.

The initial stage of evolutionary decision-making is the creation of input and output, which for life forms are the *sensory capabilities* and (typically) *locomotion*. Possibly the simplest method to improve decisions is to develop information storage, i.e. *memory*, of previous decisions. It is important to develop advanced decision techniques to adapt to a complex or changing environment, which can be done by *learning*. Some types of decisions are always favorable, while others are always unfavorable, which nature has handled by the evolution of *emotions*. Their purpose is to overrule the decision techniques. It is also important to develop decision techniques in a competitive world of other decision-making life forms, so-called *evolutionary game theory*. Decision techniques might benefit from a central location of the processing, motivating the development of a brain. Advanced brains can create *mental maps* of reality in order to evaluate the consequences of complex decisions. The mental maps can also prepare for future decisions for which the response time is crucial. For example, individuals in regions with high bear populations might mentally rehearse potential encounters to prepare for quick responses. Finally, *communication* can improve decisions. Ants, for example, use pheromones to communicate an optimal path.

As variability must have an evolutionary explanation, its cause should be found in the *italicized* characteristics underlying decision-making. The obvious candidates that can result in variability are indeed learning and game theory, supporting our previous assertion that they are the underlying causes. It should be noted, however, that primitive organisms can exhibit variability in the early stages of evolution, relying solely on basic sensory capabilities, locomotion, and a rudimentary form of memory. This will be illustrated in our subsequent discussion on E. coli.

Our conclusion that variability has an evolutionary origin is reinforced by its widespread presence in nature. Learning, for example, is always present in human life, particularly evident before we are grown-ups. This can be observed in the seemingly illogical behaviors of children and teenagers. Their random actions serve as a calibration, helping them to make better decisions later in life. Similar patterns are observable in the animal kingdom, where random behaviors during early life stages function as necessary learning experiences. The relevance of game theory and mixed strategies in natural settings is also well known [1]. For instance, an animal that follows the same path consistently becomes predictable and more susceptible to predators. Moreover, variability aids in environmental adaptation: consider animals with low variability in decision-making, all choosing to migrate in the same direction; a severe winter could eradicate the population. Variability has also been observed in simpler life forms such as fruit flies, see for example [2, 3], or [4] where the cause was attributed to experiential learning and game theory.

Finally, we will explore the enjoyment of variations and its evolutionary significance. To understand the impact of emotions in evolution, consider the fear of darkness. This seemingly irrational fear deterred our ancestors from venturing into the night—a behavior that proved advantageous given the risks and limited benefits associated with darkness due to our poor night vision. Consequently, evolution favored the survival and propagation of genes that instilled this fear. A similar evolutionary mechanism applies to other emotions. In particular, ancestors that enjoyed variations benefitted because of experiential learning and game theory, and passed their genes on. In conclusion, enjoyment of variations, while an influential driver of behavior, is not a fundamental cause of variability but rather a trait that has evolved to support and reinforce the benefits derived from experiential learning and game theory.

6 Cause of variability

We have made progress toward understanding the physical causes of variability in humans and animals by suggesting an origin from experiential learning and game theory, both shaped by Darwinian evolution. To further unravel the physical roots of this variability, it is necessary to investigate the apparent randomness that underpins experiential learning and game theory. As they are the result of evolution, studying primitive life forms that exhibit similar behaviors could provide insights. E. coli, known for its simplicity, serves as an ideal model organism for such studies, enabling us to trace the biological causes of variability.

E. coli moves through liquids in straight lines, called runs, followed by occasional stops, called tumbles, when it changes direction. Both movements manifest variability: the duration of runs follows a Poisson distribution, while the direction changes are seemingly random, albeit with a bias influenced by the direction of the previous run. T. A benefit of this particular type of motility can be understood when E. coli is exposed to a gradient of an attractant: it prolongs the runs towards favorable directions.

The motility of E. coli is facilitated by 1 to 10 flagella, which are whip-like structures. Each one is driven by a wheel-like structure called the flagellum motor. During a run, all the flagella rotate counterclockwise (viewed by an observer in pursuit of the cell) at about 130Hz. A tumble occurs when at least one flagellum's motor reverses its rotation. The decision to tumble is influenced by the bacterium's detection of attractants and repellents through protein receptors, functioning as a primitive form of short-term memory that assesses whether conditions are improving or worsening. The process occurs locally on E. coli and the information is sent to a near-by flagellum [5].

The causes of E. coli's motility are in principle understood down to molecular level due to, for example, cryogenic electron microscopy [6]. The processes causing the motility are complex and sensitive. They are affected by the variability within the bacterium and in its environment, such as thermal molecular movement. They also depend on fluctuations in the concentration of various attractants and repellents. Regarding the new direction of the bacterium after the tumble, it depends on the precise pre-tumble positions of the flagella as well as molecular variability.

The observable variability of E. coli's motility can be viewed as an amplification of the surrounding molecular variability. Such a complex process with a strong dependence on the in-data is sometimes described mathematically by chaos theory. In biological systems, it typically occurs due to positive feedback loops. The decision process of E. coli is illustrated in the figure. The inputs to its decisions are the gradient of attractants and repellents, and the internal and external variability. The outcome is its motility.

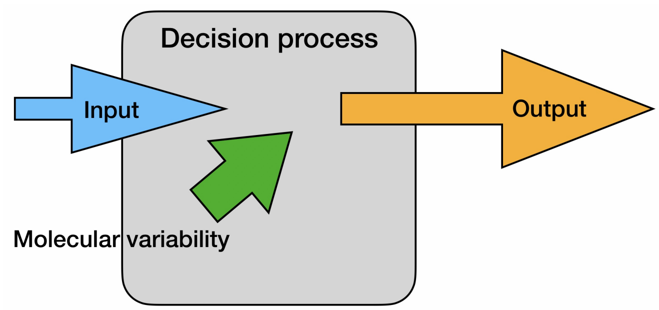

Figure 2: The decision process for E. coli's motility.

Evolutionary game theory is observed for E. coli when it takes the roles of cooperator or cheater, depending on the circumstances [7]. While experiential learning per se has not been observed in E. coli, it has been noted in other primitive organisms, such as the singular cell (but with multiple nuclei) slime mold Physarum polycephalum, which exhibits maze-solving abilities [8], and C. elegans, a tiny worm with only 302 neurons [9]. The simplicity of these life forms suggests that the variability in their learning and evolutionary game theory also derives from molecular variability, analogous to the cause of E. coli's motility. Furthermore, we note that experiential learning and evolutionary game theory have been observed in the full spectrum from primitive to advanced life forms, and to our knowledge, there have been no studies that indicate a nonsmooth transition in variability. The simplest explanation is, therefore, that the moderate variability observed in both experiential learning and evolutionary game theory in more complex life forms also results from the amplification of molecular variability.

7 A decision model

The decision-making mechanisms in more advanced, non-primitive life forms, which incorporate memories and emotions, are more complex than those observed in organisms like E. coli. In these complex organisms, variability is only one component influencing decisions, rather than the dominant factor. Consequently, the role of the inputs increases, whereas the influence of molecular variability decreases, as depicted in the figure. This molecular variability arises from phenomena such as thermal motion and molecular interactions, which are in part derived from quantum fluctuations, rendering this variability not just complex and unpredictable but also partially genuinely random. Additionally, non-primitive organisms are larger than E. coli and thus relatively less exposed to environmental variability. It is logical to assume that the molecular variability influencing decision-making primarily occurs internally, particularly within the brain. This assertion is supported by findings that demonstrate randomness in the brain, such as the stochastic opening and closing of ion channels, which are proposed to be caused by molecular variability [10].

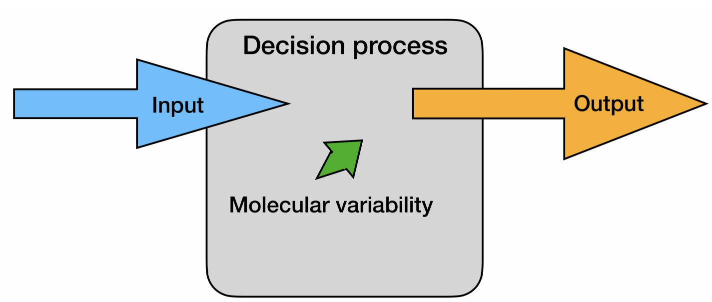

Figure 3: The decision process for advanced life forms.

Some organisms, including humans, are capable of consciously and subconsciously modulating the extent to which molecular variability influences their behavior. For example, in novel situations, such as relocating to a new city, humans often exhibit increased behavioral variability. This adaptability is beneficial for practical reasons, like finding local amenities, and also plays a role in evolutionary game theory by aiding in the formation of new social connections or finding a mate. Conversely, individuals experiencing depression may display reduced variability, to avoid the harmful negative benefits that was mentioned when we discussed learning.

Given these insights, we can address the question whether advanced life forms operate as deterministic or non-deterministic agents. If we consider only the inputs and outputs, ignoring the subtle molecular variability, life forms appear to be non-deterministic. However, if we also consider molecular variability as a form of input, they might seem to be deterministic. Yet, this dichotomy is misleading because the agent itself is a physical object. It consists of molecules and has therefore variability itself. Thus, it is impractical to separate decision-making from molecular variability. The conclusion is that the decision process will always exhibit some level of inherent variability and will therefore act as a non-deterministic agent.

An agent based on this model is a non-deterministic agent and will for an external observer appear to have variability and free will. The simplest explanation is that free will is synonymous to variability and can be understood by the above model. It has yet to be argued why humans (and possibly other life forms) are under the perception of having free will. This will be the topic for the remainder of the paper.

8 Self-awareness

The mental maps used for complex decision-making incorporate both memories of past events and simulations of possible future scenarios. The main actor in these mappings is the life form itself. It means that the mapping is most efficiently done by the ability to do self-referencing. If the organism has language capabilities, self-reference is typically facilitated by a word like "I". Thoughts and memories might then be framed as, "What do I have to do tomorrow?", "I'm afraid", or "What an eventful day I had yesterday." This mode of thinking aligns closely with human cognition. Links between language and self-referencing are supported by neurological research [11]. Furthermore, works by psychologists suggest that language plays a significant role in self-awareness [12].

If communicating with a life form capable of mental mappings and asking whether it exists, the use of self-references suggests the answer would be affirmative. As we will soon argue, it will also experience that it has the freedom of choice. It also follows from our previous arguments that it can sense emotions. This does indeed sound like a self-aware life form. The simplest explanation for self-awareness is, therefore, that it is the result of such self-referencing within the mental map [13].

We have argued that a sense of self is a byproduct of self-references within our mental map, crucial for making complex decisions. If our hypothesis is accurate, animals with a sense of self should also typically be good decision-makers. To test which animals possess self-awareness, researchers have employed methods such as the mirror test [14], where an animal is marked in a spot only visible through a mirror. Animals that recognize and attempt to investigate or remove the mark demonstrate self-awareness. Species that have passed this test include great apes, elephants, magpies, and dolphins—all noted for their decision-making capabilities.

With the ongoing development of artificial intelligence, a crucial question is whether machines can be self-aware. According to our hypothesis, self-awareness arises when having thoughts and memories about oneself. The thoughts could be future scenarios of a mental representation of the world. The artificial intelligences of today, however, are not constructed to speculate regarding their future or by imagining themselves as part of a mental representation of the world; they are built for different purposes. They will therefore not need self-references or the development of a sense of self. We infer that current artificial intelligences will not exhibit self-awareness, at least not in the way biological life forms do.

9 The perception of free will

To understand the evolutionary advantage of perceiving oneself as having free will, consider a self-aware life form that lacks this illusion. Such a being will, due to variability, occasionally make a decision of low probability. Trying to understand these kinds of decisions would be problematic. The life form knows the decision came from itself but at the same time it will not understand how it happened. Such a life form might spend time contemplating low-probability decisions, which are not the decisions that the mental map should prioritize. It is instead preferable to analyze the decisions that lead to severe consequences.

For instance, imagine a scenario where a casual remark unexpectedly upsets a close friend or partner. This event, due to its potential for severe repercussions, merits careful reflection and analysis. Conversely, a low-probability decision such as choosing to wear a long-unworn sweater color, unless it leads to notable consequences, does not warrant similar scrutiny.

A simple evolutionary solution to this problem is to introduce an emotion into the life form so that it experiences that it was the self that was responsible for the decision. This would mean that a sense of self is a necessity for the illusion of free will. If the life form knows a language, it might introduce some words that describe this emotion. In English, such words are "want" and "feel". These verbs represent an ability for the self to take independent decisions, i.e. the existence of free will. The life form will motivate its decisions with phrases such as "Because I wanted to" or "Since I felt like it". This mechanism allows life forms to allocate their analytical resources more efficiently, focusing on decisions with significant consequences rather than every low-probability event. This evolutionary perspective provides a compelling explanation for the strong human inclination to believe in free will, despite contrasting evidence. It underscores the adaptive nature of our perceptions, suggesting that evolutionary advantages may influence our belief systems.

Moreover, humans have a psychological need to maintain internal consistency. Acknowledging that one's actions are not free but determined by external factors or random internal processes could create cognitive dissonance. The illusion of free will helps resolve this discomfort by allowing individuals to attribute their actions to their own volition, thus maintaining a sense of control and consistency in their self-concept. There are plenty of examples for our demand for consistency, such as the craving of a theory for the creation of the world. Despite the lack of scientific evidence, historical civilizations came up with imaginative stories rather than admitting that they did not know. Such examples demonstrate a broader human tendency to make up explanations rather than confront uncertainty.

These discussions demonstrate that belief in free will would likely arise regardless of its actual existence. Therefore, the primary argument for free will—the subjective sensation of having it—is not a reliable indicator of its reality.

Neuroscientific research [15] supports our model, indicating that the brain first makes a decision and then constructs the impression that the self was the decision-maker. Additionally, the model proposes that if we were to manipulate the brain into initiating actions, it would subsequently fabricate explanations to attribute these actions to the self. This has indeed been observed in experiments [16, 17].

The fact that nature seems to deceive us into certain behaviors or beliefs, is not new. An example relevant to our discussion is that evolution has equipped us with the emotion to enjoy moderate variations. Another example is the experience of a stationary Earth and a rotating heaven. The illusion of geocentricity could finally be abandoned by combining heliocentricity with Occam's razor. As the model presented here is the simplest fit to observations, Occam's razor should once again be applied, this time to abandon the belief in free will.

References

[1] Smith, J.M. Evolution and the Theory of Games. Cambridge University Press, (1982)

[2] Wolf, R., Heisenberg, M. The locomotor activity of Drosophila melanogaster is controlled by a dual mode of walking and flight. *J. Comp. Physiol A* **169**, 699-705 (1991).

[3] Heisenberg, M. Is free will an illusion? *Nature* **459**, 164-165 (2009).

[4] Brembs, B. Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates. *Proceedings of the Royal Society B: Biological Sciences* **278**, 1707, 930-939 (2011).

[5] Berg, H.C. E. coli in Motion. Springer, New York (2004)

[6] Wadhwa, N. Berg, H.C. Bacterial motility: machinery and mechanisms. *Nat Rev MicroBiol* **20**(3), 161-173 (2022)

[7] Vulic M., Kolter R. Evolutionary cheating in Escherichia coli stationary phase cultures. *Genetics* **158**, 519–526 (2001)

[8] Nakagaki T., Yamada H., Tóth A. Intelligence: Maze-solving by an amoeboid organism. *Nature* **407**, 6803, 470-470 (2000).

[9] Ardiel, E.L., Rankin, C.H. An elegant mind: Learning and memory in Caenorhabditis elegans. *Learning & Memory*, **17**(4), 191-201 (2010).

[10] Braun, H.A. Stochasticity Versus Determinacy in Neurobiology: From Ion Channels to the Question of the "Free Will". *Front. Syst. Neurosci.* **15**, 629436 (2021)

[11] Morin A., Michaud J. Self-awareness and the left inferior frontal gyrus: Inner speech use during self-related processing. *Brain Research Bulletin*, **74**, 387–396 (2007).

[12] Vygotsky L. Thought and Language. MIT Press, Cambridge, MA (1986).

[13] Alexander R.D. Evolution of the human psyche. In: The human revolution. Behavioral and biological perspectives on the origins of modern humans., ed. Mellars P., Stringer C., 455-531, Princeton University Press, (1989).

[14] Gallup, G.G. Jr. Chimpanzees: Self-recognition. *Science* **167**, 3914, 86-87 (1970).

[15] Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K. 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**, 3, 623-642 (1983).

[16] Fried, I., Wilson, C.L., MacDonald, K.A., Behnke, E.J. Electric current stimulates laughter. *Nature* **391**, 6668, 650-650 (1998).

[17] Gazzaniga, M.S., In Gazzaniga, M.S. (Ed.) The Cognitive Neurosciences. MIT Press, Cambridge, MA (1995)