The Antifragile Brain

Luca Dellanna,
Luca@Luca-dellanna.com
22nd of December 2019

Abstract

This paper presents a new model for the human brain and its plasticity: the brain as an antifragile entity. In particular, parts of the brain are modeled as containers of mental patterns which undergo antifragile reactions to stressors such as emotions, pain and pleasure.

This model provides an intuitive way to think about neuroplasticity, dose-response to stimuli and conditioning.

Introduction

This paper is divided in four parts. The first part briefly introduces the concept of antifragility. The second part presents the model for the antifragile brain. In the third part, microscopic neurological processes are mapped upon the antifragile brain model, to justify its formulation as much more than a mere metaphor. The fourth part discusses some limitations to this model.

Antifragility

(Readers familiar with the concept can skip this section.)

An introduction to antifragility. A crystal glass is fragile: shocks damage it. Instead, the human body is antifragile: shocks make it stronger (provided it survives). For example, lifting weights helps building muscles.

The concepts of things which are *antifragile* and benefit nonlinearly from volatility was first proposed in (Taleb 2012).

Antifragility is due to two complementary phenomena. First, damage-triggered growth. Our muscles grow after their fibers are damaged, and lack of damage for long enough causes them to shrink (as it happens for sedentary people). Second, natural selection. In a species, the fitness of individuals is not uniformly distributed. Shocks cause the less fit members to die, raising the average fitness across the surviving population, eventually making it stronger as the newborns will be fitter on average.

What does the antifragile adapt to? The antifragile becomes stronger relative to what damaged it (it adapts to damage) and weaker relative to what didn't damage it (it adapts to lack of damage). This will be important in understanding how human behavior adapts.

The Antifragile Brain

One way to understand human behavior is to model parts of the brain – in particular, the Basal Ganglia – as a container of mental patterns, such as habits, reactions, beliefs, tastes, values, etc.

When we are exposed to a negative stressor – a negative emotion, pain, a disappointment – the mental pattern(s) responsible for exposing ourselves to it get damaged, ending up weaker or erased from our brain. As it will be described in the third section, "mental patterns getting damaged" corresponds to the synapses increasing their likelihood of being expressed becoming weaker.

Similarly, when we are exposed to a positive stressor – a positive emotion, pleasure, a nice surprise – the mental pattern(s) responsible for exposing ourselves to it get reinforced.

This seems intuitive and is a well-known process. Is it necessary to use antifragility to describe this behavior?

There are several reasons and advantages for doing so:

- Any entity comprising multiple components which can be independently damaged and (re)grown is necessarily also antifragile (Dellanna, 2019b). The brain satisfies both requirements and does so in multiple ways: both synapses and beliefs are contained in the brain, get damaged or weakened by external events and can be (re)grown.
- Antifragility predicts the sigmoid characteristic of the dose-response observed in how human behavior is often affected by external and internal events.
- Antifragile entities which do not get damaged for long enough by stressors threatening survival maladapt. This predicts the outcomes of conditions which commonly go

under the names of "growing up in a golden cage".

- Antifragility is a bottom-up phenomenon directly derived from Lindy fields such as physics and mechanics. Phenomena of this kind tend to be rather robust.

The Neurology of the Antifragile Brain

This section will present some ways in which the brain behaves like an antifragile entity.

In the Basal Ganglia, external events with a better or worse outcome than predicted generally lead to a dopamine-levels-mediated increase or decrease (respectively) of the expected outcome associated to a given behavior in a given context. For example, eating a better-than-expected ice-cream tends to increase the likelihood I'll want to eat an ice-cream whereas receiving less-than-expected thank-yous from my neighbor for having helped him start his dead car battery might decrease my desire to do the same in the future, would the occurrence arise again (the context being equal).

In antifragile entities, strengthening or weakening depend not only from the magnitude of the external event, but also on its magnitude respect to the current strength of the antifragile entity and on its subjective perception of the event. For example, whether lifting 5kg for 10 times will make my muscle grow depends on how strong I already am and on how the weight is applied (what the lifting movement is). Similarly, in the Basal Ganglia, whether an outcome reinforces or weakens a mental pattern depends on its current strength (i.e., the expected emotional outcome it produces) and on how that outcome is subjectively perceived.

The similarity described above is not superficial pattern-matching, but the result of both phenomena being instances of antifragility: what happens every time there is an entity made of components which can be (re)grown and subject to stressors. In the human brain, there are multiple instances of antifragile entities: the Basal Ganglia (comprising synapses[1]), beliefs about the expected emotional outcome of actions (comprising synapses) and macro-beliefs (comprising other beliefs).

A partial model

The brain is an extremely complex entity. This model, while useful to understand *some* of the bottom-up processes behind human behavior and the plasticity of the mind, it does not describe all processes nor all parts of the brain.

In particular, this model can be useful to understand how the Basal Ganglia adapts to external events in order to increase the fitness of its patterns.

References

[Dellanna 2019a]: Dellanna, L. (2019). Techniques for The Emergence of Meaning in ML

[Dellanna 2019b]: Dellanna, L. (2019). The Dynamics of Risk-Taking.

[Taleb 2018]: Taleb, N.N. (2012). Antifragile.

Also, the contents of this paper will be expanded and published in the author's upcoming book, "The Control Heuristic, 2nd Edition".

[1] In the case of the Basal Ganglia, for the purposes of antifragility, the components to consider are synapses and not neurons. This because external events such as pleasurable or disappointing experiences damage synapses, not neurons.