

Recensioni

Andy Clark

The experience machine: How our minds predict and shape reality

Penguin Books, New York 2024

Pagine: 304; € 13,00

In his latest book, *The experience machine*, Andy Clark presents a revolutionary paradigm that sees the human mind as a sophisticated predictive machine. This vision has its roots in a historical evolution starting from the cognitive psychology of the '50s and '60s, passing through David Marr's computational theories, and reaching contemporary perspectives, in which the brain is no longer conceived as a mere passive receiver of stimuli but as an active organ that anticipates and interprets events. At the heart of this conception is the concept of "predictive error", which is the discrepancy between internally generated expectations and sensory information. Far from being considered a flaw, this error represents a fundamental signal that enables the brain to continuously update and refine its internal model. The internal model is a dynamic representation of the world and the body, constantly updated based on past experiences and predictions about the future. It allows the brain to interpret and make sense of sensory information by comparing it with expectations and, when discrepancies arise, modifying its representations to improve accuracy. In this way, the perception of the world becomes an active construction in which predictions and inferences play a central role. Clark deepens the analysis by going beyond the basic mechanisms of predictive processing, exploring its implications for action, psychological disorders, and consciousness, and situates the discussion within a broader philosophical debate that embraces theories of the extended mind and embodied cognition, highlighting how the body, the environment, and predictive processes interact to shape our experience of the world.

The book is divided into chapters, each organized into paragraphs that make the reading fluid, pleasant, and stimulating. Clark's style is informal yet firmly anchored in scientific dissemination, allowing readers to easily follow clear and accessible arguments for most of the text. In the first chapter, the author introduces the concept of the "predictive mind", the idea that the brain does not merely passively receive stimuli, but actively anticipates and organizes incoming information. An emblematic and practical example provided to illustrate this function is what the author calls a "Mooney image". Initially, an ambiguous image composed of patches is presented, in which what it represents is not immediately recognizable. Only

after viewing an explanatory photograph – for example, that of a dalmatian – is the brain able to reorganize the ambiguous image, reinterpret it, and render it clear. In other words, the perception of the image changes radically: what was initially confused and meaningless becomes clearly identifiable thanks to the brain's predictive action, which integrates the explanatory information with its own expectations. A similar phenomenon is observed in "sine-wave speech": an acoustic signal, initially ambiguous and lacking a clear linguistic structure, is reorganized by the predictive brain after listening to the corresponding real sentence. In this way, the phenomenological experience of the signal changes dramatically, making the message much clearer and more distinct than in its initial form. Through these and many other examples, Clark demonstrates how the predictive mind uses prior knowledge and expectations to order sensory information, laying the groundwork for further exploration in the subsequent chapters.

In the second chapter, the text delves into a more technical discussion, addressing topics such as pain, schizophrenia, and post-traumatic stress disorder, and illustrating how the model of the predictive mind can provide a fundamental and, in some cases, preventive interpretative key for other psychological and psychiatric disorders. To support this, Clark introduces another crucial element for the entire theory: the brain's best estimate of the reliability and significance of stimuli. In the language of predictive processing, this estimate is defined as "precision" and is integrated as a variable weighting factor on both predictions and sensory stimuli. According to these models, predictive brains are constantly engaged in estimating precision and modifying the way they handle both sensory evidence and their own predictions. As the author states: «this means we need to think not just about our brain's predictions and the incoming sensory evidence, but also about the way these estimates of precisions flexibly alter the balance of power between them» (p. 38).

From a technical point of view, "precision" represents a weighting factor capable of amplifying or attenuating different aspects of information processing. In the brain, the precision-weighting mechanism involves, among other things, the coordinated action of complex neurochemical systems, centered on dopamine and other chemical messengers. The coordinated action of these systems amplifies certain components of neuronal activity at the expense of others, acting like a volume control that regulates downstream influence over entire populations of neurons. Since precision is continuously estimated across all neuronal populations, variations in its assessment affect post-

synaptic influence patterns, determining which pieces of information, whether originating from the external world or generated internally, will have the greatest impact in guiding further processes and actions. In other words, precision fluctuations govern, moment by moment, which elements of our knowledge and perceptions will become decisive in subsequent processing, highlighting the intimate connection between precision and attention.

In the third and fourth chapters, Clark expands the scope of the predictive mind theory by thoroughly examining the role of bodily predictions in action. The author argues that the estimates and predictions formulated by the brain in response to environmental stimuli are not limited to mere sensory processing but constitute the basis for goal-directed behavior. According to this model, achieving a goal implies using the predicted outcomes to structure the most appropriate action to realize them. For example, consider the simple act of putting on a hat. In this case, the brain predicts the “sensory signature” of the desired action – that is, the sensation of moving the arm in a coherent manner and the hat fitting on the head – and initiates the “predictive error correction cycle”. This cycle coordinates the body’s movements, making the prediction itself capable of guiding action to minimize the error between the expected and perceived state. In more complex contexts, such as pursuing long-term goals, the brain operates on two levels: on one hand, it minimizes immediate errors related to sensory predictions, and on the other, it relies on an internal model with greater “temporal depth”. This extended model enables the prediction of future consequences of current actions, steering behavior in a strategic and adaptive way. Ultimately, goal-directed behavior emerges from the brain’s ability to continuously integrate and update its predictions, guiding action through a dynamic balance between what is perceived and what is expected.

Beyond this, Clark broadens his analysis by introducing a crucial element: the ways in which information and predictions about the external world interact with those concerning the body’s internal state, such as heartbeat, thirst, or hunger. In this integrated network, internal and external predictions offer new clues about the nature and origin of emotions and feelings. The entire mental life reflects the incessant and profoundly anticipatory activity of the brain engaged in “body-budgeting”, that is, managing internal resources to maintain homeostasis. To enable this kind of anticipatory control, the predictive models that shape human and animal behavior must be as directed inward as they are outward. Every brain region involved in generating and experiencing emotions is closely connected with circuits dedicated to regulating the body’s balance.

In a brief interlude, the author presents, even though in a relatively brief manner, several accounts and theoretical proposals on how the predictive mind theory might help explain the phenomenon of consciousness. Given the overall structure of the book, it is understandable that this vast and complex area is not explored exhaustively. Clark’s intent appears to be to offer a solid and well-argued general overview of the predictive mind hypothesis, capable of providing an all-encompassing vision of the research project and spanning multiple, intricate strands of study and reasoning.

In the final chapters, the author revisits further practical and theoretical examples, offering in-depth clarifications on what has been introduced earlier. This section explores topics ranging from the world of media to the social realm and even animal ethology, presenting accounts of theoretical frameworks and empirical experiment results enriched by philosophical reflections that encompass epistemological and scientific aspects. Particular attention is devoted to the extended mind theory, which posits that many human objects and artifacts can directly influence and function as true functional extensions of cognitive processes. According to this perspective, predictive minds are naturally adept at exploiting the opportunities offered by action cycles aimed at gathering information to help them achieve their goals. In other words, thanks to the fundamental principles that enable predictive brains to select actions capable of generating useful information using every available resource, our constructed world – made of objects and artifacts – can, in certain contexts, substitute for, transform, and enhance functions that were once performed exclusively by the brain. This vision underlines how the external environment and technologies are not merely passive supports, but integrated elements in a dynamic cognitive system, capable of extending and enhancing mental capacities.

Finally, in the last chapter titled *Hacking the predictive machine* (p. 181), Clark proposes a series of theoretical and practical accounts illustrating how the brain’s predictive capacity can be modified or influenced, positively or negatively, through various procedures and phenomena. This is not so much a set of practical tips to follow as it is an explanation of how certain processes can alter our predictions. Among the examples provided, the author cites the placebo effect (repeated several times throughout the book), self-encouragement and affirmation, as well as highlighting the influence on the brain of some active principles of hallucinogens, such as psilocybin, and even the use of virtual reality as a potential treatment for certain psychopathologies.

In conclusion, the book offers a general, detailed, and well-argued overview of the predictive

mind theory. The reading is extremely stimulating and the field of research, in addition to being fascinating, appears highly promising in light of the theoretical and practical arguments presented.

However, the text touches on some significant philosophical themes but only in a brief way. For instance, the treatment of the “self” is limited, leaving unresolved crucial questions regarding personal identity and the “sense of agency”. It would be interesting to explore whether deep connections exist between the sense of agency and the predictive model, analyzing how the internal predictive model might contribute to the formation of the self and whether it represents an indispensable element in defining the experiential subject. Similarly, the discussion would benefit from a more thorough exploration of language-related implications. It would be useful to investigate how the predictive model might enhance language comprehension from both semantic and pragmatic perspectives, based on predictions made about the linguistic system.

In this context, it might be significant to analyze whether the conscious experience of “hearing a voice in one’s head” is closely linked to such linguistic system predictions. A detailed examination of this particular linguistic phenomenon, known as inner speech, could offer valuable insights into how internal predictions shape linguistic processing and thought formation. Finally, the potential of the predictive mind model to characterize

higher cognitive functions – such as abstract reasoning, creativity, and metacognition – remains an area that would be potentially useful to explore further. A detailed analysis of these functions could help clarify how predictive processing underpins the various and complex mental faculties.

While the book leans more toward cognitive science than philosophy, this is not necessarily a limitation. Although some aspects – such as the theme of consciousness and the previously mentioned ones – remain partially unresolved, the work offers fundamental insights for deeper reflection on the subject.

The predictive mind theory is gaining significant traction in both cognitive science and philosophy, and this essay clearly explains the reasons for its success. At times, the division into numerous sub-sections, accompanied by an extensive series of examples, can make the reading somewhat dense for those unfamiliar with the field, due to the abundance of technical terms and complex reflections. Nevertheless, for anyone interested in exploring the predictive mind, this essay represents an excellent starting point, both as an introduction to this fascinating perspective and as a way to expand knowledge in a field that spans multiple areas of research and reasoning.

Francesco Giovacchini
Dipartimento di Scienze Sociali, Politiche e Cognitive
Università degli Studi di Siena