



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
Main Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2021

Unlocking the secrets of self-awareness

Ruff, Christian C

DOI: <https://doi.org/10.1126/science.abh2030>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-205744>

Journal Article

Accepted Version

Originally published at:

Ruff, Christian C (2021). Unlocking the secrets of self-awareness. *Science*, 372(6541):470.

DOI: <https://doi.org/10.1126/science.abh2030>

Unlocking the secrets of self-awareness

Long neglected by researchers, metacognition is increasingly in the scientific spotlight

By Christian C. Ruff

For millennia, religious thinkers and philosophers have cited humanity's self-awareness, our ability to think about our own mind and character, as being key to the uniqueness of our species. Carl Linnaeus' groundbreaking biological taxonomy (1) likewise characterized our genus by the words: "Homo. Nosce te ipsum." ("Man. Know thyself.").

Given our long-standing interest in self-awareness, it is surprising how little science has traditionally had to say about it. What features of our brains enable us to think about ourselves? What are our strengths and weaknesses in this respect and how do they influence how we decide, learn, and interact? Can we train self-awareness and how does this improve our performance? In the last three decades, however, research addressing such questions has been picking up pace. In *Know Thyself*, cognitive neuroscientist Stephen Fleming synthesizes this multifaceted research into an admirably coherent narrative and outlines how the resulting knowledge may be applied to solve societal problems.

Writing about self-awareness is challenging since concepts like "self" and "awareness" —let alone the combination thereof— are hard to define. The book does not get lost in this epistemological Bermuda triangle but conceptualizes self-awareness as the set of mental and brain processes that keep track of our percepts, thoughts, and actions.

Not all these "meta-cognitive" processes concern the self in a philosophical sense, Fleming notes, and not all of them need to be conscious. A helpful metaphor in the book compares our brain to a flying plane that is largely guided by autopilot technology but can be flexibly controlled by the pilot whenever needed. For our behavior, the autopilot is the unconscious, "implicit" part of meta-cognition and the pilot is the "explicit" metacognition that we can consciously report.

Fleming begins by summarizing the psychology and neuroscience of these meta-cognitive processes. Implicit meta-cognition, he notes, is evident in many seemingly low-level brain processes, ranging from the sensory brain cells that signal the uncertainty associated with particular percepts, to brain cells that activate when we commit action errors (think mistyping on a keyboard). All of these implicit signals can be read out in the service of explicit meta-cognition, when, for example, we need to judge our confidence in having chosen the right action. This latter ability depends on specific brain areas in prefrontal cortex and is independent of the basic perceptual and motor abilities it serves to monitor.

Explicit meta-cognition, meanwhile, depends on our ability to think about mental states of others; an ironic twist nicely summarized by the caption of a cartoon that appears in the book. ("Of course I care about how you imagined I thought you perceived I wanted you to feel.")

It is eye-opening to realize how many fields of human endeavor depend not just on our skills and knowledge, but also on our ability to estimate our competence. Obvious examples can be found in education, politics, the legal system, corporate decision making and leadership, news and social media, and, indeed, in any collaboration in which people pool their expertise. The book illustrates the role of meta-cognition in these diverse fields with elegant combinations of philosophical considerations, basic science findings, and more applied examples.

Fleming even ventures into the near future, sketching how artificial intelligence with superhuman computational abilities but no self-awareness may become disconnected from humanity at best and outright catastrophic at worst. Emerging ideas on how we may address this problem—for example, by endowing intelligent machines with coarse self-awareness or by ensuring that self-aware humans remain at the helm—only serve to prove how little we have appreciated our own prodigious meta-cognitive abilities.

In the end, the book makes a convincing case that self-awareness is a key feature of human existence and that our growing knowledge about it will be important for addressing many of our societies' problems. One may quibble that the book somewhat understates this point, since it focuses on meta-cognition and does not cover our ability to monitor our emotions, another key aspect of self-awareness that has major implications for health and well-being. However, the literature on this topic is so diverse that doing it justice would likely require several additional volumes. As it stands, Fleming's book finally heaves meta-cognition into a long-deserved place in the scientific spotlight.

REFERENCES AND NOTES

1. C. Linnaeus, "HOMO." in *Systema naturae, sive regna tria naturae systematice proposita per classes, ordines, genera, & species* (Haak, Leiden, NL, 1735),