

Wireheading Done Right: Stay Positive Without Going Insane

Andrés Gómez Emilsson

Qualia Research Institute

San Francisco, CA

ALGEKALIPSO@GMAIL.COM

Abstract

Wireheads are beings who have changed their reward architecture in order to be happy all the time. Unfortunately, few people are making a serious effort to steel man [49] the case for wireheading [61]. The concept of wireheading tends to be a conversation stopper, and is frequently used as a *reductio-ad-absurdum* for valence utilitarianism [8]. Hedonism is a low-status philosophy at the moment, but this may be the result of what amounts to dumb reasons [13] (*i.e.*, going against it signals intellectual sophistication). Let’s be meta-contrarian for a moment and think critically about it. What would a good case for wireheading look like? In what follows I will (1) provide an account of what is know about emotional dynamics over time, (2) discuss the known pitfalls of current wireheading methods, (3) propose a system to overcome those pitfalls, and (4) make the case that combining wireheading (done right) with a systematic exploration of the state-space of consciousness might ultimately be our saving grace against the perils of Darwinism at the evolutionary limit.

The Cube of Euphoria

A little over a year ago I conducted a study to figure out the main dimensions along which psychotropic drugs influence people. The State-Space of Drug Effects [31] consists of six main dimensions: fast euphoria, slow euphoria, spiritual euphoria, clarity, perception of overall value, and external vs. internal source of interest. The first three dimensions are directly related to pleasure, which makes them relevant for our current discussion.

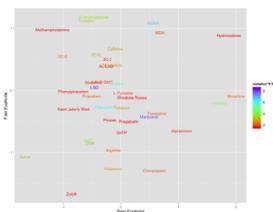


Figure 1: Slow — Fast

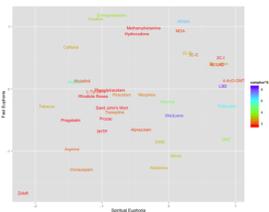


Figure 2: Spiritual — Fast

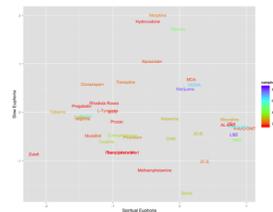


Figure 3: Spiritual Slow

Fast euphoria is what you get when you take stimulants, exercise, or anticipate that something great is about to happen. Slow euphoria is what you experience when you take opioids or depressants, receive messages, or hug a loved one. Spiritual/philosophical euphoria changes less frequently relative to the daily comings and goings of the other two. It is a state of consciousness related to how we represent “the big picture”. Those who seek it try to induce it by methods such as philosophical thinking, spiritual practices, and/or psychedelic drug use.

Two out of these three dimensions is equivalent to the well-studied emotion classification space of valence and arousal (also called core affect [63]). Valence is *how good the experience feels*, whereas arousal deals with the intensity of the experience. It turns out that one can get the slow-fast projection of the cube of euphoria by changing the basis used to represent [1] the valence-arousal space. You can get the valence-arousal space simply by rotating the slow-fast euphoria projection by 45 degrees:

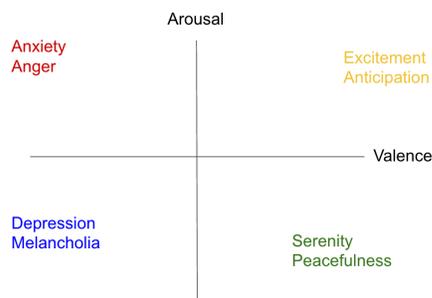


Figure 4: The valence-arousal dimensional classification of emotions

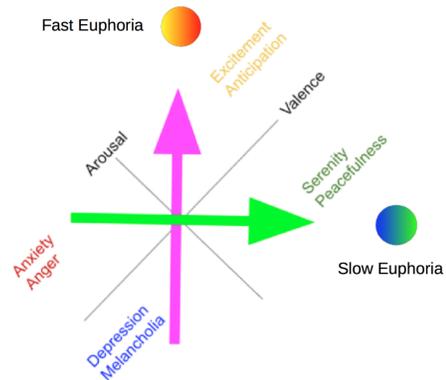


Figure 5: A change of basis by turning the space 90 degrees. Fast and Slow euphoria are two of three dimensions of the cube of euphoria.

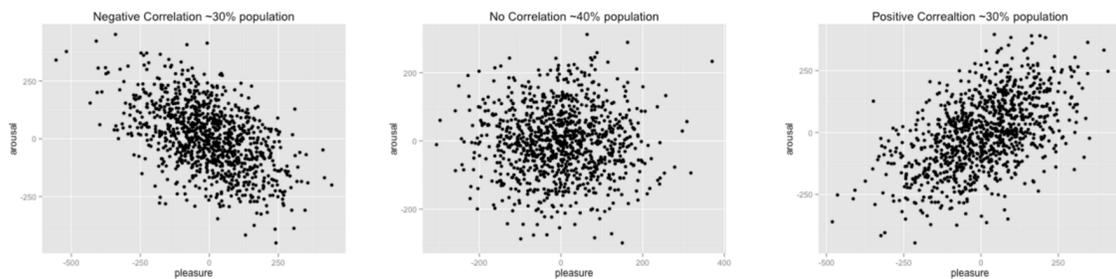
As we can see, fast euphoria is equivalent to “high-valence, high-arousal”, while slow euphoria is equivalent to “high-valence, low-arousal”. This basis is not uncommon in affective psychology where the axes are usually labelled “positive and negative activation”. We will use a yellow-red circle to represent fast euphoria and a blue-green circle to represent slow euphoria. I chose this color-coding by reasoning [2] that warm colors are a better representation of ecstatic states of consciousness whereas cool colors illustrate better the feelings of cooling off and relaxing. I happen to prefer the fast-slow basis because it highlights the different kinds of euphoria in a helpful way that captures behavioural differences. This will be important when we get to steel-manning wireheading later on.

Formalizing the Hedonic Treadmill: Negative Feedback Mechanisms

It is well known that in the long run major life events have a surprisingly small effect on your overall level of happiness. One tends to always orbit around one's hedonic set-point [3] (our mean valence and arousal value). Although our average sense of wellbeing does change from context to context (in response to variables such as stress, novelty, drug regimens, accomplishments, and opportunities for meaningful relationships), the environmental effect is usually washed out over time by one's internal negative feedback mechanisms. The ability to achieve lasting happiness, it turns out, was not as evolutionarily adaptive in the ancestral environment as the robust re-centring of affective dynamics that ended up governing out patterns of wellbeing. Thankfully, though unfairly, we are not all equally miserable; some lucky people are born hyperthymic and enjoy life the majority of the time [35]. Genetically-determined pain-thresholds [48] do not only influence how one responds to physical discomfort, but also predict the the size of one's social network [43] (presumably by making social rejection less taxing).

Less well known is that people have different values for their valence-arousal correlation. According to a 2007 study by Peter Kuppens [45], the conventional wisdom in affective psychology that valence and arousal are uncorrelated is not quite correct. For 30% of people, valence is negatively correlated with arousal, for 30% it is the opposite, and for the remaining 40% there is no correlation between these dimensions.

Individual Differences in the Relationship Between Pleasure and Arousal



Kuppens 2007

Figure 6: The study showed that the correlation between valence and arousal was person-specific (negative for 30%, positive for 30%, no correlation for 40% of people).

This means that some people usually experience high valence (*i.e.*, feel good) at the same time as being in an upbeat, high energy state, and when they feel bad they tend to also have low levels of energy. On the other extreme are those who experience bliss by turning down the energy and relaxing and who primarily experience bad feelings in the form of high-energy states (such as irritation, worry, and anger).

What else is variable between people? As it turns out, the transition patterns of core affect are related to personality factors. People’s level of variance in the valence dimension is an important component of neuroticism. Although most neurotics tend to hang out in low-valence states, there are indeed very happy neurotics whose problem is not that they feel bad, but that great feelings are too short lasting and unpredictable. It is the unpredictability of valence rather than its absolute value that results in the coping mechanisms typical of this dimension. Likewise, higher variability in arousal is a component of extraversion. Openness to experience can be understood in terms of novelty-triggered increases in valence, so that more open individuals are more likely to experience euphoria of all kinds when learning new information relative to people who describe themselves as conventional. Conscientious individuals feel very rewarded when they complete a laborious task (but may experience more intense shame if they do not finish it on time). Agreeableness is undeniably connected to a positive perception of other people. If one feels that others are right and deserve to exist one is more likely to cooperate. The way to have positive perceptions of others is to increase the hedonic tone of the interpersonal representations [56]. In brief, core affect dynamics can be used to capture otherwise hard-to-describe properties of the various personality factors. They each have a signature behaviour in the valence-arousal space.

In a paper titled A Hierarchical Latent Stochastic Differential Equation Model for Affective Dynamics [50], Oravecz, Tuerlinckx, and Vandekerckhove applied the Ornstein-Uhlenbeck process [6] to the dynamics of core affect. Their model takes into account many important features that had previously been overlooked for the sake of simplicity. As mentioned in the previous paragraph, these features turn out to be important signatures of personality factors, so having a model that incorporates them may be very useful for understanding the differences between people. Their model describes people as having: a latent *home base* (hedonic set-point), variance (for both components), a correlation between valence and arousal, an average speed, and a time-dependent relocation of the home base determined by the hour of the day. The model allows you to estimate person-specific parameters (using as input a sequence of self-reported emotional states). In turn, once you have determined someone’s latent parameters, the model can help you predict their future affect based on their current state.

This model is perhaps as good as it gets if you are restricted by a Markov assumption [4] and given only the valence and arousal dimensions of the participants over time. The state-space of emotion is far more granular, though. Even increasing the number of dimensions by one (*e.g.*, by including the dimension of spiritual euphoria) may go a long way in clarifying the nature of unexpected emotional transitions. What explains the sometimes very large effect of philosophical discoveries, religious conversions, and personal epiphanies?

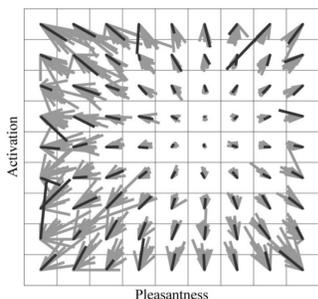


Figure 6. Estimated vector field of the core affect grid. It shows the angle and speed of leaving core affect grid positions. The black lines are calculated based on the observed data, and the gray lines are calculated based on simulated data.

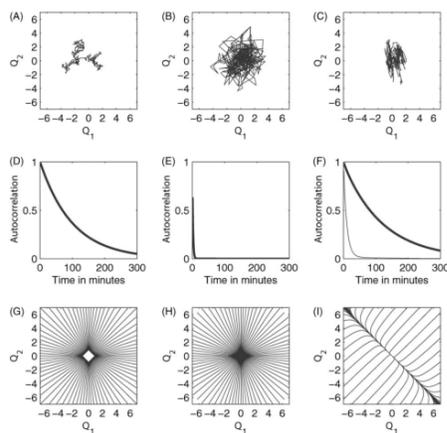


Figure 4. Shows the effect of different **B** matrices.

A Map of Emotion Attractors: Studying 176x176 Transition Probabilities

From 2012 to 2014 I worked on modelling the dynamics of emotion transitions. I did this as part of a research project for Kanjoya and then as the topic of my Master’s thesis. If you’re interested in reading more about it please see this paper I co-authored [67] on predicting future emotions based on a sequence of previous ones (together with social cues).¹

The analysis I worked on was based on a sample of hundreds of thousands of users of the now-defunct Experience Project social network. Participants would have the option to record their mood on the landing page: they would select an emotion from a list of 176 words, say how intense this emotion was at the time (from 1 to 5) and explain why they feel the way they do (open text; optional). I analyzed the transition probability between each ordered pair of emotions for different intervals of time and compressed it into a score that describes the overall *flow* of people between them. This results in a flow graph that we can analyze with tools from graph theory. I explored many ways of clustering this graph and ultimately settled on a method that generated the best predictive power on a model to forecast future emotions. This method consisted of grouping the emotions in such a way that each emotion would maximize its mean transition probability to other emotions in the same group (relative to other groups). For the paper I made this graph with all of the emotions (nodes), the transition probabilities between them (edge thickness) and the resulting clusters (colors):

1. Thanks to Chris Potts [62] for putting the papers online.

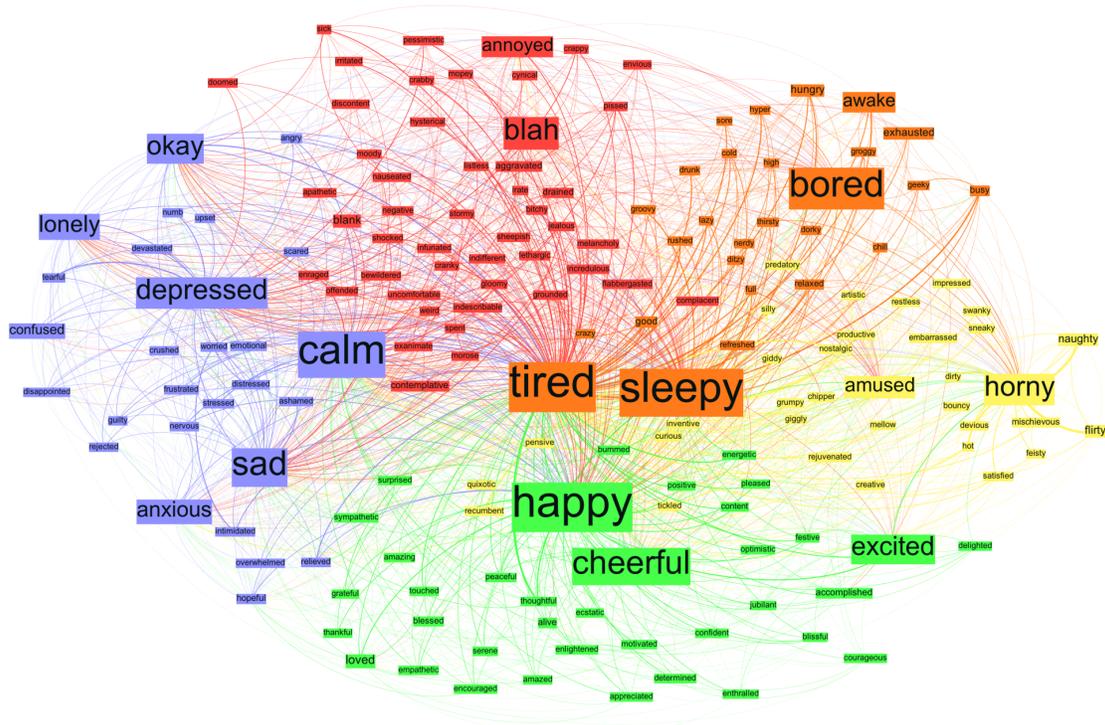


Figure 8: This is a weighted directed graph with 176 nodes, each representing a distinct state of consciousness. The edges represent the (directed) compressed transition probability between each ordered pair of states. The size of each node approximates the base rate at which the emotion occurs in the sample.

Each color represents a given “emotion attractor”. At a high level, we can say that whenever you are experiencing an emotion that is, *e.g.*, green you are more likely to transition to other emotions that are also green (relative to what would be expected from choosing an emotion randomly). This analysis is ultimately consistent with Oravecz *et al.*’s model in the sense that both analysis study the dynamic way in which people tend to get in and out of their home base. However, the granularity afforded by the 176 different options also allowed me to examine the deviations from this pattern. I investigated the question: “Which emotions take you to places that are inconsistent with the general trend of stochastically moving towards the central hedonic set-point?”

It turns out that some emotions behave in interesting ways. Some are what we called “hubs”: common stopping points that work as a route between any two colors. For example, “calm” and “tired” are hubs, and they do not give you much information about past or future emotions. Some other emotions behave like ‘gateways’ in the sense that they tend to indicate a jump from a particular color to another. For example, “hopeful” and “relieved”

are two gateway emotions: they work as stepping stones from blue (depressive) emotions to green (positive) ones.

Some emotions challenge the hedonic treadmill by virtue of predicting unexpectedly long-lasting stays on a given color. For example, the words “blessed”, “blissful” and “loved” were great predictors of long-lasting green emotions. By examining the text of these mood updates we determined that on average the people listing religious and spiritual themes as the cause of their feelings were more likely to stay for longer periods of time in the zone of positive emotions than most other people in the sample. I suppose that people’s spiritual euphoria may hack the pattern of hedonic habituation to some extent in a few lucky ones. I personally do not think that this is a scalable solution for everyone, though, since not everyone is spiritually oriented or has their endogenous opioid system wired up properly for meditation. The outstanding effect sizes we may see in some people who benefit from a particular, *e.g.*, meditation technique rarely generalize to everyone else. That said, it is certainly neat to see some evidence of some (spiritual/philosophical) sabotage at the mill [53].

How can we feel better in the long term?

A few years ago I abandoned hope in the idea that psychological interventions are sufficient to increase our wellbeing (philosophy, spirituality and exposure therapy can only take you so far in making you feel better). So what is next? The trick will be to combine psychological, chemical, electrical and genetic methods together in a balanced and healthy way and forget about relying on a single method. Can we be happy all the time? Let us move on to the subject of wireheading [61] more directly. Given what we have discussed about core affect, emotion dynamics and the resilience of the hedonic set-point, is it possible to wirehead oneself in a non-regrettable way? I think that the answer is yes, but we will need to avoid some crucial dangers...

Wireheading Done Wrong I: Forgetting About The Negative Feedback

Fast and slow euphoria can be reliably triggered by sensorial or chemical methods. However, doing so kick-starts two negative feedback mechanisms.

The first one is that the effect is reduced (shown in the image as the little loops with a minus sign) with each use. And the second one is that withdrawing from these euphoric states kindles circuits [22] that do the opposite of what was intended (as shown by the arrows with a positive sign). Too much of something that calms you is going to bring about a long and withdrawn state of constant low-level anxiety. Too much of anything that makes you up-beat and ecstatic is going to induce a long and withdrawn state of low-level depression.

Amphetamines, traditional opioids, barbiturates and empathogens can be ruled out as wise tools for positive hedonic recalibration. They are not comprehensive life enrichers precisely because it is not possible (at least as of 2020) to control the negative feedback mechanisms they kick-start. Simply pushing the button of pleasure and hoping it will all be alright is not

an intelligent strategy given our physiological implementation. The onset of this negative feedback often triggers addictive behavior and physiological changes that shape the brain [5] to expect the substance.

The case of spiritual/philosophical euphoria is a lot trickier. It is clear that there is a negative feedback that may be described (more or less) as a sort of philosophical boredom. Psychedelics are capable of changing our brain so as to increase the range of possible valence (*i.e.*, they can enable states of extreme pleasure but also extreme suffering) in a way that sidesteps the need to directly interact with our pleasure centers. I think it is extremely important to figure out the mechanism of action of psychedelic bliss. We will in fact address it in another article. For now it will suffice to say that psychedelic pleasure does not seem to induce cravings or withdrawal. We should take a close look at it because it may be the key to understanding how to produce unlimited positive valence with no negative repercussions. Unfortunately, producing philosophical, spiritual and psychedelic bliss nowadays is still more of an art than a science; these methods are unreliable and can backfire tremendously.

In summary, we might say that if one is oblivious to negative feedback, then meth addiction is an attempt at *fast euphoria wireheading*, whereas opioid dependence is the result of trying (ineffectively) to obtain boundless *slow euphoria*. *Spiritual euphoria wireheading* attempts usually involve activities such as philosophy, meditation, prayer and psychedelic drug use. Even though attempting spiritual euphoria wireheading on oneself is a hell of a lot healthier than doing meth or heroin, it is certainly not free from possible psychological side effects (such as acquiring bizarre beliefs [18] or experiencing—sometimes profoundly distressing—events of spiritual dysphoria [72] and unwanted changes in one’s belief system).

Wireheading Done Wrong II: Seduced By A World Of Your Own

One simple approach to wireheading effectively is to remove either one or both of the aforementioned negative feedback mechanisms. Wiring electrodes into one’s pleasure centres apparently removes both [58]. It turns out that the mechanism for generating physiological tolerance is bypassed by direct electric (rather than chemical) stimulation to the nucleus accumbens. Bliss obtained this way does not seem to stop pouring nor diminish in intensity over time. Unfortunately this method has profound pitfalls. Most salient of all is that if given the choice, mice (and some, but not all, people) will continuously self-stimulate this way as frequently and as intensely as possible, neglecting both physiological needs (like food and sleep) and social demands (like feeding one’s children or paying taxes).

In the case of humans, people feel compelled to self-stimulate when suffering, but under normal circumstances (*i.e.*, if they already feel good) they can hold off from pressing the button in order to carry out other activities. Admittedly, this is an improvement over drugs, which make you feel terrible in the long run and in turn make you seek relief with the same method that brought you there. With electrical rather than chemical stimulation we can at least avoid this pitfall. That said, people do not like to have objects implanted in their brain, and our infection-prone future will thank us for not developing an addictive

technology that requires a constant stream of ineffective antibiotics to keep it plugged in place. Thankfully, future wireheading may be minimally invasive. Attractive alternatives to old-fashioned electrodes include body-powered wireless implants [42], optogenetics, and genetically-encoded magnetic triggers of neural activity [66].

A much more subtle way to try to improve one’s hedonic set point is by counteracting the activation of the post-pleasure dysphoria. SSRIs and the less well-known fast-acting aminoguanidine agmatine [20] help prevent gross kindling of circuits that produce unpleasant sensations. This method may ultimately come down to increasing the amount of noise in the entire system² and thus reducing the survivability of highly-ordered states [32] (such as pain and pleasure) in one’s consciousness. Unfortunately, preventing withdrawal by this method comes at the cost of blunting high-valence states [41]. Prolonged SSRI use often makes people anhedonic and feel like they have lost all lust for life. In contrast, ibogaine [19] and low-dose opioid antagonists are promising chemical avenues to attack the same problem in a very different way without such side-effects. These compounds work by rebalancing one’s proportion of the various opioid receptor subtypes and in turn ramping up one’s hedonic capacity.

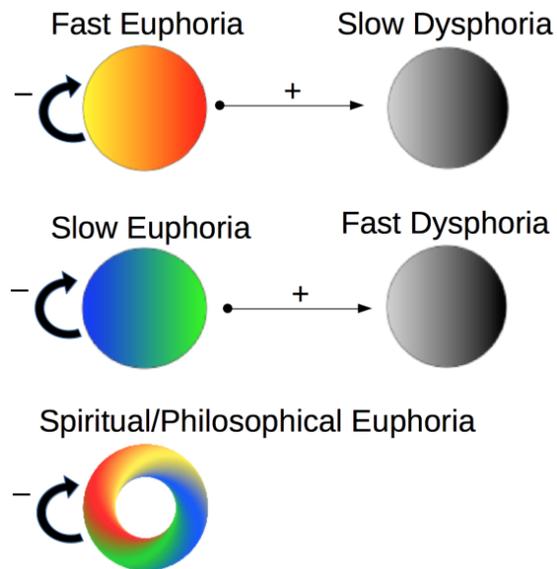


Figure 9: Current negative hedonic feedback dynamics

A whole generation of people will probably be “lost” to what I call *single euphoria wireheading*. Let’s say that you have mastered the ability to experience a high level of fast euphoria in a sustainable way. You can (in principle) stop at any point and come down without feeling like you are missing out. But whenever you do activate the fast euphoria you are about ten times more motivated to go out, explore the world, work on projects and meet great people who also share your newfound interests and values. You may end up choosing to join a community of other people who value living fast and staying hyper-motivated, just as you do now.

Fast euphoria in particular is extremely tricky to program correctly, since it deals so directly with behavioral reinforcement. Many people get hooked on *meth + X* rather than on *just meth*: whether X is music, gaming, sex, gambling, porn and/or alcohol, during a meth binge

2. I owe this theoretical framework to Mike Johnson and his magnificent work on the nature of valence. [44]

people often end up doing the exact same repetitive but pleasant task for tens of hours. In other words, fast euphoria not only reinforces itself, but it also reinforces whatever activity you do while you experience it, and this is especially the case if the activity is more enjoyable as a result of the fast euphoria. Stimulant addiction [9], deep brain stimulation [69], and manic states in bipolar sufferers [64] share a core personality-changing effect driven by an excessive interest in a few rewarding activities at the expense of all other interests and responsibilities. It is extremely tricky to rationally use one's reinforcement system directly in order to recalibrate one's hedonic tone. Without (as-of-yet-uninvented) safeguards, doing so tends to increase impulsivity in the long term and mess up one's preference architecture.

We could in principle block the metabolic pathways that lead to changes in one's motivational system as a response to fast euphoria. If we did this, then people might be able to master side-effect-free hyper-motivation [55]. Does this mean that a straightforward road to super-happiness [60] is short-cutting to perpetual motivation?

The main problem is that motivation and one's implicit notion of our self-in-time interact in unpredictable ways. One of the ways something like meth transforms your preference architecture is by forcibly re-defining your self-model [47] (*c.f.*, Ontological Qualia [29]). Fast euphoria makes one attend to the present moment. In high amounts, absorption can be so intense it feels like the present moment *is who you fundamentally are*. Returning to typical human life conditions after an experience of feeling *truly alive* can be felt as a sort of death—a death not of the biological body but of the fleeting self-models that inhabited the subjectively more worthwhile state-spaces of consciousness.

If one's implicit self-model is not robust against sudden changes of fast euphoria, then one will not be good at surviving and being productive in a social economy. Let us say that person A identifies with her future self in 2059 and thus saves for retirement, but person B on meth has a very hard time thinking of herself in any other terms than “me, right now, for as long as I can stay in this state of mind.” The present moment, the immediate future, and the pleasure opportunities available in them can be so salient that they eclipse every other interest. If we do not find a way to prevent this shift in perspective it may be impossible to safeguard rationality when showered with streams of high-grade fast euphoria.

How about slow euphoria wireheading? [57] I suspect that it is in principle possible to master hyper-relaxation without being incapacitated. In the meantime, trying to slow down too much does seem to reduce one's productivity by a good margin, so wireheading of the slow euphoria type is not currently advisable. That said, achieving hedonic recalibration by guaranteeing a minimum of slow euphoria is, as I see it, a lot more feasible than doing so through fast euphoria. Slow euphoria does not have the explosive effects on one's motivational architecture and self-models that fast euphoria does. On the contrary, relaxation can allow us to reconceive of ourselves as beings who inhabit much longer timelines (to really grasp our decades-long lifespan and know how to pace ourselves rather than feeling pressed to identify with our present moment exclusively).

Spiritual euphoria may or may not necessarily imply changes in one’s belief structure. Currently, peak spiritual/philosophical states (including high levels of psychedelia) are a rather different kind of subjective wellbeing than the other two that dominate our everyday life. This bliss is often associated with extreme changes in the quality of one’s conceptualization of reality [37], which limits its effective incorporation into a rational and economically productive being. (Unless, of course, one is producing useful information in those states [28]. More about this further below.)

In summary, if a device is ever discovered that allows people to enjoy fast, slow, or spiritual euphoria without implicitly influencing their worldview and economic capacity, then that device will probably become a staple of life. Issues of authorship and agency aside [10], single euphoria wireheading without serious engineering to counter its problems is an evolutionary dead-end. Whether controlled single euphoria wireheading can be adaptive is still up for debate.

Wireheading Done Wrong III: Becoming A Pure Replicator (Even If You Love It)

Look, we’re all friends here. We are trying to delay for as long as possible the development of a Singleton [23] (*i.e.*, a state of complete control by one system), while we also try to keep at bay the problem of Moloch [15] (*i.e.*, complete lack of control).³ We are trying to find a sustainable solution against both extremes. In our ideal world, all beings should have the freedom to explore the state-space of consciousness [52] however they want (or live in an archipelago of societies [14] at the very least). We need to work together to design a future that avoids evolutionary extremes and safeguards the freedom of consciousness. Now, who is the enemy?

The Threat of Pure Replicators

I will define a *pure replicator*, in the context of agents and minds, to be an intelligence that is indifferent towards the valence of its conscious states and those of others. A pure replicator invests all of its energy and resources into surviving and reproducing, even at the cost of continuous suffering to themselves or others. Its main evolutionary advantage is that it does not need to spend any resources making the world a better place.

If given the choice, please don’t become a pure replicator and throw under the bus all the hard work that people throughout history have put into making the world not an entirely horrible place. Pure replicators come in many guises. While the term pure replicator may invoke images of cockroaches and viruses, the truth is that your modafinil-fueled, income-maximizing coworker may already be turning into one.

Considering that the dimension of *spiritual euphoria* is the most intense (and subjectively profound) source of conscious value, it would be a shame if our society exclusively optimized for linear logico-linguistic “high clarity” states of consciousness. Of all the drugs

3. Tragedy of the commons (*i.e.*, failures of cooperation). [16]

available, when balancing side effects and overall effectiveness, it is likely that modafinil-like compounds (*e.g.*, custom nootropics) give you the single largest economic edge within this society. Caffeine is already available to everyone, speed slowly kills you, and micro-dosed LSD makes you (believe it or not) too creative for most paying jobs. Is it possible to make the interesting and valuable states of consciousness the ones that are economically rewarded? Are we going to let the economic incentives in our society silently maximize the presence of modafinil-like states of consciousness?

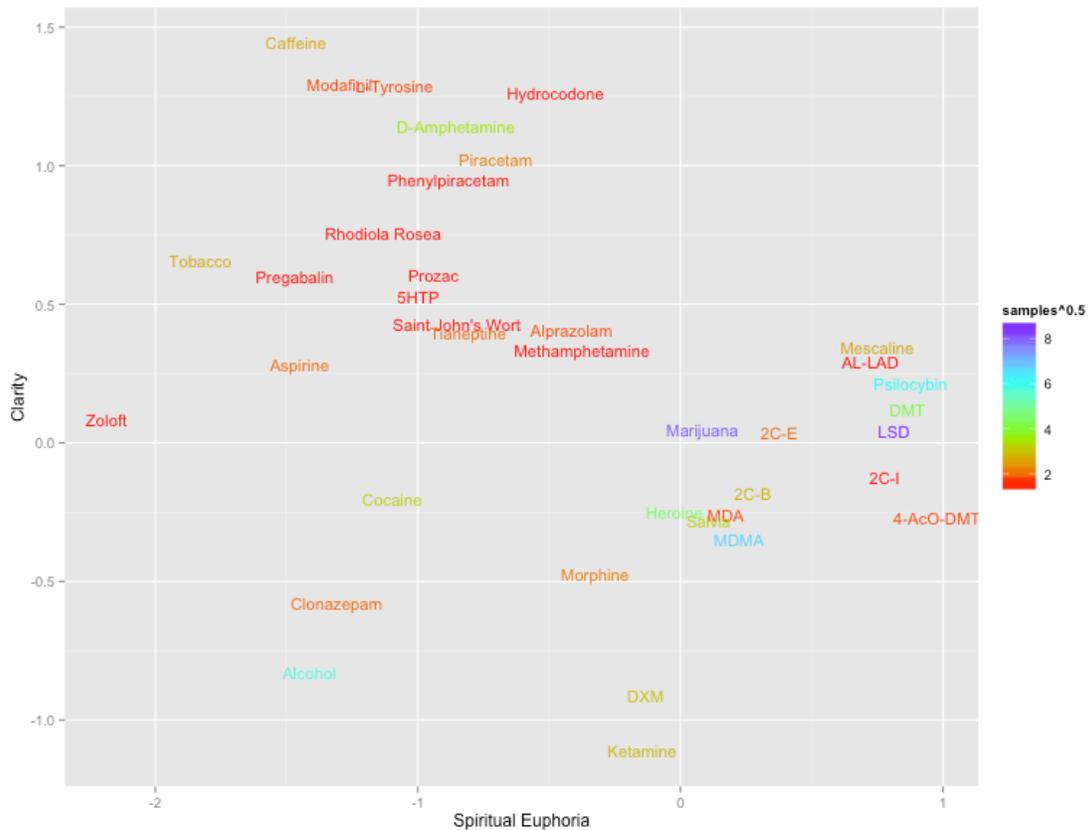


Figure 10: There is no known substance that enhances both “clarity” and “spiritual/philosophical euphoria” at the same time. It would be a shame if all the economy cared about was your level of clarity, for that would mean that modafinil [24] junkies [68] users [51] will rule the world. (Oh, wait... [21]). At the limit, such a world may be impervious to conceptual revolutions or caring about valence research.

In practice, unless digital AGI or nanotechnology takes over [11], pure replicators are going to need to interface with human and posthuman markets to gain any power. Although fashionable to think about nowadays, exotic nanotech and/or AI pure replicators may ulti-

mately be far easier to stop than pure replicators that disguise themselves as humans (*i.e.*, people who turn into empty shells of their former selves by embracing hyper-competitive Moloch memes [40] and their associated technologies). As we will see, the nature of future economic selection pressures may be the most important factor in whether or not we are taken over by armies of pure replicators.

Aren't We All Pure Replicators Already?

Tautologically, natural selection can *only* produce pure replicators. But this would be to think of the term in an unhelpful way that is not true to the spirit of the idea. This is why we defined *pure replicators* in terms of *indifference towards conscious states*. Most animals do indeed care a great deal about the valence of their own consciousness; after all, the motivational power of the pleasure-pain axis is the very reason why evolution recruited conscious valence to begin with. More so, sexual selection happens to have recruited introspection, aesthetics, benevolence and intelligence as fitness-indicators [32] (which explains why we are so keen on advertising these traits [38]). Brian Tomasik calls our times the *Age of Spandrels* [70] because we live in a period that is reaping the benefits of surplus production (still being below carrying capacity [46]) while silly non-optimal aesthetics inherited from our evolutionary past still survive. Interpersonal love, sexually selected hedonistic social rituals and ingrained prosocial implicit values may be evolutionary spandrels in the context of our economy, but (perhaps surprisingly) they are still part of our society. Hence, we today can enjoy watching movies, making love and thinking about philosophy. Our drive to delight in life is powerful enough to distract us from optimal economic participation, and our emotional wellbeing (which affects our economic participation) is still linked to events dealing with our level of pleasure outside work.

In contrast, the intelligent agents of the future may not be constrained to using the pleasure-pain axis to implement goal-oriented behaviors. One could envision scenarios like Robin Hanson's Age of EM [17] in which the most productive (and abundant) minds do 99.9999% of the work, and this work is boring 99.9999% of the time. These minds may work while in near-neutral states of consciousness that have either negligibly positive or even outright negative valence. The employees of this massive workforce are those individuals who are willing to do whatever they are told for 0.00001% vacation time and the opportunity to stay alive and multiply (in this case by copying the minds in digital servers). The employers may themselves not be particularly happy because they are also competing against other companies that cut down on costs as much as possible. If smiling does not increase one's productivity at one's job but it does waste precious calories and units of attention, then smiling will be abolished for purely economic reasons. In this scenario everyone is either employed and miserable (relative to our current standards), or unemployed and dying of starvation. We can thank those entities who were willing to completely sacrifice their own psychological depth (and freedom to explore the state-space of consciousness) for the sake of merely existing. The world now fails to produce any actual value in the form of meaningful states of consciousness and is over-saturated with modafinil-like consciousness.

Singleton and Moloch end-of-times scenarios tend to look pretty terrible because the worlds they present don't seem to contain reasons for anyone to care about valence.

But in this day and age we may be on the brink of reverse-engineering valence itself. Once we figure out the equation that takes as input quantum fields and outputs the conscious valence present in them [44], we will be able to quantify just exactly how bad our possible futures-at-the-limit will be depending on the economic selection pressures that we put in place today.

A desirable Singleton should at the very least care about states of high-valence and avoid negative valence states as much as possible. In a future article we will discuss some ideas for how to design an economic system based on cooperation that increases our chances of having ecologies of sustainable conscious entities who have the following properties: (1) they are free to explore the state-space of consciousness, (2) are social, and (3) have access to practically unlimited positive valence. But what if we are headed towards a perpetual Moloch (failure of cooperation) scenario?

Surviving in the Sundown of the Age of Spandrels

'In Time, any being that is spontaneous and alive will wither and die like an old joke.'

–WL 111. [25]

What would be a list of desirable traits that we want to have after acquiring complete control over our individual pleasure-pain axis? David Pearce [54] doesn't get tired of pointing out that the future does not belong to anti-natalists [59]. Their compassionate genes will be weeded out of the gene pool, and it will be their own compassionate sentimental fault. Similarly, full-blown single euphoria wireheading (as discussed above) is destined for oblivion unless it also happens to give you marketable skills.

We want to be able to both feel good and at the same time remain economically competitive (or we are going to be crowded out by pure replicators). Here is a list of traits that would help us have lives worth living without sacrificing our economic value:

1. Always being in a positive valence state (*i.e.*, remaining above hedonic zero).
2. A faithful/good enough internal simulation of one's environment (both physical and social).
3. The freedom to explore at will the known state-space of consciousness.
4. Capable of producing socially-useful information.
5. Free from unconscious bonds and protected against mind control..
6. Capable of exiting attractor states (affective, cognitive, and social, *e.g.*, thought loops).

7. Able to make others happier in ways they did not know were possible.

The first trait matters for ethical reasons: one needs to guarantee that the entities you bring into existence will always be happy to be alive. One should never compromise on the wellbeing of the beings one designs and gives birth to. If someone does, then we are again off to the races against pure replicators willing to suffer for a chance to exist.

The second trait is a requirement to survive in a physical world and a social economy (for obvious reasons).

The third trait is motivated by both ethical and practical reasons: as I understand it, having the ability to explore the known state-space of consciousness guarantees that you yourself can benefit from whatever awesome things people have made and discovered already. It guarantees that each individual will be able to experience the most valuable states (as judged by themselves at the time) without a preconceived notion of which states are the most ideal before experimenting on their own.

Being capable of experiencing any state of consciousness already discovered and understood will hopefully also turn out to be economically desirable. In order to be of relevance in the market of information about the state-space of consciousness you yourself will need to be an explorer and be up to date with what is in vogue. This opens up the possibility of a full-fledged qualia economy: when people have spare resources and are interested in new states of consciousness, anyone good at mining the state-space for precious gems will have an economic advantage.

In principle, the whole economy may eventually be entirely based on exploring the state-space of consciousness and trading information about the most valuable contents discovered doing so. The traits 4 through 7 are intended to address the complications that arise from the need to have social competence to survive in such an economy.

Society may ultimately converge into a system in which people are constantly in hyper-valuable states and the only way to become powerful is to invent new ways to improve upon the already highly-optimal state-spaces people are free to roam all the time. In this economy, people would also be motivated to help others succeed: Everyone benefits from making discoveries since every discovered state is made accessible to everyone.

How could we implement a conscious mind with these attributes? The task is indeed extremely demanding, and billions of dollars in R&D [34] will have to be invested before we have a silver-bullet genetic intervention that takes us in that direction. In the next centuries we are likely to see hundreds of thousands of researchers [30] experimenting with various cocktails, implants and genetic vectors hacking themselves in order to reliably improve their hedonic tone while also increasing their economic value.

In order to have any chance at living in such a society we need to make sure we won't be overrun by pure replicators in any of their gazillion different guises. To do so, we need to make sure we do not fall prey to any of the wireheading mistakes outlined above. And we

also need to make sure that we can give back to the world more than we take, so that the world is happy to have us around.

Wireheading Done Right: Stay Positive Without Going Insane

To remain economically relevant and subvert the rise of pure replicators, it is quintessential that one’s capacity to explore the state-space of consciousness is a marketable skill. Your imagination (for lack of a better term), should therefore work at an acceptable depth and speed from the point of view of the current economy of social relationships.

For example, we may want to wirehead ourselves so that our ability to experience fast euphoria is gated by slow euphoria: unless you have “satiated” your psychological need for resting you will not be able to feel hyper-motivated. Our desires are already state-specific, but the current network of transition probabilities between emotions facilitates the reinforcement of toxic local attractors (also called “death spirals” like states of depression or generalized anxiety). By re-engineering the network of transition probabilities between emotions and extracting the dysphoric components we might be able to guarantee a continuous flow between functionally and phenomenologically distinct modes of wellbeing.

Wireheading done right consists of having wonderful experiences all the time, but in such a way that you never feel compelled to stay where you are for too long. In addition, a good wireheading procedure should also allow you to keep learning useful information about the state-space of consciousness. Wireheading should not imply the end of learning. In brief, we suggest that we should change our brains so that by feeling great in a certain way you temporarily reduce the response to that particular kind of euphoria but also make it easier to enjoy some other kind. One would thus be incentivized to keep moving, and to never give up or to get stuck in loops.

Naturally one may be skeptical that perpetual (but varied) bliss is at all possible. After all, shouldn’t we already be there if such states were actually evolutionarily advantageous?

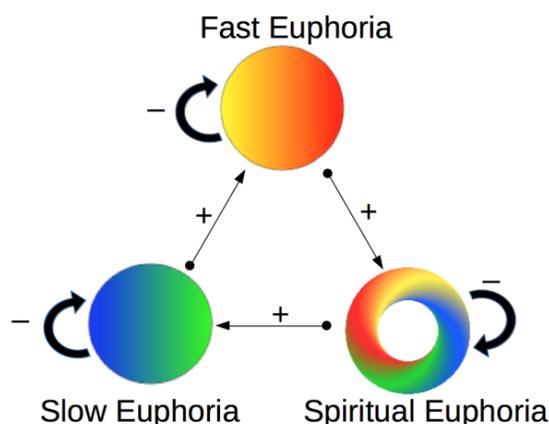


Figure 11: We want to rewire our reward architecture in such a way that as each kind of euphoria is instantiated a different one becomes more accessible. It is a virtuous cycle that delivers hearty amounts of craving-free euphoria. What’s reinforced is the flow between the types of euphoria rather than each kind on its own.

The problem is that the high-valence states we experience evolved to increase our inclusive fitness in the ancestral environment, and not in an economy based on gradients of bliss. Experiences are calorically expensive; in the African Savannah it may cost too many calories to keep people in novelty-producing hyperthymic states (even if one is kept psychologically balanced) relative to the potential benefits of having our brains working at the minimal acceptable capacity. In today's environment we have a surplus of calories which can be put to good use, *i.e.*, to exploring the state-space of consciousness and discovering socially (and hedonically) valuable states. Exploring consciousness may thus not only be aligned with real value (*c.f.*, valence realism [44]) but it might also turn out to be a good, rational time investment if you live in an experience-oriented economy. We are not particularly calorie-constrained nowadays; our states of consciousness should be enriched to reflect this fact.



Link and Leona [33] are two successful wireheads who are now happier than ever. They chose to have the following feedback network for their valence: fast makes it easier to feel spiritual, spiritual makes it easier to feel slow, and slow makes it easier to feel fast. Their primary state of consciousness cycles over a period of 24 hours. Here is their routine: They wake up and experience intense zest for life and work at full capacity making others happy and having fun. Then they go crazy creative in the afternoon, usually spending that time alone or with friends, and explore (and share) strange but always awesome psychedelic-

like states of consciousness. Finally, at night they just *relax to the max* (the healthy and genetically-encoded phenomenological equivalent of shooting heroin). They report having more agency than before, since now they feel that there is time to do everything they want, and moving from one activity to the next is easy and spontaneous. This kind of wireheading allows them to avoid loops, drops in motivation, or impoverished creativity and introspection. They are genetically engineered to *live the good life* without getting stuck in bad loops. The only thing they had to accept was that “hey, you don’t need to have all of the euphorias at once all the time”. By enjoying them one at a time you can guarantee a healthy mind, a healthy social life and a healthy economic output.

Positive Wireheading at the Evolutionary Limit

One of the main insights of evolutionary game theory is that strategies that have the best shot at being dominant for long periods of time have three properties: (1) they do well on average against other strong strategies, (2) do well against themselves, and (3) have an immune system against custom anti-strategies. This last condition may be skipped if we are not going to play for too long, but in the long run it is absolutely necessary. Custom anti-strategies may themselves be terrible against other strong strategies and even against themselves, so you may not realize they exist at first. For a while your strategy may be dominating the space with no signs of anything changing. But then you may start noticing that a tiny population of contrarians are beginning to grow exponentially. In no time you are defeated and the world breaks into chaos (since the contrarians may not be good at holding power). A classical illustration of this phenomenon comes from the evolutionary setup [26] of the Iterated Prisoner’s Dilemma. Here we can find some strategies that satisfy (1) and (2) but do not have a good immune system against free-loading strategies. The strategy *always-defect* is surprisingly effective in an ecosystem dominated by *always-cooperate*. Likewise the *Pavlov* [7] strategy can exploit the vulnerabilities of *tit-for-tat-with-forgiveness* [65] (though improved versions can counter it). Over time we see sporadic population booms and busts caused by cycles of cooperative eras collapsing under parasitic breakthroughs.

I posit that an ecology of wirehead minds that dedicate themselves to exploring (“mining”) the state-space of consciousness can be economically powerful if other people are willing to pay for the information about how to instantiate the high-valence qualia discovered during these explorations. A large group of cooperators that help each other explore the state-space of consciousness [36] satisfies condition (2) from the previous paragraph. But in order to satisfy condition (1) we need an environment in which knowledge about consciousness is marketable. The Super-Shulgin Academies [39] (*i.e.*, rational psychonautic collectives) of the future may concentrate all the *qualia research talent* in the world and reap the highest economic benefits while producing the largest amount of value, but they will only be able to do so if the surrounding society values their output. A society of pure replicators has no use for Super-Shulgin Academies, let alone Manhattan Projects of Consciousness [36] (*i.e.*, global concerted efforts to find and recruit benevolence-enhancing states of consciousness). But given the moral and hedonistic pursuits of our fellow humans we may still have a chance: making people happier than they currently are is a trillion-dollar industry [12] nowadays, and Super-Shulgin Academies may capitalize on this demand by selling valence-enhancing

technologies to the masses.

With regards to (3) we may happen to be lucky this time: Knowing all there is to know about the state-space of consciousness is the best way to prevent oneself from being outsmarted. Super-Shulgin Academies would invest heavily in researching ways to defend themselves against pure replicators. As part of its immune system, a Super-Shulgin Academy should only admit benevolent individuals as researchers. Benevolence, perhaps, is best implemented at the level of ontological qualia [29]: someone who believes that we are all the same consciousness is a lot less dangerous to others than someone who is solipsistic or self-centered. “Technology is destructive only in the hands of people who do not realize they are one and the same process as the universe.” [71] Rational agency and super-sentience in the hands of Open Individualists [27] (people who believe that we are all the same subject of experience [37]) could eventually allow us to bring about a good *qualia and valence-centric Singleton*.

But to bootstrap our way there we need to make sure that the organization would not die out even in an economy that isn’t already completely focused on consciousness (*i.e.*, to fulfill condition 1).

We are very lucky to live in the Age of Spandrels. We take for granted the fact that people around us like to watch movies, go to sports events, read novels, get drunk, listen to music, have sex, *etc.*, without realizing they could be investing all of that energy in figuring out how to make clones. We don’t usually realize that people’s atavistic inefficient hedonism is in fact our saving grace. (As an aside, I hope I am not inspiring anyone to go wild and do *silly non-optimal things*. My readership is capable of much more than that. Let us do *silly non-optimal things in the most optimal way possible*, by which I mean, let’s try to ensure that future beings care about valence and consciousness.) We should be thankful that we still have residual sexually-selected preferences for experientially rich lifestyles over pure and efficient dullness.

As long as we make intelligent use of today’s collective interest in exploring consciousness (in all of its guises, *e.g.*, art, philosophy, drugs) we still have a chance to create a sustainable economy of well-rounded wireheads that is worth living. The wirehead psychonaut collective would obtain most of its economic power from the discoveries made during the systematic exploration of the state-space of consciousness. If the public consumes these discoveries, then the strategy may be perfectly self-sustaining. The process itself would be beneficial, as we would discover new ways to make people happy, extend the freedom to inhabit any known state of consciousness, and increase our understanding of the universe.

One can even imagine that if Super-Shulgin Academies become the most powerful economic forces in the world, they may choose to create a massive “*wiki-consciousness*”: a library of all known states of consciousness completely accessible to anyone free of charge. Why would they do this given that their power comes from being able to sell this information? On the one hand it stabilizes people’s ability to gain power over each other (since the only way to gain power in such an economy is to sell information about the state-space). This incen-

tivizes people to actually find something new and valuable for everyone if they are aiming to become more powerful. And on the other hand, making the information freely available would also increase the quality of prospective members of the psychonaut collective due to widespread consciousness literacy.

If we play our cards right we may still have a chance of avoiding the pure replicators, Molochs, and Singletons that lurk in our forward light-cone. But to do so we need to stay grounded and avoid the pitfalls here discussed.

Conclusion

If we are to have a chance at surviving with a good quality of life in the sundown of the Age of Spandrels we will need to preemptively outcompete pure replicators. To do so we must avoid wireheading traps and take seriously future economic selection pressures, as they will determine who or what survives at the evolutionary limit. It is imperative that we take advantage of the current collective demand for valence and information about consciousness to fund ambitious consciousness research programs. Such programs will capitalize on this demand and kick-start a valence-centric market. In turn, scientific breakthroughs in this area may increase the percentage of the economy that is dedicated to exploring consciousness, which may reduce the opportunities for pure replicators to participate in the economy.

We need to act fast: if the economic demand for valence technologies disappears (or is low enough), we will find ourselves in a world in which exploring the state-space of consciousness is not profitable, a world where pure replicators win.

References

- [1] Change of basis, . URL https://en.wikipedia.org/w/index.php?title=Change_of_basis&oldid=949783669. Page Version ID: 949783669.
- [2] Color theory, . URL https://en.wikipedia.org/w/index.php?title=Color_theory&oldid=960841376. Page Version ID: 960841376.
- [3] Hedonic treadmill, . URL https://en.wikipedia.org/w/index.php?title=Hedonic_treadmill&oldid=962033777. Page Version ID: 962033777.
- [4] Markov property, . URL https://en.wikipedia.org/w/index.php?title=Markov_property&oldid=919906798. Page Version ID: 919906798.
- [5] New insight into how brain 'learns' cocaine addiction, . URL <https://www.sciencedaily.com/releases/2013/08/130801125035.htm>.
- [6] Ornstein–Uhlenbeck process, . URL https://en.wikipedia.org/w/index.php?title=Ornstein%E2%80%93Uhlenbeck_process&oldid=955419756. Page Version ID: 955419756.
- [7] Tit For Tat vs Pavlov, . URL <http://wiki.c2.com/?TitForTatVsPavlov>.

- [8] Valence utilitarianism, . URL <http://www.valence-utilitarianism.com/>.
- [9] What are some characteristics of meth addicts?, . URL <https://www.quora.com/What-are-some-characteristics-of-meth-addicts>.
- [10] Why Happiness Eludes Us...For Now, . URL <https://neurohacker.com/happiness-eludes-us-now>.
- [11] Gray goo, Apr. 2020. URL https://en.wikipedia.org/w/index.php?title=Gray_goo&oldid=950394098. Page Version ID: 950394098.
- [12] Happiness economics, June 2020. URL https://en.wikipedia.org/w/index.php?title=Happiness_economics&oldid=961482257. Page Version ID: 961482257.
- [13] S. Alexander. Intellectual Hipsters and Meta-Contrarianism. URL <https://www.lesswrong.com/posts/9kcTNWopvXFncXgPy/intellectual-hipsters-and-meta-contrarianism>.
- [14] S. Alexander. Archipelago and Atomic Communitarianism, June 2014. URL <https://slatestarcodex.com/2014/06/07/archipelago-and-atomic-communitarianism/>.
- [15] S. Alexander. Meditations On Moloch, July 2014. URL <https://slatestarcodex.com/2014/07/30/meditations-on-moloch/>.
- [16] S. Alexander. Misperceptions On Moloch, Aug. 2014. URL <https://slatestarcodex.com/2014/08/01/misperceptions-on-moloch/>.
- [17] S. Alexander. Book Review: Age of Em, May 2016. URL <https://slatestarcodex.com/2016/05/28/book-review-age-of-em/>.
- [18] S. Alexander. Why Were Early Psychedelists So Weird?, Apr. 2016. URL <https://slatestarcodex.com/2016/04/28/why-were-early-psychedelists-so-weird/>.
- [19] T. Antonio, S. R. Childers, R. B. Rothman, C. M. Dersch, C. King, M. Kuehne, W. G. Bornmann, A. J. Eshleman, A. Janowsky, E. R. Simon, M. E. A. Reith, and K. Alper. Effect of Iboga Alkaloids on μ -Opioid Receptor-Coupled G Protein Activation. *PLoS ONE*, 8(10), Oct. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0077262. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3818563/>.
- [20] F. Aricioglu, A. Means, and S. Regunathan. Effect of agmatine on the development of morphine dependence in rats: potential role of cAMP system. *European journal of pharmacology*, 504(3):191–197, Nov. 2004. ISSN 0014-2999. doi: 10.1016/j.ejphar.2004.10.011. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2923207/>.
- [21] M. Arrington. How Many Silicon Valley Startup Executives Are Hopped Up On Provigil? URL <https://social.techcrunch.com/2008/07/15/how-many-of-our-startup-executives-are-hopped-up-on-provigil/>.
- [22] H. C. Becker. Kindling in Alcohol Withdrawal. *Alcohol Health and Research World*, 22(1):25–33, 1998. ISSN 0090-838X. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6761822/>.

- [23] N. Bostrom. What is a Singleton?, 2005. URL <https://www.nickbostrom.com/fut/singleton.html>.
- [24] G. Branwen. Modafinil. Feb. 2009. URL <https://www.gwern.net/Modafinil>.
- [25] CCRU. Lemurian Time War. URL <http://www.ccru.net/archive/burroughs.htm>.
- [26] J. Chen, S.-I. Lu, and D. Vekhter. Axelrod’s Tournament. URL <https://cs.stanford.edu/people/eroberts/courses/soco/projects/1998-99/game-theory/axelrod.html>.
- [27] G. Cornwall. Open Individualism – Being Everyone, Oct. 2011. URL <http://phantomself.org/open-individualism-%e2%80%93-being-everyone/>.
- [28] A. G. Emilsson. How to secretly communicate with people on LSD, May 2015. URL <https://qualiacomputing.com/2015/05/22/how-to-secretly-communicate-with-people-on-lsd/>.
- [29] A. G. Emilsson. Ontological Qualia: The Future of Personal Identity, Dec. 2015. URL <https://qualiacomputing.com/2015/12/17/ontological-qualia-the-future-of-personal-identity/>.
- [30] A. G. Emilsson. The psychedelic future of consciousness, Mar. 2015. URL <https://qualiacomputing.com/2015/03/07/the-immediate-future-of-consciousness/>.
- [31] A. G. Emilsson. State-Space of Drug Effects: Results, June 2015. URL <https://qualiacomputing.com/2015/06/09/state-space-of-drug-effects-results/>.
- [32] A. G. Emilsson. Algorithmic Reduction of Psychedelic States, June 2016. URL <https://qualiacomputing.com/2016/06/20/algorithmic-reduction-of-psychedelic-states/>.
- [33] A. G. Emilsson. Beyond Turing: A Solution to the Problem of Other Minds Using Mindmelding and Phenomenal Puzzles, Nov. 2016. URL <https://qualiacomputing.com/2016/11/12/beyond-turing-a-solution-to-the-problem-of-other-minds-using-mindmelding-and-phenomenal-puzzles/>.
- [34] A. G. Emilsson. David Pearce on “Making Sentience Great”, Apr. 2016. URL <https://qualiacomputing.com/2016/04/13/david-pearce-on-making-sentience-great/>.
- [35] A. G. Emilsson. Information-Sensitive Gradients of Bliss, Aug. 2016. URL <https://qualiacomputing.com/2016/08/03/information-sensitive-gradients-of-bliss/>.
- [36] A. G. Emilsson. Peaceful Qualia: The Manhattan Project of Consciousness, Mar. 2016. URL <https://qualiacomputing.com/2016/03/29/peaceful-qualia-the-manhattan-project-of-consciousness/>.
- [37] A. G. Emilsson. Samadhi, Nov. 2016. URL <https://qualiacomputing.com/2016/11/03/samadhi/>.

- [38] A. G. Emilsson. So Why Can't My Boyfriend Communicate?, June 2016. URL <https://qualiacomputing.com/2016/06/17/so-why-cant-my-boyfriend-communicate/>.
- [39] A. G. Emilsson. The Super-Shulgin Academy: A Singularity I Can Believe In, Feb. 2016. URL <https://qualiacomputing.com/2016/02/04/the-super-shulgin-academy-a-singularity-i-can-believe-in/>.
- [40] A. G. Emilsson. Work Religion, May 2016. URL <https://qualiacomputing.com/2016/05/11/work-religion/>.
- [41] A. Higgins, M. Nash, and A. M. Lynch. Antidepressant-associated sexual dysfunction: impact, effects, and treatment. *Drug, healthcare and patient safety*, 2:141–150, Sept. 2010. ISSN 1179-1365. doi: 10.2147/DHPS.S7634. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3108697/>.
- [42] C. Iozzio. No Batteries Here: New Implants Can Charge Through Your Body's Own Tissue. URL <https://www.smithsonianmag.com/innovation/no-batteries-here-these-wireless-medical-implants-use-bodys-own-tissue-charge-180951612/>.
- [43] K. V.-A. Johnson and R. I. M. Dunbar. Pain tolerance predicts human social network size. *Scientific Reports*, 6(1):1–5, Apr. 2016. ISSN 2045-2322. doi: 10.1038/srep25267. URL <https://www.nature.com/articles/srep25267>.
- [44] M. Johnson. Principia Qualia. URL <https://opentheory.net/2016/11/principia-qualia/>.
- [45] P. Kuppens. Individual differences in the relationship between pleasure and arousal. *Journal of Research in Personality*, 42(4):1053–1059, Aug. 2008. ISSN 0092-6566. doi: 10.1016/j.jrp.2007.10.007. URL <http://www.sciencedirect.com/science/article/pii/S0092656607001365>.
- [46] J. Layton. Has Earth reached its carrying capacity?, Dec. 2008. URL <https://science.howstuffworks.com/environmental/green-science/earth-carrying-capacity.htm>.
- [47] T. Metzinger. Being No One, Jan. 2003. URL <https://mitpress.mit.edu/books/being-no-one>.
- [48] M. Mientka. Your Low Tolerance For Pain Can Be Blamed On Your Genes, Apr. 2014. URL <https://www.medicaldaily.com/pain-tolerance-has-genetic-basis-study-suggests-new-avenues-drug-treatment-277840>.
- [49] L. Muehlhauser. Better Disagreement. URL <https://www.lesswrong.com/posts/FhH8m5n8qGSShAgG/better-disagreement>.
- [50] Z. Oravecz, F. Tuerlinckx, and J. Vandekerckhove. A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*, 16(4):468–490, 2011. ISSN 1939-1463, 1082-989X. doi: 10.1037/a0024375. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0024375>.

- [51] Ozymandias. The World Is Mad, Oct. 2015. URL <https://thingofthings.wordpress.com/2015/10/30/the-world-is-mad/>.
- [52] D. Pearce. The Biointelligence Explosion How recursively self-improving organic robots will modify their own source code and bootstrap our way to full-spectrum superintelligence by, . URL <https://www.biointelligence-explosion.com/>.
- [53] D. Pearce. The Hedonistic Imperative, . URL <https://www.hedweb.com/hedethic/hedon1.htm>.
- [54] D. Pearce. HedWeb : paradise-engineering, . URL <https://www.hedweb.com/confi le.htm>.
- [55] D. Pearce. Hypermotivation : drugs, genes and electrodes, . URL <https://www.hedweb.com/wirehead/hypermotivation.html>.
- [56] D. Pearce. MDMA / Ecstasy : Utopian Pharmacology, . URL <https://www.mdma.net/>.
- [57] D. Pearce. Opioids: past, present and future, . URL <https://www.opioids.com/>.
- [58] D. Pearce. The Orgasmic Brain, . URL <https://www.paradise-engineering.com/brain/>.
- [59] D. Pearce. Review: Better Never To Have Been: the harm of coming into existence by David Benatar, . URL <https://www.abolitionist.com/anti-natalism.html>.
- [60] D. Pearce. Superhappiness: Ten Objections To Radical Mood-Enrichment, . URL <https://www.superhappiness.com/>.
- [61] D. Pearce. Wirehead hedonism versus paradise-engineering, . URL <https://www.hedweb.com/wirehead/index.html>.
- [62] C. Potts. Christopher Potts: Papers. URL <http://web.stanford.edu/~cgpotts/papers.html>.
- [63] J. A. Russell. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172, Jan. 2003. ISSN 0033-295X. doi: 10.1037/0033-295x.110.1.145.
- [64] sammibunny23. How I Feel When I'm Manic (Bipolar Disorder). URL <https://www.youtube.com/watch?v=50e9QRZZn1c>.
- [65] R. Sapolsky. Forgiving tit for tat and Pavlov strategy. URL https://www.youtube.com/watch?v=puN_vAsWipQ.
- [66] N. Scharping. Positive Vibes: Scientists Control Mouse Brains With Magnets. URL <https://www.discovermagazine.com/mind/positive-vibes-scientists-control-mouse-brains-with-magnets>.

- [67] M. Sudhof, A. Gómez Emilsson, A. L. Maas, and C. Potts. Sentiment expression conditioned by affective transitions and social forces. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, pages 1136–1145, New York, New York, USA, 2014. ACM Press. ISBN 9781450329569. doi: 10.1145/2623330.2623687. URL <http://dl.acm.org/citation.cfm?doid=2623330.2623687>.
- [68] supposedlysober. Modafinil - 'Completely Addictive'. URL <https://erowid.org/experiences/exp.php?ID=82145>.
- [69] S. Surve. Deep Brain Stimulation for Pleasure. URL <http://www.brainblogger.com/2009/01/08/deep-brain-stimulation-for-pleasure/>.
- [70] B. Tomasik. The Future of Darwinism. URL <https://reducing-suffering.org/the-future-of-darwinism/>.
- [71] A. Watts and A. C.-l. Huang. *Tao: the watercourse way*. Pantheon Books, New York, 1st ed edition, 1975. ISBN 9780394489018.
- [72] S. Young. The Dark Night, Nov. 2011. URL <http://shinzenyoung.blogspot.com/2011/11/dark-night.html>.